



**MODEL ADITIF TERGENERALISIR SEMIPARAMETRIK
DENGAN MENGGUNAKAN PEMULUSAN *LOESS* PADA
KASUS PENDUGAAN CURAH HUJAN**

SKRIPSI

oleh

**Rifka Wahyuni
111810101011**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016**



**MODEL ADITIF TERGENERALISIR SEMIPARAMETRIK
DENGAN MENGGUNAKAN PEMULUSAN *LOESS* PADA
KASUS PENDUGAAN CURAH HUJAN**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan Program Studi Matematika (S1)
dan mencapai gelar Sarjana Sains

oleh

**Rifka Wahyuni
111810101011**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016**

PERSEMBAHAN

Skripsi ini saya persembahkan untuk :

1. Allah SWT yang telah memberikan semua kemudahan dan kesempurnaan dalam kehidupan ini;
2. Ayahanda Muji Slamet dan Ibunda Ninik Mulyani, SE. yang telah memberikan doa, cinta dan semangat;
3. Adik tercinta Widya Puji Saputro yang telah memberikan dukungan;
4. Seluruh guru dan dosen sejak teman kanak-kanak hingga perguruan tinggi yang telah membimbing keingintahuan serta memberikan ilmu kepada saya dengan penuh perhatian dan kasih sayang;
5. Almamater Jurusan Matematika FMIPA Universitas Jember, SMA Negeri 4 Jember, SMP Negeri 1 Rambipuji, SDN 2 Rambipuji, serta TK ABA 1.

MOTTO

“Musuh yang paling berbahaya atas dunia ini adalah penakut dan bimbang. Teman yang paling setia, hanyalah keberanian dan keyakinan yang teguh.”

(Andrew Jackson) *)

“Maka sesungguhnya bersama kesulitan ada kemudahan. Sesungguhnya bersama kesulitan ada kemudahan. Maka apabila engkau telah selesai (dari sesuatu urusan), tetaplah bekerja keras (untuk urusan yang lain). Dan hanya kepada Tuhanmulah engkau berharap.”

(Q.S. Al-Insyirah, 6-8) **)

*) <http://www.maribelajarbk.web.id/2015/03/contoh-motto-terbaru-dalam-skripsi.html>

**) Departemen Agama Republik Indonesia. 1997. Al-Qur'an – Al-Karim dan Terjemahannya. Solo: CV. Pustaka Mantiq.

PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Rifka Wahyuni

NIM : 111810101011

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Model Aditif Tergeneralisir Semiparametrik Dengan Menggunakan Pemulusan *Loess* Pada Kasus Pendugaan Curah Hujan“ adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada instansi mana pun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenar-benarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Maret 2016

Yang menyatakan,

Rifka Wahyuni
NIM. 111810101011

SKRIPSI

**MODEL ADITIF TERGENERALISIR SEMIPARAMETRIK DENGAN
MENGUNAKAN PENGHALUS *LOESS* PADA
KASUS PENDUGAAN CURAH HUJAN**

Oleh

Rifka Wahyuni
111810101011

Dosen Pembimbing Utama : Drs. Prof. I Made Tirta, M.Sc, Ph.D

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si, M,Si

PENGESAHAN

Skripsi yang berjudul “**Model Aditif Tergeneralisir Semiparametrik Dengan Menggunakan Pemulusan *Loess* Pada Kasus Pendugaan Curah Hujan**” telah diuji dan disahkan pada:

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas
Jember

Tim Penguji,

Ketua,

Sekretaris,

Prof. Drs. I Made Tirta, M.Sc., Ph.D
NIP. 193912201985031002

Dian Anggraeni, S.Si., M.Si.
NIP. 198202162006042002

Penguji I,

Penguji II,

Dr. Alfian Futuhul Hadi, S.Si., M.Si.
NIP. 197407192000121001

Kosala Dwija Purnomo, S.Si., M.Si.
NIP. 196908281998021001

Mengesahkan,

Dekan,

Drs. Sujito, Ph.D
NIP. 196102041987111001

RINGKASAN

Model Aditif Tergeneralisir Semiparametrik dengan Menggunakan Penghalus Loess Pada Kasus Pendugaan Curah Hujan; Rifka Wahyuni 111810101011; 2015; 41 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Model aditif tergeneralisir atau yang sering disebut dengan *Generalized Additive Models* (GAM) adalah model statistika yang merupakan perluasan dari GLM dan model aditif dengan mengganti prediktor linier menjadi prediktor aditif serta distribusi yang digunakan adalah distribusi keluarga eksponensial. Dalam penggunaannya, GAM sering digunakan untuk memodelkan data dengan menggunakan estimasi parameternya secara semiparametrik maupun nonparametrik. Pendekatan estimasi parameter ini ada 2, yaitu parametrik dan nonparametrik. Pendekatan semiparametrik merupakan gabungan antara parametrik dan nonparametrik. Pada GAM, pengaplikasian bentuk semiparametrik pada data juga dapat dilakukan dengan menggunakan pemulusan sebagai prediktor aditifnya ketika data yang digunakan tidak memenuhi asumsi kelinieran. Pemulus yang digunakan dalam penelitian ini adalah pemulus loess.

Pada penelitian ini, data yang digunakan adalah intensitas curah hujan yang terjadi di Indonesia yang termasuk dalam data kontinu. Tujuannya adalah untuk mengaplikasikan GAM dengan pemulus loess serta mengetahui pengaruh tekanan udara, suhu udara dan kelembaban udara terhadap curah hujan dengan menggunakan model aditif tergeneralisir semiparametrik dengan pemulus *loess*. Selanjutnya, dianalisis data curah hujan sebagai variabel respon dengan kelembaban (X_1), suhu (X_2) dan tekanan (X_3). Analisis ini dilakukan menggunakan bantuan software R dengan paket `gam` dan `MASS`.

Hasil analisis pada penelitian ini menunjukkan bahwa dengan menggunakan aplikasi GAM dengan pemulus loess didapatkan distribusi yang terbaik yaitu distribusi Gamma dengan fungsi penghubung log, *span* (rentang) terbaik yaitu dan derajat polinomial yang cocok untuk data curah hujan yaitu 2 dengan variabel yang diberi pemulus adalah 2 sekaligus, yaitu kelembaban dan suhu. Berdasarkan model terbaik yang telah didapatkan menunjukkan bahwa hasil dari pemberian pemulus loess mengakibatkan bahwa kelembaban, suhu dan tekanan di udara berpengaruh signifikan terhadap curah hujan yang terjadi di Indonesia. Model terbaik untuk curah hujan di Indonesia dengan menggunakan GAM dan pemulus loess adalah :

$$\log(Y) = 14,192 + 0,065X_1^* + 0,285X_2^* - 0,019X_3$$

PRAKATA

Puji syukur kehadirat Tuhan Yang Maha Esa atas segala rahmat, taufik dan hidayahNya sehingga penulis dapat menyelesaikan penulisan tugas akhir yang berjudul “Model Aditif Tergeneralisir Semiparametrik dengan Menggunakan Pemulusan *Loess* Pada Kasus Pendugaan Curah Hujan”. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan strata satu (S1) pada Jurusan Matematika, Fakultas Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan tugas akhir ini tidak luput dari bantuan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada :

1. bapak Prof. Drs. I Made Tirta, M.Sc., Ph.D selaku Dosen Pembimbing Utama dan Ibu Dian Anggraeni, S.Si., M.Si. selaku Dosen Pembimbing Anggota yang telah bersedia meluangkan waktu, memberikan saran serta bimbingannya dalam penulisan skripsi ini;
2. bapak Dr. Alfian Futuhul Hadi, S.Si., M.Si. selaku Dosen Penguji I dan Bapak Kosala Dwija Purnomo, S.Si., M.Si. selaku Dosen Penguji II yang telah memberikan kritik serta sarannya terhadap penulisan skripsi ini;
3. seluruh staf pengajar jurusan Matematika Fakultas MIPA Universitas Jember yang telah memberikan ilmu serta bimbingannya sehingga penulis dapat menyelesaikan skripsi ini;
4. kakek Almarhum Wakirun dan Almarhum Rabil serta nenek Almarhumah Sri Utami dan Almarhumah Tiamah tercinta yang senantiasa memberi inspirasi;
5. ayahanda Muji Slamet dan ibunda Ninik Mulyani tercinta yang senantiasa memberi doa, semangat dan kasih sayang;
6. adikku tercinta Widya Puji Saputro serta seluruh keluarga yang telah memberikan doa dan semangat;
7. sahabat-sahabatku Aviv, Arif, Hiqma, Nisa, Alan dan Heni yang telah memberikan semangat dan motivasi;

8. keluarga besar KRAMAT'11, sahabat REMPONGERS (Rani, Rima, Retna, Rika, Renny, Anjar, Fia, Diana dan Diah), juga KKN Pondokrejo (Uuk, Eza, Muliawan, Diana, Dila, Ita) yang telah meluangkan waktu serta memberikan dukungan, keceriaan dan canda tawa yang telah diberikan selama masa studi;
9. semua pihak-pihak yang tidak dapat saya sebutkan satu per satu.

Skripsi ini juga tidak lepas dari kekurangan dan kesalahan baik isi maupun susunannya. Oleh karena itu, penulis menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Penulis berharap semoga skripsi ini dapat memberi manfaat bagi pembaca dan kehidupan.

Jember, Maret 2016

Penulis

DAFTAR ISI

| | Halaman |
|--|---------|
| HALAMAN JUDUL | i |
| HALAMAN PERSEMBAHAN | ii |
| HALAMAN MOTTO | iii |
| HALAMAN PERNYATAAN | iv |
| HALAMAN PEMBIMBING | v |
| HALAMAN PENGESAHAN | vi |
| RINGKASAN | vii |
| PRAKATA | ix |
| DAFTAR ISI | xi |
| DAFTAR GAMBAR | xiii |
| DAFTAR TABEL | xiv |
| BAB 1. PENDAHULUAN | 1 |
| 1.1. Latar Belakang | 1 |
| 1.2. Rumusan Masalah | 3 |
| 1.3. Tujuan Penelitian | 3 |
| 1.4. Manfaat Penelitian | 3 |
| BAB 2. TINJAUAN PUSTAKA | 4 |
| 2.1. Model Linier | 4 |
| 2.2. Model Linier Tergeneralisir | 4 |
| 2.3. Model Aditif | 5 |
| 2.4. Model Aditif Tergeneralisir | 6 |
| 2.5. Local Regression (LOESS) | 9 |
| 2.6. Akaike's Information Criterion (AIC) | 12 |
| 2.7. Model Semiparametrik | 12 |
| 2.8. Curah Hujan | 13 |

| | |
|--|----|
| BAB 3. METODOLOGI PENELITIAN | 15 |
| 3.1. Ilustrasi Data | 15 |
| 3.2. Analisis Data | 15 |
| 3.2.1. Struktur Fungsi GAM pada R..... | 15 |
| 3.2.2. Metode Analisis Data..... | 17 |
| BAB 4. HASIL DAN PEMBAHASAN | 19 |
| 4.1. Deskripsi Data | 19 |
| 4.2. Eksplorasi Data Curah Hujan | 20 |
| 4.3. Pengepasan Model Menggunakan GAM dengan Metode Pemulusan Loess | 24 |
| 4.3.1. Pemilihan Distribusi Data..... | 24 |
| 4.3.2. Analisis Data Menggunakan Pemulus Loess | 27 |
| a. Penentuan Derajat Polinomial..... | 27 |
| b. Penentuan <i>Span</i> | 28 |
| BAB 5. PENUTUP | 37 |
| 5.1. Kesimpulan | 37 |
| 5.2. Saran | 37 |
| DAFTAR PUSTAKA | 38 |
| LAMPIRAN | 39 |

DAFTAR GAMBAR

| | Halaman |
|--|---------|
| Gambar 2.1 Ilustrasi Pemulus Loess | 9 |
| Gambar 2.2 Ilustrasi Pemulus Loess Dengan <i>Span</i> Yang Berbeda | 10 |
| Gambar 3.1 Langkah-langkah Menganalisis Data | 18 |
| Gambar 4.1 Hubungan Antarvariabel pada Data Tentang Curah Hujan dan Variabel Eksplanatori | 21 |
| Gambar 4.2 Surface yang Dihasilkan Setelah Diberi Pemulusan | 30 |
| Gambar 4.3 Kontur Hasil Model Setelah Diberi Pemulus Loess | 30 |

DAFTAR TABEL

| | Halaman |
|--|---------|
| Tabel 4.1 Nilai AIC Model Terbaik Berdasarkan Fungsi Penghubung Distribusinya | 26 |
| Tabel 4.2 Nilai AIC untuk Pemilihan <i>Degree</i> | 27 |
| Tabel 4.3 Nilai AIC setelah Penentuan <i>Span</i> | 28 |
| Tabel 4.4 <i>P-value</i> Variabel Pada Anova Model Keseluruhan..... | 31 |
| Tabel 4.5 Nilai AIC Model Sebelum dan Sesudah Pemulusan..... | 32 |
| Tabel 4.6 Hasil Estimasi Parameter Model..... | 34 |

BAB 1. PENDAHULUAN

1.1. Latar Belakang

Regresi merupakan salah satu metode analisis data yang digunakan untuk melihat pengaruh antara dua model atau lebih. Hubungan antar variabel ini akan dibentuk menjadi suatu model matematika yang bersifat fungsional, atau yang sering dikenal dengan analisis regresi. Analisis regresi, yang juga disebut model statistika (*statistical models*), terdiri atas 2 variabel, yaitu variabel respon (*response variable*) dan variabel penduga (*explanatory variable*). Tujuan adanya pemodelan ini adalah untuk mengetahui adanya hubungan antara beberapa variabel dan melakukan prediksi. Adapun bentuk model yang dihasilkan dalam analisis regresi ini yaitu $y_i = \mu(t_i) + \varepsilon_i$, dengan $i = 1, 2, \dots, n$, y_i variabel respon, dan (t_i, y_i) adalah elemen t . Analisis regresi sendiri dibedakan menjadi tiga, yaitu regresi parametrik, regresi nonparametrik dan regresi semiparametrik.

Menurut Eubank (1999), regresi parametrik merupakan regresi yang mengandaikan bahwa bentuk μ sudah diketahui kecuali untuk parameter dalam jumlah banyak parameter yang tidak diketahui secara berhingga. Bentuk μ untuk regresi parametrik yaitu $X\beta$ dengan β adalah vektor parameter yang merupakan komponen parametrik dari X . β dari kombinasi linier dari X . Bentuk ini dapat diaplikasikan pada data yang memiliki pola tertentu.

Model regresi nonparametrik merupakan model regresi yang lebih fleksibel jika dibandingkan dengan model regresi parametrik. Pada regresi ini, penduga β digantikan oleh elemen s . Eubank (1999) mengasumsikan bahwa s terbentuk dari suatu fungsi tertentu, misalnya fungsi turunan atau fungsi turunan kedua dari kuadrat. Regresi ini mengabaikan asumsi-asumsi yang melandasi penggunaan regresi parametrik, terutama yang berkaitan dengan distribusi normal.

Dari kedua analisis regresi di atas, model regresi tersebut dapat digabungkan menjadi model regresi semiparametrik. Model regresi semiparametrik merupakan

gabungan dari 2 regresi, yaitu parametrik dan semiparametrik. Dalam model ini juga memiliki kedua komponen parameter, yaitu β dan s . Regresi semiparametrik digunakan jika pola hubungan antara variabel independen terhadap variabel dependen ada yang polanya diketahui dan ada pula yang pola datanya tidak diketahui. Salah satu bentuk model yang menggunakan regresi semiparametrik adalah model aditif. Model aditif ini merupakan perluasan dari model linier. Model aditif merupakan hasil pengembangan dari model linier, yaitu mengganti prediktor linier $\sum_j \beta_j X_j$ menjadi prediktor aditif, $\sum_j f_j(X_j)$. Model ini juga dapat diklasifikasikan menjadi model semiparametrik.

Hastie dan Tibsirani (1986) mengembangkan model aditif menjadi model yang lebih baru yang dikenal dengan model aditif tergeneralisir atau sering disebut dengan *Generalized Additive Models (GAM)*. Model ini merupakan perluasan dari model aditif dan model linier tergeneralisir yang mengasumsikan distribusi dari variabel respon tidak harus normal dan fungsi pada parameter yang dihasilkan tidak harus linier.

Model aditif tergeneralisir dapat diaplikasikan dalam berbagai jenis data. Ini karena model aditif tergeneralisir lebih umum dibandingkan dengan model linier dan model linier tergeneralisir. Hastie (1990) menyebutkan bahwa metode penghalusan dan algoritma yang digunakan dalam model aditif tergeneralisir terdiri dari pemulus diagram pencar (*scatterplot smoother*), algoritma *backfitting* dan algoritma *local scoring* (dalam Handayani, 2006). Algoritma-algoritma ini digunakan sebagai dasar untuk mempartisi atau membagi kelas-kelas dari pemulus.

Pemulus dibutuhkan dalam memodelkan suatu data dikarenakan kurva yang dihasilkan memiliki nilai estimasi yang besar, sehingga menghasilkan kurva yang kurang halus ataupun terlalu halus. Hastie dan Tibshirani (1989) menggunakan pemulus *spline* untuk mengestimasi parameter model aditif nonparametrik. Pada tahun 2006, Handayani mengaplikasikan model aditif tergeneralisir semiparametrik tersebut pada data agroklimatologi dengan menggunakan pemulus *spline*. Kemudian dengan metode dan pemulus yang sama, Laome (2009) memodelkan data longitudinal pada kasus kadar CD4 penderita HIV.

Selain *spline*, pemulus yang dapat digunakan lainnya yakni *loess*. Cleveland (1979) memperkenalkan pemulus ini yang merupakan singkatan dari *local regression* dan perluasan dari *lowess*. Pada tahun 2000, Jacoby menggunakan pemulus *loess* untuk mengetahui tingkat keikutsertaan siswa SMA dalam pemilihan presiden di Amerika Serikat yang juga menerangkan tentang penggunaan model aditif tergeneralisir dengan pemulus *loess*. Lalu pemulus ini digunakan oleh Safi'i (2010) untuk menganalisis data univariat tentang obligasi.

Salah satu keuntungan jika kita menggunakan *loess* yaitu dapat memuluskan dua variabel secara bersamaan atau tidak terpisah, sedangkan jika di *spline* itu hanya dapat digunakan 1 variabel yang dimuluskan. Maka dari itu, pada penelitian ini akan membahas tentang penggunaan pemulus *loess* pada model aditif tergeneralisir semiparametrik dengan menggunakan data tentang curah hujan di Indonesia sebagai variabel respon sedangkan untuk variabel prediktornya yaitu suhu, kelembaban udara dan tekanan udara.

1.2. Rumusan Masalah

Permasalahan yang akan dibahas pada penelitian ini dapat dirumuskan adalah bagaimana bentuk estimasi parameter dan pemulus *loess* pada model aditif tergeneralisir semiparametrik dan juga bagaimana aplikasinya pada data curah hujan.

1.3. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mendapatkan model terbaik dari data curah hujan dan mengetahui pengaruh tekanan udara, suhu udara dan kelembaban udara terhadap curah hujan dengan menggunakan GAM semiparametrik dengan pemulus *loess*.

1.4. Manfaat Penelitian

Manfaat yang dapat diambil oleh pembaca dari penelitian skripsi ini yaitu pembaca dapat menambah wawasan tentang model statistika yang menggunakan

pendekatan pemulus dan mengetahui bentuk GAM semiparametrik dengan menggunakan pemulus *loess* pada data jumlah curah hujan pada tahun 2009.



BAB 2. TINJAUAN PUSTAKA

2.1 Model Linier

Menurut Ratih (dalam Handayani, 2006), model linier memegang peranan penting dalam percobaan yang melibatkan lebih dari satu variabel yaitu melakukan analisis hubungan antarvariabel. Model linier paling sederhana yang sudah dikenal yaitu model regresi linier, yang hanya melibatkan satu variabel respon (Y) dan satu variabel prediktor (X).

Model linier memiliki bentuk umum :

$$Y = \sum_{j=0}^p \beta_j X_{ij} + \varepsilon \quad (2.1)$$

dengan ε merupakan matriks galat yang diasumsikan merupakan peubah acak yang memenuhi distribusi tertentu. Chambers & Hastie (1997) mengatakan bahwa asumsi-asumsi yang dipenuhi oleh model linier:

- ε adalah variabel independen dan terdistribusi identik;
- ε memiliki $\mu = 0$ dan varians terdefinisi;
- ε terdistribusi normal.

Peubah X adalah matriks peubah tetap yang tidak bersifat acak dan β adalah parameter yang menentukan koefisien dari peubah tetap tadi yang didapatkan dengan melakukan menggunakan metode kuadrat terkecil.

2.2 Model Linier Tergeneralisir

Model linier tergeneralisir atau yang sering dikenal *Generalized Linear Models* (GLM) merupakan pengembangan dari model linier yang mengandung komponen acak, komponen sistematik dan fungsi penghubung. Menurut Tirta (2009), GLM menggunakan asumsi bahwa respon memiliki distribusi keluarga eksponensial. Asumsi model linier tergeneralisir mengandung 3 hal, antara lain:

- komponen tetap yang disebut prediktor linier $\eta_i = \sum_{j=0}^p \beta_j X_{ij}$;
- respon y_i berdistribusi saling bebas dalam keluarga eksponensial;
- hubungan antara mean dan prediktor linier ditunjukkan dengan fungsi $g(.)$

yang disebut fungsi hubungan (*link function*) sedemikian hingga $g(\cdot) = \eta_i$.

Jika variabel respon berdistribusi normal, maka (\cdot) adalah identitas. Sehingga dari ketiga hal di atas komponen-komponen penting dalam model linier tergeneralisir yaitu:

1. adanya prediktor linier,
2. adanya distribusi keluarga eksponensial,
3. adanya hubungan fungsi yang differensiabel.

(Tirta, 2009).

2.3 Model Aditif

Perkembangan lain dari model statistika tidak mewajibkan adanya kombinasi linier, tetapi mengadopsi bentuk yang lebih luas yaitu polinomial atau bentuk aditif (Tirta, 2009). Hastie dan Tibshirani (1990) mengatakan bahwa model aditif (*Additive Model*) merupakan hasil generalisir atau pengembangan dari model linier. Sama halnya dengan model linier, model aditif juga diasumsikan berdistribusi normal. Perbedaan model aditif dan model linier terletak pada fungsi yang ada pada masing-masing model. Fungsi pada model linier mengharuskan fungsi yang digunakan linier, sedangkan pada model aditif tidak harus linier. Prediktor pada model aditif disebut prediktor aditif. Model aditif dapat dituliskan sebagai berikut:

$$Y = \beta_0 + \sum_{j=1}^p f_j(X_j) + \varepsilon \quad (2.2)$$

dengan Y adalah variabel respon; X_1, X_2, \dots, X_p adalah variabel prediktor; f_1, f_2, \dots, f_p adalah fungsi dari prediktor; dan ε adalah galat. Menurut Basset & Bishop (2005), model aditif disebut sebagai model nonparametrik jika koefisien linier $\beta_1, \beta_2, \dots, \beta_p$ pada persamaan (2.1) digantikan oleh fungsi *smooth* s_1, s_2, \dots, s_p . Sehingga model aditif juga dapat dituliskan dalam bentuk persamaan sebagai berikut:

$$\eta(X) = s_0 + \sum_{j=1}^p s_j(X_j) \quad (2.3)$$

dengan $s_j(X_j)$, untuk $j = 1, 2, 3, \dots, p$ sebagai fungsi *smooth*. Fungsi *smooth* s_j memenuhi kondisi standar yaitu $E[s_j(X_j)] = 0$. Model linier dapat digunakan untuk menganalisis berbagai jenis data, begitu juga dengan model aditif. Tetap

terdapat beberapa permasalahan, sehingga data tidak dapat dianalisis menggunakan model aditif. Salah satu permasalahan tersebut adalah distribusi normal digunakan untuk data diskrit sehingga dibutuhkan model dengan distribusi yang lebih luas dan fleksibel yakni model aditif tergeneralisir.

2.4 Model Aditif Tergeneralisir

Hastie & Tibshirani (1986) menjelaskan model aditif tergeneralisir merupakan pengembangan dari model linier tergeneralisir dengan mengganti fungsi linier $\sum_{j=1}^p \beta_j X_{ij}$ dengan fungsi aditif $\sum_{j=1}^p s_j(X_j)$ sehingga prediktor dalam model ini disebut prediktor aditif. Model ini menggeneralisir model aditif ke dalam bentuk distribusi keluarga eksponensial selain distribusi normal.

Model aditif tergeneralisir memuat komponen acak, komponen tetap yakni prediktor aditif dan fungsi *link* yang menghubungkan komponen acak dengan prediktor aditif tersebut. Variabel respon Y sebagai komponen acak diasumsikan memiliki kepadatan keluarga eksponensial sebagai berikut:

$$f_Y(y; \theta; \phi) = \exp \left\{ \left(\frac{y(\theta) - b(\theta)}{a(\phi)} \right) + c(y, \phi) \right\} \quad (2.4)$$

dengan θ sebagai parameter natural dan ϕ sebagai parameter dispersinya. Hubungan antara mean μ dari variabel respon dengan komponen aditif $\eta(X) = s_0 + \sum_{j=1}^p s_j(X_j)$ didefinisikan oleh fungsi *link* $g(\mu) = \theta$ dan secara umum fungsi *link* yang digunakan adalah fungsi kanonik dengan $\eta = \theta$.

Bentuk umum dari model aditif tergeneralisir dapat diformulasikan sebagai berikut:

$$g(\mu) = s_0 + \sum_{j=1}^p s_j(X_j) + \varepsilon \quad (2.5)$$

dengan fungsi $s(\cdot)$ merupakan fungsi pemulus dan juga diperlukan algoritma dalam memodelkan GAM ini. Variabel respon yang digunakan dalam persamaan (2.5) diasumsikan berdistribusi keluarga eksponensial.

Model aditif tergeneralisir adalah hasil perluasan dari model linier tergeneralisir. Perluasan ini dilakukan dengan cara yang sama dengan perluasan model linier menjadi model aditif, yaitu dengan menggantikan prediktor linier $\sum_j \beta_j X_j$ dengan prediktor aditif, $\sum_j f_j(X_j)$. Fungsi $s(\cdot)$ merupakan model yang didapatkan dari pemulus dengan mengestimasi secara berulang yang disesuaikan

dengan variabel X . Estimasi ini dapat disebut prosedur *local scoring* karena nilai *Fisher Scoring* dihitung menggunakan estimasi lokal dari *score* (Hastie & Tibshirani, 1986).

Handayani (2006) menjelaskan bahwa dalam model linier tergeneralisir, estimasi *maximum likelihood* dari nilai parameter β didefinisikan oleh persamaan *score* (*score equation*) sebagai berikut.

$$\sum_{i=1}^n x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_j} \right) V_i^{-1} (y_i - \mu_i) = 0 \quad (2.6)$$

dengan $V_i = \text{var}(Y_i)$ dan x_{i0} diasumsikan 1. Prosedur *Fisher Scoring* merupakan metode umum yang digunakan untuk menyelesaikan persamaan tersebut. Prosedur yang ekuivalen dengan persoalan ini disebut dengan regresi variabel dependen. Rumusan variabel dependen adalah

$$z = \hat{\eta} + (y - \hat{\mu}) \left(\frac{\partial \eta}{\partial \mu} \right) \quad (2.7)$$

dengan $\hat{\eta}$ prediktor linier dan bobot (*weight*) W didefinisikan sebagai:

$$(W)^{-1} = \left(\frac{\partial \eta}{\partial \mu} \right)^2 V \quad (2.8)$$

untuk V varian dari Y pada $\mu = \hat{\mu}$. Algoritma dihasilkan dengan meregresi z pada x dengan bobot W guna mendapatkan kembali nilai estimasi $\hat{\beta}$. Dengan menggunakan $\hat{\beta}$, didapatkan $\hat{\mu}$ dan $\hat{\eta}$ yang baru sehingga nilai z juga dihitung dan didapatkan nilai baru, begitu seterusnya hingga devian berubah

$$\text{dev}(y, \hat{\mu}) = 2[l(y) - l(\hat{\mu})] \quad (2.9)$$

menjadi cukup kecil (Hastie dan Tibshirani, 1990). Penduga untuk $\hat{\beta}$ setelah di estimasi menggunakan maksimum *likelihood* akan identik dengan menggunakan metode kuadrat terkecil terbobot, yaitu

$$b^{(m)} = b^{(m-1)} + \left(\mathbf{X}^T \left(\frac{\partial \mu}{\partial \eta} \right) \left(\frac{1}{\text{var}(Y)} \right) \left(\frac{\partial \mu}{\partial \eta} \right)^T \mathbf{X} \right)^{-1} \left(\mathbf{X}^T \left(\frac{1}{\text{var}(Y)} \right) \left(\frac{\partial \mu}{\partial \eta} \right) (Y - \mu) \right) \quad (2.10)$$

Jika fungsi pemulus yang digunakan adalah parametrik, maka dalam penentuan estimasinya menggunakan *weighted least square*, sehingga GAM ekuivalen dengan GLM. Apabila fungsi pemulus pada prediktor linier adalah nonparametrik, maka pengepasan model regresi aditifnya menggunakan algoritma *backfitting*. Siklus algoritma *backfitting* melalui variabel dan menghitung masing-masing fungsi koordinat dengan menggunakan *smoothing partial residual*

sehingga ketika fungsi pemulus yang dimasukkan dalam model, maka algoritma *backfitting* memberikan solusi bentuk tertutup dari estimasi parameter.

Menurut Hastie & Tibshirani (1986), algoritma *local scoring* dapat dituliskan seperti berikut ini.

(i) Inisialisasi

$$s_i = g(E(Y)), s_1^0(\cdot) = s_2^0(\cdot) = \dots = s_p^0(\cdot) = 0, m = 0$$

(ii) Iterasi

$$m = m + 1$$

Persamaan variabel dependen biasa; variabel prediktor; dan *mean* adalah sebagai berikut:

$$Z_i = \eta^{m-1} + (Y - \mu^{m-1}) \left(\frac{\partial \eta}{\partial \mu^{m-1}} \right)$$

$$\eta^{m-1} = s_0 + \sum_{j=1}^p s_j^{m-1}(X_j)$$

$$\mu^{m-1} = g^{-1}(\eta^{m-1})$$

persamaan bobot:

$$w_i = \left(\frac{\partial \eta}{\partial \mu^{m-1}} \right)^2 V_i^{-1}$$

Masukkan model aditif pada Z menggunakan algoritma *backfitting* dengan bobot w guna mendapatkan fungsi estimasi $s_j^m(\cdot)$.

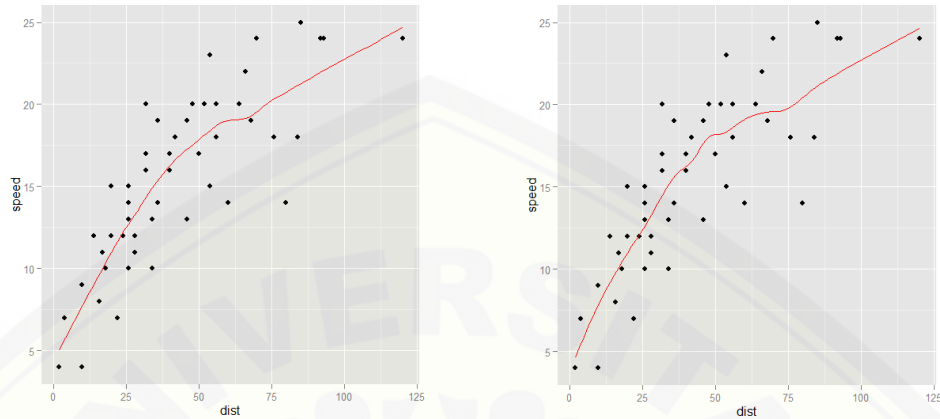
(iii) Ulangi langkah (ii) hingga memenuhi kriteria kekonvergenan sebagai berikut:

$$\frac{\sum_{j=1}^n \sum_{i=1}^k w(x_i) (s_i^{m-1}(x_j) - s_i^m(x_j))^2}{\sum_{j=1}^n \sum_{i=1}^k w(x_i) (s_i^{m-1}(x_j))^2} \leq \varepsilon$$

2.5 Local Regression (Loess)

Salah satu aspek yang unik dari model aditif tergeneralisir adalah fungsi nonparametrik yang diestimasi oleh fungsi pemulus yakni pemulus diagram pencar. Ada beberapa pemulus yang digunakan dalam menentukan fungsi nonparametrik. Hastie & Tibshirani (1990) menggunakan pemulus *spline* untuk menghasilkan model yang baik. Namun, pada tahun sebelumnya, Cleveland (1979) memperkenalkan pemulus lain yaitu loess.

Loess, *local regression* merupakan generalisir dari *lowess (local weighted regression)*. Proses pemulusan ini dianggap lokal karena setiap nilai yang dimuluskan ditentukan oleh titik data yang berdekatan yang didefinisikan dalam rentang tersebut. Model regresi yang digunakan dalam regresi ini dapat berupa



(a) *Degree 1*

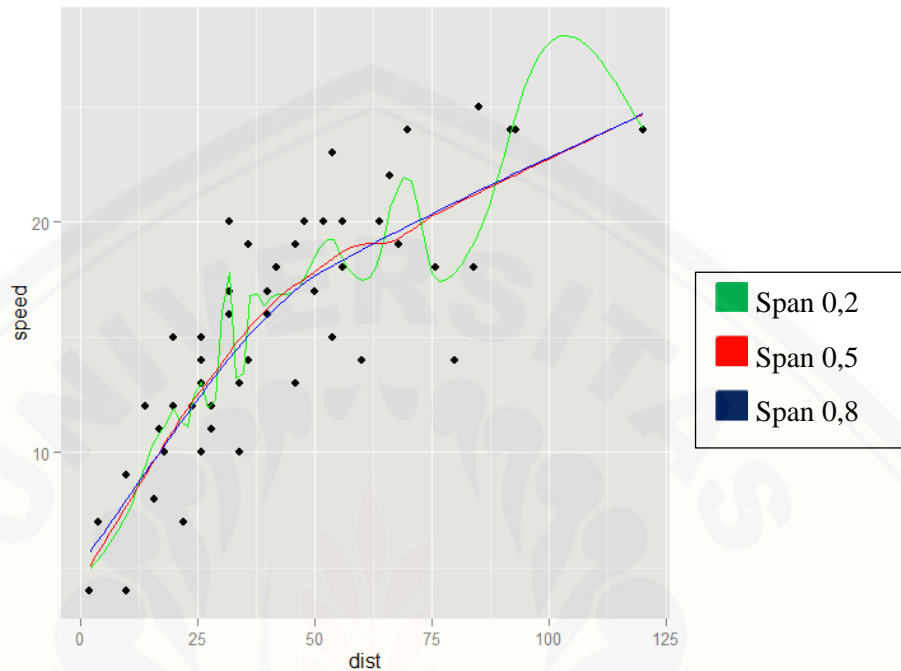
(b) *Degree 2*

Gambar 2.1 Ilustrasi Pemulus Loess dengan *Degree* yang Berbeda. Gambar (a) menggunakan *degree 1* dan Gambar (b) menggunakan *degree 2*.

Parameter yang digunakan untuk menghasilkan pemulus yang baik bagi model adalah *span* (f) dan *degree* (d). *Degree* merupakan derajat polinomial yang akan digunakan, yakni 1 atau 2. Jika derajat polinomial yang digunakan adalah 1, maka model yang digunakan pada setiap jendelanya adalah linier. Ketika derajat polinomial adalah 2, maka persamaan kuadrat yang digunakan. Gambar 2.1 merupakan ilustrasi pemulus loess dengan derajat berbeda. Gambar (a) merupakan pemulus loess dengan menggunakan derajat 1, sedangkan gambar (b) pemulus loess yang menggunakan derajat 2. Dari gambar tersebut, dapat dilihat terdapat perbedaan pada bagian tengah gambar (a) dan gambar (b). Pada gambar (b) bagian tengahnya lebih kasar daripada gambar (a), tetapi garis regresi yang dihasilkan lebih pas dengan data adalah gambar (b). Ini dapat dilihat dari pola garis patahan yang terbentuk mengikuti sebaran data yang diplotkan.

Selain derajat polinomial, parameter lain yang digunakan adalah rentang (*span*). Parameter rentang (*span*) dibutuhkan untuk menentukan lebar jendela. Ilustrasi dapat dilihat pada Gambar 2.2. Jacoby (2000) menjelaskan *span* ini

berkisar antara nilai 0 sampai 1, sedangkan Cleveland (1979) menyebutkan bahwa *span* terbaik yang dapat digunakan adalah dari 0.2 sampai 0.8. Berikut adalah gambar ilustrasi penggunaan *span* pada pemulus loess.



Gambar 2.2 Ilustrasi Pemulus Loess Dengan *Span* Yang Berbeda

Pada Gambar 2.2 terdapat 3 garis dengan warna yang berbeda, yaitu hijau, biru dan kuning. Garis regresi yang paling mendekati dengan daerah data adalah garis dengan warna merah dan biru. Garis hijau tidak dapat digunakan karena lekukan yang dihasilkan terlalu banyak dan tidak mengikuti sebaran data yang ada. Ini dapat dilihat dari lekukan terakhir yang ada pada garis tersebut dimana terdapat puncak yang disekitar puncak tersebut tidak terdapat sebaran data. Garis merah merupakan garis yang paling pas bila digunakan untuk mewakili data karena lekukan yang dihasilkan sudah pas dengan plot sebaran data.

Adapun prosedur-prosedur yang dibutuhkan dalam pemulus loess yang didefinisikan oleh Cleveland (1979) sebagai berikut.

- Untuk setiap i , menghitung h_i sebagai jarak dari x_i ke r terdekat persekitaran x_i . Oleh karena itu, h_i adalah bilangan r diantara $|x_i - x_j|$ terkecil dengan

$i = j = 1, 2, \dots, n$. Untuk $k = 1, 2, \dots, n$, pembobot $w_k(x_i)$ dapat ditentukan dengan persamaan :

$$w_k(x_i) = W\left(\frac{x_k - x_i}{h_i}\right) \quad (2.11)$$

dengan W adalah fungsi pembobot *tricube*,

$$W(x) = \begin{cases} (1 - x^3)^3, & |x| < 1 \\ 0, & |x| \geq 1 \end{cases} \quad (2.12)$$

- b. Untuk setiap i , dapat diestimasi parameter $\hat{\beta}_j(x_j), j = 1, 2, \dots, d$, yaitu parameter di regresi polinomial dari y_k ke x_k dengan derajat d . Kemudian melakukan analisis regresi polinomial dengan pembobot $w_k(x_i)$ untuk (x_k, y_k) .

Parameter $\hat{\beta}_j(x_j)$ didapatkan dengan meminimalisasi errornya dari persamaan:

$$\sum_{k=1}^n w_k(x_i)(y_k - \beta_0(x_i) - \beta_1(x_i)x_k - \dots - \beta_d(x_i)x_k^d)^2 \quad (2.13)$$

Jika $d = 1$, maka polinomial akan berbentuk linier sedangkan jika $d = 2$, maka polinomial akan berbentuk kuadratik.

- c. Nilai parameter $\hat{\beta}_j(x_j)$ telah didapatkan di langkah 2. Langkah selanjutnya adalah menentukan nilai pembobot *robust*, δ_k :

$$\delta_k = B\left(\frac{e_i}{6s}\right) \quad (2.14)$$

- d. Nilai aksiran \hat{y}_i dihitung kembali (seperti pada langkah 2) menggunakan *weighted least squares* dengan pembobot $\delta_k w_k(x_i)$.

(Safi'i, 2010).

2.6 Akaike's Information Criterion(AIC)

Pemeriksaan model terbaik dapat dilakukan dengan banyak cara. Jika menggunakan bantuan *software*, maka dapat dilakukan dengan menggunakan kriteria informasi *Akaike* atau yang sering disebut *Akaike's Information Criterion* (AIC). AIC adalah salah satu pertimbangan dalam membentuk model terbaik menghitung pertimbangan antara *likelihood* dengan banyaknya variabel dalam model. Besarnya AIC dihitung melalui rumus

$$AIC = -2l(\hat{\theta}) + 2q \quad (2.15)$$

dengan $l(\hat{\theta})$ adalah nilai *likelihood* dari model yang dihadapi dan q adalah banyaknya parameter dalam model. Secara umum, semakin kecil AIC, model yang dipakai semakin cocok (Tirta, 2009).

2.7 Model Semiparametrik

Model semiparametrik merupakan model gabungan antara model yang menggunakan pendekatan parametrik dengan model yang menggunakan pendekatan nonparametrik. Persamaan model parametrik yaitu

$$Y = \beta_0 + \sum_{i=1}^p \beta_j X_j \quad (2.16)$$

dengan β merupakan kombinasi linier dari variabel X. Model parametrik ini didasarkan pada asumsi-asumsi yang mengikuti suatu bentuk distribusi tertentu. Sebaliknya, jika koefisien β tidak didasarkan pada suatu bentuk distribusi tertentu dengan hubungan fungsi yang juga tidak diketahui, linier atau tidak, maka model ini disebut sebagai model nonparametrik.

Persamaan model nonparametrik dapat dituliskan sebagai berikut :

$$\eta_i = \gamma(X) + \varepsilon \quad (2.17)$$

dengan $\gamma(X)$ merupakan fungsi tidak tetap atau berubah-ubah. Dari persamaan (2.10) dan (2.11) sehingga persamaan model semiparametrik dapat dituliskan sebagai berikut (Hastie dan Tibsirani, 1990) :

$$\eta(X, t) = \alpha + \sum_{j=1}^p X_j \beta_j + \gamma(t) \quad (2.18)$$

dimana t adalah vektor kovariat skalar dan γ adalah fungsi yang berubah-ubah.

2.8 Curah Hujan

Curah hujan adalah banyaknya air yang jatuh ke permukaan bumi. Derajat curah hujan dinyatakan dengan jumlah curah hujan dalam suatu satuan waktu. Biasanya satuan yang digunakan adalah mm/jam. Dalam meteorologi butiran hujan dengan diameter lebih dari 0,5 mm disebut hujan dan diameter antara 0,5 – 0,1 mm disebut gerimis. Semakin besar ukuran butiran hujan maka semakin besar pula kecepatan jatuhnya. Ketelitian alat ukur curah hujan adalah 1/10 mm. Pembacaan dilakukan satu 22 kali dalam sehari dan dicatat sebagai curah hujan hari terdahulu/kemarin (Suyono,1985:8). Faktor-faktor yang mempengaruhi curah hujan antara lain:

a. Kelembaban Udara

Kelembaban adalah perbandingan antara massa uap dan satuan volume dengan massa uap yang jenuh dalam satuan volume itu pada suhu yang sama. Secara umum kelembaban menyatakan banyaknya kadar air yang ada di udara. Banyaknya uap yang bergerak di dalam atmosfer berpengaruh terhadap besarnya hujan, lamanya hujan dan intensitas cahaya. Kelembaban tertinggi umumnya terjadi pada musim penghujan dan paling rendah pada musim kemarau. Variasi kelembaban bergantung pada suhu udara, jika di pagi hari suhu rendah maka kelembaban akan lebih tinggi jika dibanding siang hari saat suhu tinggi.

Umumnya semakin tinggi suatu daerah permukaan laut maka kelembaban udaranya semakin tinggi. Makin tinggi kelembaban udara akan dapat menyebabkan bertambah banyak uap air yang dapat diserap awan. Uap air itu akan menghasilkan tekanan yang dinyatakan dengan satuan tinggi air raksa ($1 \text{ mmHg} = 1,33 \text{ milibar}$). Tekanan yang diberikan oleh uap air tersebut disebut dengan tekanan uap air (Suyono, 1985:12).

b. Tekanan Udara

Tekanan udara merupakan tenaga yang bekerja untuk menggerakkan massa udara dalam setiap tahun dalam setiap satuan uas tertentu. Diukur dengan menggunakan barometer. Satuan tekanan udara adalah milibar. Garis yang menghubungkan tempat-tempat yang sama tekanannya disebut sebagai isobar.

Tekanan udara dibatasi oleh ruang dan waktu. Artinya pada tempat dan waktu berbeda, besarnya juga berbeda. Semakin tinggi suatu tempat, maka tekanan udaranya semakin turun, sedangkan tekanan udara pada daerah yang mempunyai ketinggian rata-rata sama maka tekanan udara dipengaruhi oleh suhu udara. Daerah yang suhu udaranya tinggi akan bertekanan rendah dan daerah yang suhu udaranya rendah tekanannya tinggi (Soewarno, 2000:10).

c. Suhu Udara

Suhu udara adalah keadaan panas atau dinginnya udara. Suhu juga disebut temperatur yang diukur dengan alat termometer. Beberapa faktor yang mempengaruhi suhu udara diantaranya; tinggi tempat, daratan/lautan, radiasi

matahari, indeks datang matahari dan angin. Pengukuran biasa dinyatakan dalam skala Celcius (C), Reamur (R), Fahrenheit (F). Suhu udara tertinggi di permukaan bumi adalah di daerah tropis (sekitar equator) dan makin ke kutub makin dingin (Soewarno, 2000:12).

d. Kecepatan Angin

Angin adalah udara yang bergerak akibat adanya perbedaan tekanan udara dengan arah aliran angin dari tempat yang memiliki suhu/temperatur rendah ke wilayah bersuhu tinggi. Angin memiliki hubungan erat dengan sinar matahari karena daerah yang terkena banyak paparan sinar matahari akan memiliki suhu yang lebih tinggi serta tekanan udara yang lebih rendah dari daerah lain di sekitarnya sehingga menyebabkan terjadinya aliran udara. Angin juga dapat disebabkan oleh pergerakan benda sehingga mendorong udara di sekitarnya untuk bergerak ke tempat lain (Soewarno, 2000:15).

BAB 3. METODE PENELITIAN

Metode penelitian merupakan kerangka acuan yang akan digunakan dalam menganalisis suatu data serta menerangkan tentang langkah-langkah kerja yang harus dilakukan agar menghasilkan hasil analisis yang diinginkan.

3.1. Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data riil yang diambil dari laman <http://www.bps.go.id/>. Data ini merupakan data tentang curah hujan yang ada di setiap daerah di Indonesia sebanyak 33 data yang diambil [er harinya selama 1 tahun. Curah hujan dipengaruhi oleh beberapa variabel, antarlain kelembaban udara, suhu udara dan tekanan udara. Adapun variabel yang digunakan dalam analisis riil ini yaitu:

- a. data curah hujan yang menjadi variabel respon (Y),
- b. data kelembaban udara (X_1), suhu udara (X_2), dan tekanan udara (X_3) menjadi variabel prediktor.

3.2. Analisis Data

Metode analisis yang digunakan dalam penelitian ini meliputi program Rcommander sebagai *software* komputasinya menggunakan struktur fungsi *gam* dan analisis data riil.

3.2.1. Struktur Fungsi GAM pada R

Program yang digunakan dalam bidang statistika cukup banyak. Salah satunya yaitu Rcommander 3.1.3. *Software* R merupakan salah satu program yang digunakan dalam menganalisis data dalam statistika.

Adapun struktur fungsi dari model aditif tergeneralisir yang digunakan dalam paket R adalah sebagai berikut:

```
gam(formula = y ~ lo(x, span=..., degree=...), family =
    gaussian, data, weights, subset, na.action, start,
    etastart, mustart, control = gam.control (...), model =
    FALSE, method, x = FALSE, y = TRUE, ...)
```

Keterangan:

- formula : suatu rumusan tentang model regresi, bentuk response ~ predictors. pemulus nonparametrik diindikasikan oleh s untuk pemulus *splines* dan lo untuk *loess*.
- family : deskripsi dari distribusi error dan fungsi link yang digunakan dalam model. Rumusan ini dapat berupa karakter string yang memberikan nama pada fungsi *family*.
- data : suatu datat frame yang memuat variabel-variabel dalam model, variabel-variabel itu diambil dari environment (formula), khususnya *environment* dimana gam dipanggil.
- weight : suatu vektor dari bobot (weights) yang digunakan dalam proses pencocokan.
- subset : suatu optimalitas vektor yang secara khusus merupakan bagian dari pengamatan yang digunakan dalam proses pencocokan.
- na.action : sebuah fungsi yang mengindikasikan tentang apa yang terjadi ketika data memuat NA. Kesalahan itu dikumpulkan oleh *na.action*, disusun dalam *options*, dan jika *na.fail* maka tidak dikumpulkan.
- start : nilai awal parameter dalam prediktor aditif.
- etastart : nilai awal untuk prediktor aditif.
- mustart : nilai awal untuk vektor dari mean.
- offset : rumusan ini dapat digunakan untuk menetapkan suatu a priori,

- sebagai komponen yang diketahui untuk dimasukkan ke dalam prediktor aditif selama proses pencocokan.
- `control` : daftar dari parameter-parameter untuk mengontrol proses pencocokan.
- `model` : nilai logis yang mengindikasikan apakah model frame harus dimasukkan sebagai komponen awal.
- `method` : metode yang digunakan dalam pencocokan bagian parametrik dari model.
- `x, y` : untuk gam: nilai logis yang mengindikasikan apakah vektor respon dan model matriks yang digunakan dalam pencocokan, harus dikembalikan sebagai komponen dari nilai awal.

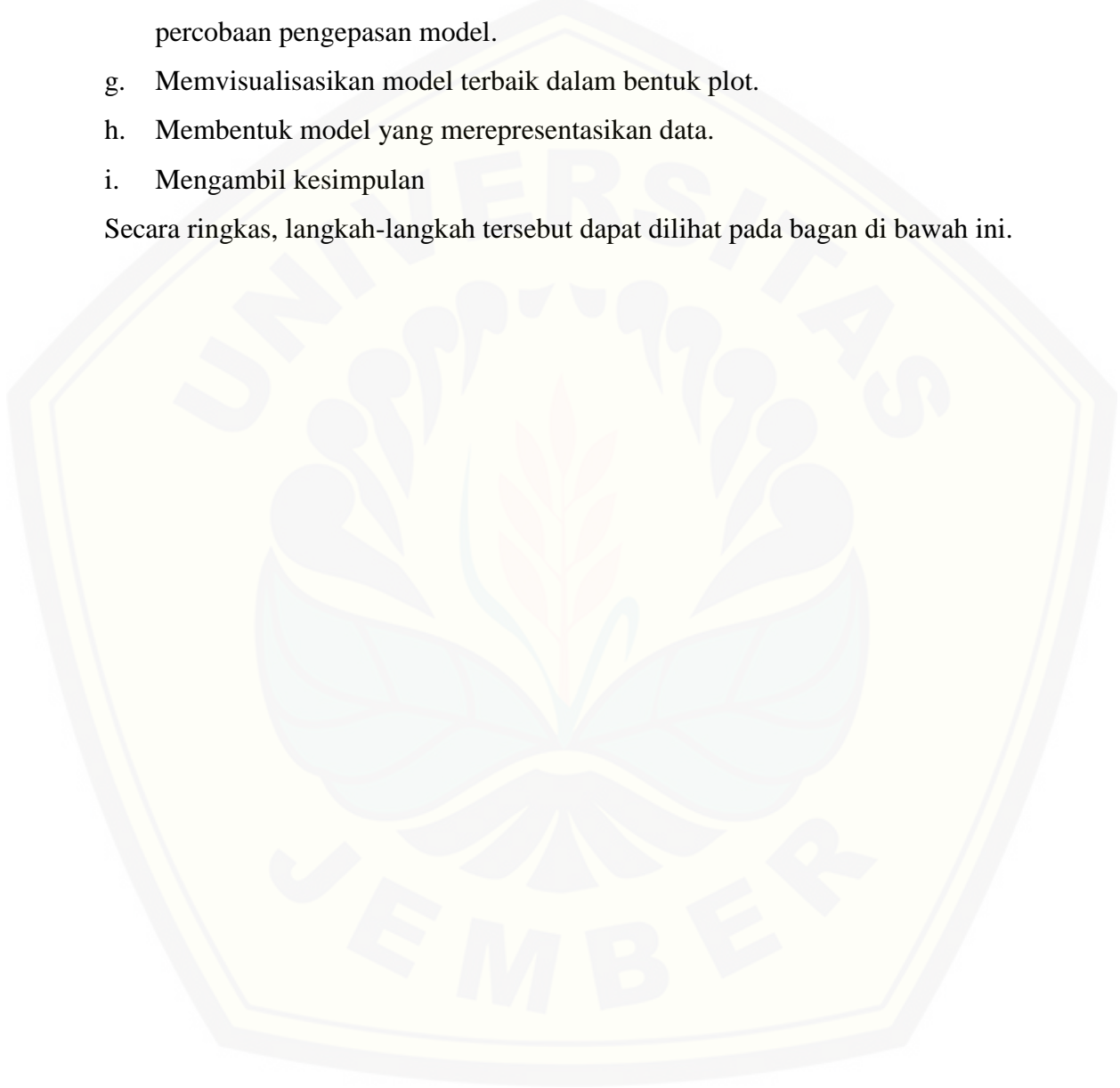
3.2.2. Metode Analisis Data

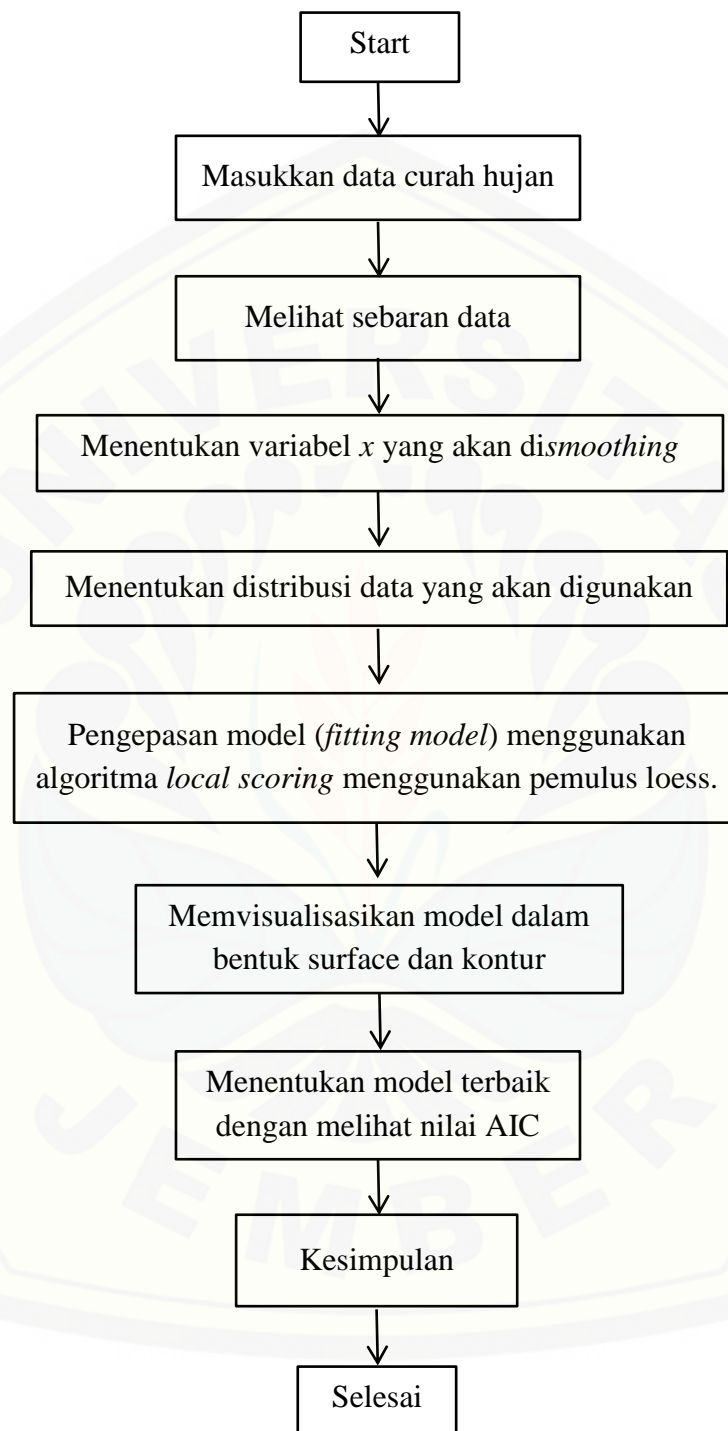
Data yang digunakan dalam penelitian ini merupakan data curah hujan udara di seluruh Indonesia selama tahun 2009. Ilustrasi data dapat dilihat pada Lampiran A. Adapun langkah-langkah dalam menganalisis data riil curah hujan tersebut adalah sebagai berikut.

- Memasukkan data yang akan dianalisis pada program R. Data yang akan dimasukkan pada program R adalah data curah hujan yang terdiri atas variabel-variabel yang telah dijelaskan sebelumnya.
- Melihat sebaran data yang digunakan untuk menentukan distribusi data yang akan digunakan.
- Melihat *scatterplot* data lalu menentukan variabel mana yang akan diberi pemulus loess.
- Menentukan distribusi data yang sesuai dengan data tersebut. Pada tahap ini, distribusi yang digunakan adalah keluarga eksponensial, diantaranya yaitu Gamma dan Normal untuk data kontinu. Pada tahap ini, GAM digunakan untuk menentukan fungsi penghubung dari distribusi.

- e. Pengepasan model (*fitting model*) menggunakan algoritma *local scoring* menggunakan pemulus loess.
- f. Menentukan model yang baik dengan melihat AIC yang paling kecil dari percobaan pengepasan model.
- g. Memvisualisasikan model terbaik dalam bentuk plot.
- h. Membentuk model yang merepresentasikan data.
- i. Mengambil kesimpulan

Secara ringkas, langkah-langkah tersebut dapat dilihat pada bagan di bawah ini.





Gambar 3.1 Langkah-langkah menganalisis data

BAB 5. PENUTUP

5.1. Kesimpulan

Berdasarkan hasil dan pembahasan, dapat diambil kesimpulan bahwa variabel prediktor kelembaban, suhu udara dan tekanan memberikan hubungan yang signifikan pada curah hujan di Indonesia. Semakin tinggi tingkat kelembaban dan suhu, semakin tinggi curah hujan yang dihasilkan. Sedangkan semakin rendah tekanan udara, curah hujan yang dihasilkan akan semakin tinggi.

Dalam penggunaan GAM dengan pemulus loess ini, didapatkan model terbaik setelah menentukan distribusi data berikut dengan fungsi penghubungnya, juga dengan pemilihan parameter pemulus, *degree* dan *span*. Jadi, distribusi yang digunakan juga dengan fungsi penghubungnya mempengaruhi baik tidaknya suatu model jika diaplikasikan pada data, selain parameter dari pemulus loess, *degree* dan *span*.

5.2. Saran

Hasil penelitian ini menggunakan pemulus loess untuk mengestimasi parameter nonparametriknya dengan mencari *degree* dan *span* yang pas untuk model yang diinginkan. Penggunaan pemulus ini dapat dieksplorasi lagi dengan menggunakan model statistika lainnya, misalnya GAMLSS. Selain menggunakan metode lain, pemulus yang digunakan juga bisa diubah sehingga dapat dibandingkan antara pemulus loess dan spline.

DAFTAR PUSTAKA

- Chambers, J. M., & Hastie, T. J. 1997. *Statistical Models in S*. London: Chapman & Hall.
- Cleveland, W. S. 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of The American Statistical Association*, **74**:829-836.
- Eubank, Randall L. 1999. *Nonparametric Regression And Spline Smoothing: Second Edition*. USA: Marcel Dekker.
- Handayani, E. W. 2006. *Pencocokan Model Aditif Tergeneralisir Semiparametrik dengan Menggunakan Estimator Spline*. Tidak dipublikasikan (Skripsi). Jember: Universitas Jember.
- Hastie, T. J., & Tibshirani, R. 1986. Generalized Additive Models. *Statistical Science*, **1**(3):297-318.
- Hastie, T. J., & Tibshirani R. 1990. *Generalized Additive Models*. London: Chapman & Hall.
- Jacoby, W. 2000. Loess: A Nonparametric, Graphic Tool for Depiciting Relationship Between Variabels. *Electoral Studies*, 577-613.
- Laome, L. 2009. Model Regresi Semiparametrik Spline Untuk Data Longitudinal Pada Kasus Kadar CD4 Penderita HIV. *Paradigma*, **13**(2) : 189-194.
- Ling, Xihong dan Carrol, Raymond J. 2001. Semiparametric Regression for Clustered Data Using Generalized Estimating Equation. *Journal of American Statistical Association*. **96**(455): 1045.
- Safi'i, W. 2010. *Analisa Penerapan Metode Robust Locally Weight Regression Smoothing Scatterplots Pada Obligasi*. <http://digilib.its.ac.id/public/ITS-Undergraduate-9292-Paper.pdf>. [31 Maret 2015]
- Soewarno. 2000. *Hidrologi Operasional*. Jilid kesatu. Bandung: Citra Aditya Bakti.
- Suyono, S. 1985. *Hidrologi*. Jakarta: PT Pradnya Paramita.

Tirta, I. M. 2009. *Analisis Regresi dengan R*. Jember: UPT Penerbitan Universitas Jember.



Lampiran A. Data Jumlah Curah Hujan di Indonesia Pada Tahun 2009

| No. | Provinsi | Curah Hujan (Y) | Kelembaban (X ₁) | Suhu (X ₂) | Tekanan Udara (X ₃) |
|-----|---------------------------|-----------------|------------------------------|------------------------|---------------------------------|
| 1 | Aceh | 1577,00 | 78,70 | 26,90 | 1008,80 |
| 2 | Sumatera Utara | 2184,00 | 86,0 | 27,10 | 1011,80 |
| 3 | Sumatera Barat | 4691,00 | 86,20 | 25,50 | 997,00 |
| 4 | Riau | 3390,00 | 76,20 | 27,70 | 995,50 |
| 5 | Kepulauan Riau | 2739,00 | 84,10 | 26,90 | 1010,50 |
| 6 | Jambi | 2299,00 | 83,80 | 26,90 | 1011,20 |
| 7 | Sumatera Selatan | 2389,00 | 82,30 | 27,40 | 1009,70 |
| 8 | Kepulauan Bangka Belitung | 1866,00 | 81,0 | 27,30 | 1009,60 |
| 9 | Bengkulu | 3850,00 | 84,10 | 26,50 | 1008,00 |
| 10 | Lampung | 1789,00 | 79,10 | 26,70 | 1011,00 |
| 11 | DKI Jakarta | 1973,00 | 74,40 | 28,30 | 1010,00 |
| 12 | Jawa Barat | 2098,00 | 78,60 | 23,40 | 922,90 |
| 13 | Banten | 1414,00 | 81,30 | 27,20 | 1008,70 |
| 14 | Jawa Tengah | 2807,00 | 74,90 | 27,90 | 1010,30 |
| 15 | DI Yogyakarta | 1584,00 | 79,50 | 26,10 | 994,60 |
| 16 | Jawa Timur | 1987,00 | 68,80 | 28,20 | 1010,60 |
| 17 | Bali | 1987,00 | 68,60 | 28,20 | 1009,50 |
| 18 | Nusa Tenggara Barat | 1440,00 | 77,60 | 27,40 | 1008,70 |
| 19 | Nusa Tenggara Timur | 1513,00 | 76,60 | 27,30 | 1010,70 |
| 20 | Kalimantan Barat | 2989,00 | 84,30 | 27,10 | 1009,10 |
| 21 | Kalimantan Tengah | 2771,00 | 81,80 | 26,90 | 1012,70 |
| 22 | Kalimantan Selatan | 2321,00 | 81,60 | 26,70 | 1011,70 |
| 23 | Kalimantan Timur | 2163,00 | 82,50 | 28,40 | 1007,60 |
| 24 | Sulawesi Utara | 2884,00 | 82,80 | 26,60 | 1010,00 |
| 25 | Gorontalo | 1245,00 | 78,60 | 27,30 | 1009,6 |
| 26 | Sulawesi Tengah | 563,00 | 74,90 | 27,60 | 1010,20 |
| 27 | Sulawesi Selatan | 2908,00 | 80,90 | 27,30 | 1011,30 |

| | | | | | |
|----|-------------------|---------|-------|-------|---------|
| 28 | Sulawesi Barat | 1713,00 | 78,40 | 27,50 | 1010,60 |
| 29 | Sulawesi Tenggara | 1783,00 | 81,50 | 27,70 | 1011,30 |
| 30 | Maluku | 1920,00 | 84,40 | 26,60 | 1010,40 |
| 31 | Maluku Utara | 2024,00 | 80,0 | 27,40 | 1008,20 |
| 32 | Papua | 2323,00 | 85,50 | 27,0 | 1008,20 |
| 33 | Papua Barat | 1907,00 | 81,30 | 27,10 | 1008,40 |



Lampiran B. Skrip Program Analisis Data

Eksplorasi Data

```
library(car) #digunakan untuk mengerjakan scatterplotmatriks
data(hujan)
scatterplotMatrix(~CurahHujan+Kelembaban+Suhu+Tekanan,
reg.line=lm, smooth=TRUE, spread=FALSE, diagonal = 'histogram',
data=hujan)
```

Analisis Data Menggunakan GAM

1. Penentuan Link Distribusi Data

```
gam.new1 <- gam(CurahHujan ~ Kelembaban + Suhu + Tekanan,
family=Gamma(link=identity), data=hujan)
gam.new2 <- gam(CurahHujan ~ Kelembaban + Suhu + Tekanan,
family=Gamma(link=log), data=hujan)
```

2. Penentuan Derajat Polinomial Pemulus Loess

```
gama <- gam(CurahHujan ~ Kelembaban + lo(Suhu, Tekanan, degree =
1), family = Gamma(link = log), data = hujan)
gamb <- gam(CurahHujan ~ Kelembaban + lo(Suhu, Tekanan, degree =
2), family = Gamma(link = log), data = hujan)
```

3. Penentuan *Span*

```
lk1 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.3) +
Tekanan, family=Gamma(link=log), data=hujan)
lk2 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.4) +
Tekanan, family=Gamma(link=log), data=hujan)
lk3 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.5) +
Tekanan, family=Gamma(link=log), data=hujan)
lk4 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.6) +
Tekanan, family=Gamma(link=log), data=hujan)
lk5 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.7) +
Tekanan, family=Gamma(link=log), data=hujan)
lk6 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.8) +
Tekanan, family=Gamma(link=log), data=hujan)
```

Plot Surface Model

```
library(asbio)
data(hujan)
hujan.df <- data.frame(x=hujan$Kelembaban, y=hujan$Suhu,
z=hujan$CurahHujan)
loess.surf(hujan.df$z, cbind(hujan.df$x, hujan.df$y), phi = 15,
theta = 2)
```

Plot Kontur Variabel

```
require(MASS)
data(hujan)
hujan
```

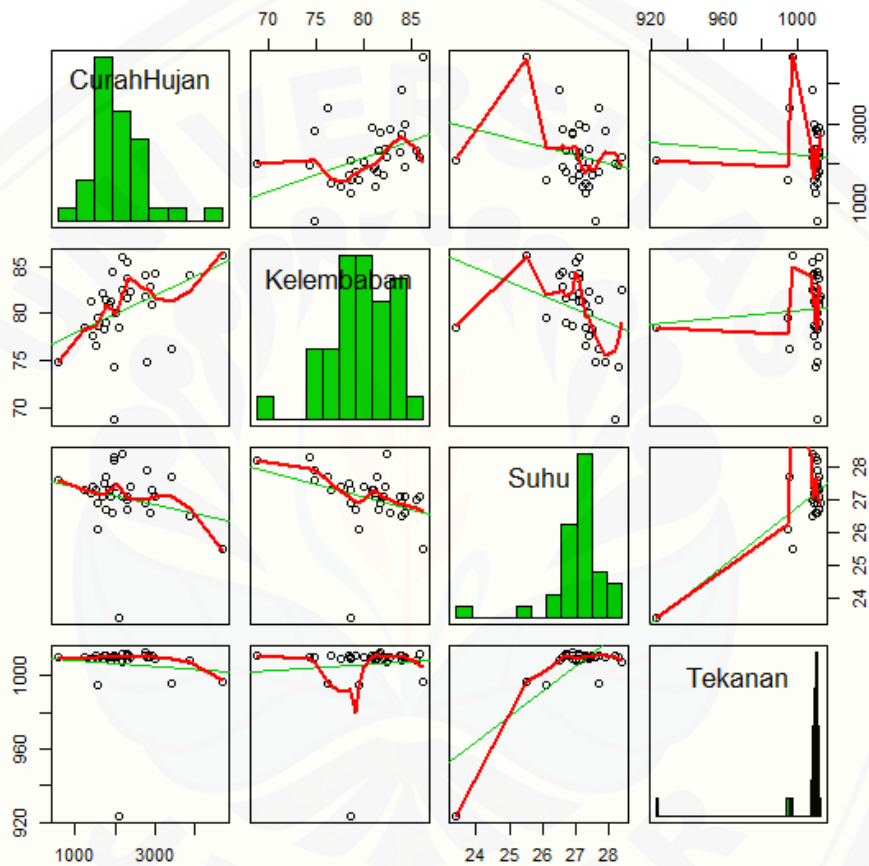
```
hujan.df <- data.frame(x=hujan$Kelembaban, y=hujan$Suhu,  
z=hujan$CurahHujan)  
hujan.lo <- loess(z ~ x*y, hujan.df, degree=1, span=0.7, normalize  
=FALSE)  
hujan.mar <- list(x=seq(68,87,length = 100), y=seq(24,28,length =  
100))  
new.dat <- expand.grid(hujan.mar)  
hujan.pred <- predict(hujan.lo, new.dat)  
eqscplot(hujan.mar, type = "n")  
contour(hujan.mar$x, hujan.mar$y, hujan.pred, ylim =  
range(c(20,30), finite = TRUE),  
levels = seq(500, 4700, 100), add = TRUE)  
## original points  
points(hujan.df, col = "red")
```



Lampiran C. Hasil Analisis Data

Hasil Eksplorasi Data

```
>library(car) #digunakan untuk mengerjakan scatterplotmatriks
>data(hujan)
>scatterplotMatrix(~CurahHujan+Kelembaban+Suhu+Tekanan,
reg.line=lm, smooth=TRUE, spread=FALSE, diagonal = 'histogram',
data=hujan)
```



Hasil Analisis Data Menggunakan Prosedur GAM

1. Penentuan Link Distribusi Data

```
> gam.new1 <- gam(CurahHujan ~ Kelembaban + Suhu + Tekanan,
family=Gamma(link=identity), data=hujan)
> summary(gam.new1)
```

```
Call: gam(formula = CurahHujan ~ Kelembaban + Suhu + Tekanan,
family = Gamma(link = identity),
data = hujan)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|--------|
| -0.9788 | -0.2184 | -0.1165 | 0.2029 | 0.6758 |

(Dispersion Parameter for Gamma family taken to be 0.1226)

Null Deviance: 4.255 on 32 degrees of freedom
Residual Deviance: 3.633 on 29 degrees of freedom
AIC: 535.324

Number of Local Scoring Iterations: 16

Anova for Parametric Effects

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------|----|--------|---------|---------|-----------|
| Kelembaban | 1 | 0.5317 | 0.53173 | 4.3381 | 0.04619 * |
| Suhu | 1 | 0.0110 | 0.01102 | 0.0899 | 0.76646 |
| Tekanan | 1 | 0.0425 | 0.04248 | 0.3466 | 0.56061 |
| Residuals | 29 | 3.5546 | 0.12257 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> gam.new2 <- gam(CurahHujan ~ Kelembaban + Suhu + Tekanan,
family=Gamma(link=log), data=hujan)
> summary(gam.new2)
```

```
Call: gam(formula = CurahHujan ~ Kelembaban + Suhu + Tekanan,
family = Gamma(link = log),
data = hujan)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|---------|---------|--------|--------|
| -0.9589 | -0.2318 | -0.1171 | 0.1801 | 0.6043 |

(Dispersion Parameter for Gamma family taken to be 0.12)

Null Deviance: 4.255 on 32 degrees of freedom
Residual Deviance: 3.5433 on 29 degrees of freedom
AIC: 534.4847

Number of Local Scoring Iterations: 7

Anova for Parametric Effects

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|------------|----|--------|---------|---------|-----------|
| Kelembaban | 1 | 0.5936 | 0.59355 | 4.9443 | 0.03413 * |
| Suhu | 1 | 0.0252 | 0.02521 | 0.2100 | 0.65020 |
| Tekanan | 1 | 0.0290 | 0.02904 | 0.2419 | 0.62653 |
| Residuals | 29 | 3.4814 | 0.12005 | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> AIC(gam.new1, gam.new2)
```

| | df | AIC |
|----------|----|----------|
| gam.new1 | 5 | 535.3240 |
| gam.new2 | 5 | 534.4847 |

2. Penentuan Derajat Polinomial

```
> gama <- gam(CurahHujan ~ Kelembaban + lo(Suhu, Tekanan, degree =
1), family = Gamma(link = log), data = hujan)
```

```
> gamb <- gam(CurahHujan ~ Kelembaban + lo(Suhu, Tekanan, degree =
2), family = Gamma(link = log), data = hujan)
```

3. Penentuan Span

```
> lk1 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.3)
+ Tekanan, family=Gamma(link=log), data=hujan)
```

Warning messages:

```
1: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
2: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
3: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
4: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
5: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
```

```
> summary(lk1)
```

```
Call: gam(formula = CurahHujan ~ lo(Kelembaban, Suhu, degree = 1, span = 0.3) +
  Tekanan, family = Gamma(link = log), data = hujan)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-0.773748 -0.138711  0.001576  0.159502  0.403566
```

(Dispersion Parameter for Gamma family taken to be 0.1051)

Null Deviance: 4.255 on 32 degrees of freedom

Residual Deviance: 2.0108 on 16.6382 degrees of freedom

AIC: 540.258

Number of Local Scoring Iterations: 9

Anova for Parametric Effects

| | Df | Sum Sq | Mean Sq | F value |
|--|--------|---------|---------|---------|
| lo(Kelembaban, Suhu, degree = 1, span = 0.3) | 2.000 | 0.71035 | 0.35517 | 3.3795 |
| Tekanan | 1.000 | 0.53138 | 0.53138 | 5.0561 |
| Residuals | 16.638 | 1.74862 | 0.10510 | |
| | | | Pr(>F) | |
| lo(Kelembaban, Suhu, degree = 1, span = 0.3) | | | 0.05865 | . |
| Tekanan | | | 0.03840 | * |
| Residuals | | | | |
| --- | | | | |
| Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 | | | | |

Anova for Nonparametric Effects

| | Npar | Df | Npar F | Pr(F) |
|--|------|--------|--------|-------|
| (Intercept) | | | | |
| lo(Kelembaban, Suhu, degree = 1, span = 0.3) | 12.4 | 1.1328 | 0.3983 | |
| Tekanan | | | | |

```
> lk2 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.4)
+ Tekanan, family=Gamma(link=log), data=hujan)
```

Warning messages:

```
1: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
2: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
3: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
4: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
```

```
> summary(lk2)
```

```
Call: gam(formula = CurahHujan ~ lo(Kelembaban, Suhu, degree = 1, span = 0.4) +
  Tekanan, family = Gamma(link = log), data = hujan)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-0.851092 -0.114340 -0.003981  0.076558  0.516973
```

(Dispersion Parameter for Gamma family taken to be 0.0965)

Null Deviance: 4.255 on 32 degrees of freedom
 Residual Deviance: 2.1166 on 19.4851 degrees of freedom
 AIC: 536.2741

Number of Local Scoring Iterations: 9

Anova for Parametric Effects

| | Df | Sum Sq | Mean Sq | F value |
|--|--------|--------|---------|---------|
| lo(Kelembaban, Suhu, degree = 1, span = 0.4) | 2.000 | 0.6635 | 0.33175 | 3.4391 |
| Tekanan | 1.000 | 1.1186 | 1.11863 | 11.5964 |
| Residuals | 19.485 | 1.8796 | 0.09646 | |

| | Pr(>F) |
|--|-------------|
| lo(Kelembaban, Suhu, degree = 1, span = 0.4) | 0.052580 . |
| Tekanan | 0.002887 ** |
| Residuals | |

Residuals

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects

| | Npar | Df | Npar F | Pr(F) |
|--|------|--------|--------|-------|
| (Intercept) | | | | |
| lo(Kelembaban, Suhu, degree = 1, span = 0.4) | 9.5 | 1.4381 | 0.2375 | |
| Tekanan | | | | |

```
> lk3 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.5)
+ Tekanan, family=Gamma(link=log), data=hujan)
```

Warning messages:

```
1: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
2: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
3: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
```

```
> summary(lk4)
```

```
Call: gam(formula = CurahHujan ~ lo(Kelembaban, Suhu, degree = 1, span = 0.6) +
  Tekanan, family = Gamma(link = log), data = hujan)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|----------|----------|---------|---------|---------|
| -0.85235 | -0.14733 | 0.01279 | 0.08958 | 0.58827 |

(Dispersion Parameter for Gamma family taken to be 0.0896)

Null Deviance: 4.255 on 32 degrees of freedom
 Residual Deviance: 2.2454 on 22.9258 degrees of freedom
 AIC: 531.3626

Number of Local Scoring Iterations: 8

Anova for Parametric Effects

| | Df | Sum Sq | Mean Sq | F value |
|--|--------|---------|---------|---------|
| lo(Kelembaban, Suhu, degree = 1, span = 0.6) | 2.000 | 0.67526 | 0.33763 | 3.7665 |
| Tekanan | 1.000 | 1.47021 | 1.47021 | 16.4012 |
| Residuals | 22.926 | 2.05507 | 0.08964 | |

| | Pr(>F) |
|--|---------------|
| lo(Kelembaban, Suhu, degree = 1, span = 0.6) | 0.0385130 * |
| Tekanan | 0.0004996 *** |
| Residuals | |

Residuals

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects

| | Npar | Df | Npar F | Pr(F) |
|--|------|--------|--------|-------|
| (Intercept) | | | | |
| lo(Kelembaban, Suhu, degree = 1, span = 0.6) | 6.1 | 2.0971 | 0.0925 | . |
| Tekanan | | | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> lk4 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.6)
+ Tekanan, family=Gamma(link=log), data=hujan)
```

Warning messages:

```
1: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
2: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
3: In lo.wam(x, z, wz, fit$smooth, which, fit$smooth.frame, bf.maxit, :
  lo.wam convergence not obtained in 30 iterations
```

```
> summary(lk4)
```

```
Call: gam(formula = CurahHujan ~ lo(Kelembaban, Suhu, degree = 1, span = 0.6) +
  Tekanan, family = Gamma(link = log), data = hujan)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-0.85235 -0.14733  0.01279  0.08958  0.58827
```

(Dispersion Parameter for Gamma family taken to be 0.0896)

Null Deviance: 4.255 on 32 degrees of freedom

Residual Deviance: 2.2454 on 22.9258 degrees of freedom

AIC: 531.3626

Number of Local Scoring Iterations: 8

Anova for Parametric Effects

| | Df | Sum Sq | Mean Sq | F value |
|--|--------|---------|---------|---------|
| lo(Kelembaban, Suhu, degree = 1, span = 0.6) | 2.000 | 0.67526 | 0.33763 | 3.7665 |
| Tekanan | 1.000 | 1.47021 | 1.47021 | 16.4012 |
| Residuals | 22.926 | 2.05507 | 0.08964 | |

Pr(>F)

```
lo(Kelembaban, Suhu, degree = 1, span = 0.6) 0.0385130 *
Tekanan 0.0004996 ***
Residuals
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Anova for Nonparametric Effects

| | Npar | Df | Npar F | Pr(F) |
|--|------|--------|--------|-------|
| (Intercept) | | | | |
| lo(Kelembaban, Suhu, degree = 1, span = 0.6) | 6.1 | 2.0971 | 0.0925 | . |
| Tekanan | | | | |

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> lk5 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.7)
+ Tekanan, family=Gamma(link=log), data=hujan)
```

```
> summary(lk5)
```

```
Call: gam(formula = CurahHujan ~ lo(Kelembaban, Suhu, degree = 1, span = 0.7) +
  Tekanan, family = Gamma(link = log), data = hujan)
```

Deviance Residuals:

```
      Min       1Q   Median       3Q      Max
-0.91990 -0.13526 -0.03211  0.11397  0.42482
```

(Dispersion Parameter for Gamma family taken to be 0.0762)

Null Deviance: 4.255 on 32 degrees of freedom

Residual Deviance: 2.2459 on 25.1607 degrees of freedom

AIC: 526.9003

Number of Local Scoring Iterations: 7

Anova for Parametric Effects

| | Df | Sum Sq | Mean Sq | F value |
|--|--------|---------|---------|---------|
| lo(Kelembaban, Suhu, degree = 1, span = 0.7) | 2.000 | 0.65827 | 0.32914 | 4.3207 |
| Tekanan | 1.000 | 0.75882 | 0.75882 | 9.9615 |
| Residuals | 25.161 | 1.91663 | 0.07618 | |

```

                                Pr(>F)
lo(Kelembaban, Suhu, degree = 1, span = 0.7) 0.024374 *
Tekanan                                       0.004116 **
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects

                                Npar Df Npar F   Pr(F)
(Intercept)
lo(Kelembaban, Suhu, degree = 1, span = 0.7)   3.8 4.0229 0.01255 *
Tekanan
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> lk6 <- gam(CurahHujan ~ lo(Kelembaban, Suhu, degree=1, span=0.8)
+ Tekanan, family=Gamma(link=log), data=hujan)
> summary(lk6)

Call: gam(formula = CurahHujan ~ lo(Kelembaban, Suhu, degree = 1, span = 0.8) +
  Tekanan, family = Gamma(link = log), data = hujan)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.92553 -0.15021 -0.04483  0.10507  0.44657

(Dispersion Parameter for Gamma family taken to be 0.0802)

Null Deviance: 4.255 on 32 degrees of freedom
Residual Deviance: 2.3919 on 26.0792 degrees of freedom
AIC: 527.1662

Number of Local Scoring Iterations: 9

Anova for Parametric Effects

                                Df Sum Sq Mean Sq F value
lo(Kelembaban, Suhu, degree = 1, span = 0.8) 2.000 0.65240 0.32620 4.0652
Tekanan                                       1.000 0.52945 0.52945 6.5982
Residuals                                   26.079 2.09264 0.08024
                                Pr(>F)
lo(Kelembaban, Suhu, degree = 1, span = 0.8) 0.02906 *
Tekanan                                       0.01628 *
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova for Nonparametric Effects

                                Npar Df Npar F   Pr(F)
(Intercept)
lo(Kelembaban, Suhu, degree = 1, span = 0.8)   2.9 4.5226 0.01159 *
Tekanan
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Plot Surface Model

```

>library(asbio)
>data(hujan)
>hujan.df <- data.frame(x=hujan$Kelembaban, y=hujan$Suhu,
z=hujan$CurahHujan)
>loess.surf(hujan.df$z, cbind(hujan.df$x, hujan.df$y), phi = 15,
theta = 2)

```


Plot Kontur Variabel

```
>require(MASS)
>data(hujan)
>hujan
>hujan.df <- data.frame(x=hujan$Kelembaban, y=hujan$Suhu,
z=hujan$CurahHujan)
>hujan.lo <- loess(z ~ x*y, hujan.df, degree=1, span=0.7,
normalize =FALSE)
>hujan.mar <- list(x=seq(68,87,length = 100), y=seq(24,28,length =
100))
>new.dat <- expand.grid(hujan.mar)
>hujan.pred <- predict(hujan.lo, new.dat)
>eqscplot(hujan.mar, type = "n")
>contour(hujan.mar$x, hujan.mar$y, hujan.pred, ylim =
range(c(20,30), finite = TRUE),
levels = seq(500, 4700, 100), add = TRUE)
## original points
>points(hujan.df, col = "red")
```