BAB III

GEOGRAPHICALLY WEIGHTED POISSON REGRESSION SEMIPARAMETRIC (GWPRS)

3.1 Model *Geographically Weighted Poisson Regression Semiparametric* (GWPRS)

Geographically Weighted Poisson Regression Semiparametric (GWPRS) merupakan model GWPR yang mengkombinasikan parameter-parameter bersifat lokal dengan parameter-parameter bersifat global terhadap lokasi (Nakaya, dkk, 2005). Model GWPRS dinyatakan sebagai berikut:

$$y_i \sim Poisson\ (\mu_i)$$

dengan

$$\mu_i = exp\left(\sum_{j=0}^{k^*} \beta_j(u_i, v_i) x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g\, x_{ig}\right) \qquad \ldots(3.1)$$

demikian sehingga

$$y_i \sim Poisson\ \left(exp\left(\sum_{j=0}^{k^*} \beta_j(u_i, v_i) x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g\, x_{ig}\right)\right) \qquad \ldots(3.2)$$

dimana :

y_i merupakan nilai observasi peubah terikat ke-i

x_{ij} merupakan nilai observasi peubah bebas lokal pada lokasi (u_i, v_i)

$\beta_j(u_i, v_i)$ merupakan parameter model peubah bebas lokal pada lokasi (u_i, v_i)

γ_g merupakan parameter model peubah bebas global pada lokasi (u_i, v_i)

(u_i, v_i) merupakan titik koordinat (*longitude, latitude)*

x_{ig} merupakan nilai observasi peubah bebas global pada lokasi (u_i, v_i)

3.1.1 Penaksiran Parameter Model GWPRS

Parameter model GWPRS ditaksir oleh Metode Kemungkinan Maksimum dengan bentuk fungsi *likelihood* sebagai berikut:

$$L(\beta_j(u_i, v_i), \gamma_g) = \prod_{i=1}^{n} \frac{e^{-\mu_i} {\mu_i}^{y_i}}{y_i!}$$

Persamaan ln *likelihood* :

$$\ln L(\beta_j(u_i, v_i), \gamma_g) = \ln \prod_{i=1}^{n} \frac{e^{-\mu_i} {\mu_i}^{y_i}}{y_i!}$$

$$= \sum_{i=1}^{n} \ln \frac{e^{-\mu_i} {\mu_i}^{y_i}}{y_i!}$$

$$= \sum_{i=1}^{n} (\ln(e^{-\mu_i}) + \ln({\mu_i}^{y_i}) - \ln(y_i!))$$

$$= \sum_{i=1}^{n} (-\mu_i + y_i \ln(\mu_i) - \ln(y_i!))$$

Kemudian subtitusikan $\mu_i = exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right)$

$$= \sum_{i=1}^{n} \left(-exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right)\right)$$

$$+ y_i \ln(exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right) - \ln(y_i!)$$

$$= \sum_{i=1}^{n} \left(-exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right)\right)$$

$$+ y_i\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right) - \ln(y_i!) \qquad \ldots(3.3)$$

Letak geografis merupakan pembobot pada model GWPRS, selanjutnya pembobot dimasukkan pada persamaan (3.3) untuk memperoleh model GWPRS :

$$\ln L^*\left(\beta_j(u_i,v_i),\gamma_g\right) = \left(-\sum_{i=1}^{n} exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right)\right)$$

$$+ \sum_{i=1}^{n} y_i\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right)$$

$$- \sum_{i=1}^{n} \ln(y_i!)) \, W_{ij}(u_i,v_i) \qquad \ldots(3.4)$$

Untuk memaksimumkan fungsi ln *likelihood*, persamaan (3.4) didiferensialkan terhadap $\beta_j(u_i,v_i)$ dan γ_g. Selanjutnya hasil pendiferensialan tersebut dibuat sama dengan 0, sehingga diperoleh :

$$\frac{\partial \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial \beta_j(u_i,v_i)} = -\sum_{i=1}^{n} x_{ij} W_{ij}(u_i,v_i) exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij}\right)$$

$$+ \sum_{g=k^*+1}^{k} \gamma_g x_{ig}) + \sum_{i=1}^{n} y_i x_{ij} W_{ij}(u_i,v_i) = 0$$

$$\frac{\partial \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial \gamma_g} = -\sum_{i=1}^{n} x_{ig} W_{ij}(u_i,v_i) exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij}\right)$$

$$+ \sum_{g=k^*+1}^{k} \gamma_g x_{ig}) + \sum_{i=1}^{n} y_i x_{ig} W_{ij}(u_i,v_i) = 0$$

Proses penyelesaian persamaan di atas dilakukan dengan menggunakan iterasi Newton Raphson. Prosedur penaksiran parameter model GWPRS adalah dengan menaksir $\beta_j(u_i,v_i)$ dan γ_g (Nakaya, dkk, 2005). Bentuk umum persamaan penaksir parameter-parameter model GWPRS adalah sebagai berikut :

$$\left(\beta^{(t+1)}(u_i,v_i),\gamma^{(m+1)}\right) = \left(\beta^{(t)}(u_i,v_i),\gamma^{(m)}\right)$$

$$- \left({H^{(t)}}^{-1}\left(\beta^{(t)}(u_i,v_i),\gamma^{(m)}\right)\right)\left(g^{(t)}\left(\beta^{(t)}(u_i,v_i),\gamma^{(m)}\right)\right)$$

dengan

$$g^{(t)}\left(\beta^{(t)}(u_i,v_i),\gamma^{(m)}\right)=\begin{bmatrix}\dfrac{\partial \ln L^*\left(\beta_j(u_i,v_i),\gamma\right)}{\partial \beta_0(u_i,v_i)}\\ \dfrac{\partial \ln L^*\left(\beta_j(u_i,v_i),\gamma\right)}{\partial \beta_1(u_i,v_i)}\\ \vdots \\ \dfrac{\partial \ln L^*\left(\beta_j(u_i,v_i),\gamma\right)}{\partial \beta_{k*}(u_i,v_i)}\\ \dfrac{\partial \ln L^*\left(\beta_j(u_i,v_i),\gamma\right)}{\partial \gamma_{k*+1}}\\ \vdots \\ \dfrac{\partial \ln L^*\left(\beta_j(u_i,v_i),\gamma\right)}{\partial \gamma_k}\end{bmatrix}$$

$$H^{(t)^{-1}}\left(\beta^{(t)}(u_i,v_i),\gamma^{(m)}\right)=$$

$$\begin{bmatrix}
\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_0(u_i,v_i)\beta_0(u_i,v_i)} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_1(u_i,v_i)\beta_0(u_i,v_i)} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_{k*}(u_i,v_i)\beta_0(u_i,v_i)} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_{k*+1}\beta_0(u_i,v_i)} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_k\beta_0(u_i,v_i)} \\
\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_0(u_i,v_i)\beta_1(u_i,v_i)} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_1(u_i,v_i)\beta_1(u_i,v_i)} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_{k*}(u_i,v_i)\beta_1(u_i,v_i)} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_{k*+1}\beta_1(u_i,v_i)} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_k\beta_1(u_i,v_i)} \\
\vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\
\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_0(u_i,v_i)\beta_{k*}(u_i,v_i)} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_1(u_i,v_i)\beta_{k*}(u_i,v_i)} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_{k*}(u_i,v_i)\beta_{k*}(u_i,v_i)} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_{k*+1}\beta_{k*}(u_i,v_i)} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_k\beta_{k*}(u_i,v_i)} \\
\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_0(u_i,v_i)\gamma_{k*+1}} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_1(u_i,v_i)\gamma_{k*+1}} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_{k*}(u_i,v_i)\gamma_{k*+1}} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_{k*+1}\gamma_{k*+1}} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_k\gamma_{k*+1}} \\
\vdots & \vdots & \ddots & \vdots & \vdots & & \vdots \\
\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_0(u_i,v_i)\gamma_k} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_1(u_i,v_i)\gamma_k} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\beta_{k*}(u_i,v_i)\gamma_k} & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_{k*+1}\gamma_k} & \cdots & \frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_k\gamma_k}
\end{bmatrix}$$

Untuk setiap langkah iterasi ke-t berlaku :

$$\frac{\partial \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial \beta_j(u_i,v_i)}=-\sum_{i=1}^{n}x_{ij}W_{ij}(u_i,v_i)$$

$$\left(y_i-exp\left(\sum_{j=0}^{k^*}\beta_j(u_i,v_i)x_{ij}+\sum_{g=k^*+1}^{k}\gamma_g x_{ig}\right)\right)=0$$

$$\frac{\partial \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial \gamma_g}=-\sum_{i=1}^{n}x_{ig}W_{ij}(u_i,v_i)$$

$$\left(y_i-exp\left(\sum_{j=0}^{k^*}\beta_j(u_i,v_i)x_{ij}+\sum_{g=k^*+1}^{k}\gamma_g x_{ig}\right)\right)=0$$

$$\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial \beta^T(u_i,v_i)\beta_j(u_i,v_i)}=-\sum_{i=1}^{n}x_{ij}W_{ij}(u_i,v_i)$$

$$\left(x_{ij}exp\left(\sum_{j=0}^{k^*}\beta_j(u_i,v_i)x_{ij}+\sum_{g=k^*+1}^{k}\gamma_g x_{ig}\right)\right)=0$$

$$\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial \gamma^T\gamma_g}=-\sum_{i=1}^{n}x_{ig}W_{ij}(u_i,v_i)$$

$$\left(x_{ig} exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right)\right) = 0$$

$$\frac{\partial^2 \ln L^*(\beta_j(u_i,v_i),\gamma)}{\partial\gamma_g \beta_j(u_i,v_i)} = -\sum_{i=1}^{n} x_{ig}x_{ij}W_{ij}(u_i,v_i)$$

$$exp\left(\sum_{j=0}^{k^*} \beta_j(u_i,v_i)x_{ij} + \sum_{g=k^*+1}^{k} \gamma_g x_{ig}\right) = 0$$

Dengan mengulang prosedur iterasi untuk setiap titik regresi ke-i, maka penaksir parameter lokal dan penaksir parameter global akan diperoleh. Iterasi berhenti pada saat konvergen, yaitu pada saat $\left\|\beta^{(t+1)}(u_i,v_i) - \beta^{(t)}(u_i,v_i)\right\| \le \varepsilon$ dan $\left\|\gamma^{(t+1)} - \gamma^{(t)}\right\| \le \varepsilon$. Penyelesaian persamaan iterasi tersebut memerlukan bantuan *software* GWR4.

3.1.2 Pengujian Parameter Model GWPRS

Pengujian parameter GWPRS dilakukan secara parsial. Pengujian ini untuk mengetahui peubah bebas mana saja yang berpengaruh terhadap variebel terikat. Langkah-langkah pengujian hipotesis untuk parameter model GWPRS secara parsial pada parameter global (Ningsih, 2016):

1. Perumusan Hipotesis

   $H_0 : \gamma_g = 0$ (parameter peubah yang bersifat global tidak signifikan)

   $H_1 : \gamma_g \ne 0$; (parameter peubah yang bersifat global signifikan)

2. Statistik Uji

$$Z = \frac{\hat{\gamma}_g}{SE(\hat{\gamma}_g)}$$

3. Kriteria Pengujian

   Dengan taraf kepercayaan sebesar $\alpha = 5\%$, Tolak H_0, apabila $|Z_{hitung}| > Z_{(\frac{\alpha}{2})}$

4. Kesimpulan

   Penafsiran H_0 diterima atau ditolak.

Langkah-langkah pengujian hipotesis untuk parameter model GWPRS secara parsial pada parameter lokal (Ningsih, 2016):

1. Perumusan Hipotesis

   $H_0 :\ \beta_j(u_i,v_i) = 0$ (parameter peubah yang bersifat lokal tidak signifikan)

   $H_1 :\ \beta_j(u_i,v_i) \ne 0$ (parameter peubah yang bersifat lokal signifikan)

2. Statistik Uji

$$Z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

3. Kriteria Pengujian

   Dengan taraf kepercayaan sebesar $\alpha = 5\%$, tolak H_0, apabila $|Z_{hitung}| > Z_{(\frac{\alpha}{2})}$

4. Kesimpulan

   Penafsiran H_0 diterima atau ditolak.

3.2 Pembobot fungsi

Pembobot lokasi ke-i pada koordinat (u_i, v_i) dinyatakan dengan $W(u_i, v_i)$. Pembobotan digunakan untuk memberikan nilai yang berbeda di setiap lokasi pengamatan. Pembobot $W(u_i, v_i)$ dapat ditentukan dengan menggunakan fungsi kernel. Fungsi kernel memberikan pembobot sesuai *bandwidth* optimum yang nilainya bergantung pada kondisi data. Menurut Fischer, dkk. (2009) terdapat dua tipe pembobot fungsi kernel yaitu :

1. Fungsi *fixed* kernel

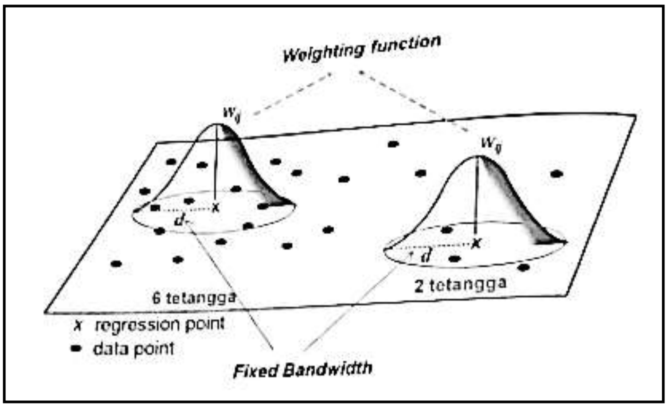
   Fungsi *fixed* kernel, memiliki *bandwidth* optimum yang sama pada setiap lokasi pengamatan. Terdapat dua jenis fungsi *fixed* kernel yaitu :

   a. Fungsi *fixed* Gaussian :

$$w_{ij} = exp\left[-\left(\frac{d_{ij}}{b}\right)^2\right] \qquad \ldots(3.5)$$

   b. Fungsi *Fixed Bisquare* :

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 ; d_{ij} < b \\ 0 \; ; d_{ij} > b \end{cases} \qquad \ldots(3.6)$$



Gambar 3. 1 Ilustrasi *Fixed* Kernel
Sumber : Badan Pusat Statistik

2. Fungsi *adaptive* kernel

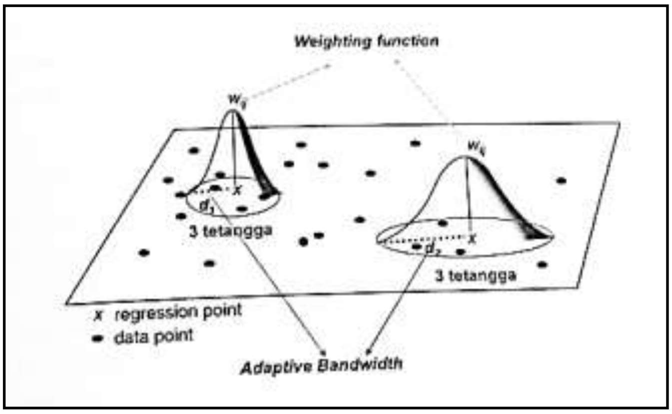
   Fungsi *adaptive* kernel memiliki *bandwidth* optimum yang berbeda-beda sesuai dengan lokasi pengamatan. Terdapat dua jenis fungsi *adaptive* kernel yaitu :

   a. Fungsi *Adaptive Bisquare* :

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b_i}\right)^2\right)^2 ; d_{ij} < b_i \\ 0 \ ; d_{ij} > b_i \end{cases} \qquad \ldots(3.7)$$

   b. Fungsi *Adaptive Gaussian* :

$$w_{ij} = exp\left[-\left(\frac{d_{ij}}{b_i}\right)^2\right] \qquad \ldots(3.8)$$



Gambar 3. 2 Ilustrasi *Adaptive* Kernel
Sumber : Badan Pusat Statistik

dimana :

w_{ij} merupakan nilai pembobot dari lokasi pengamatan ke-j terhadap titik lokasi pengamatan ke-i

d_{ij} merupakan jarak e*uclidean* antara lokasi pengamatan ke-j terhadap titik lokasi pengamatan ke-i

b merupakan *bandwidth*

b_i merupakan *bandwidth* yang diadaptasi sebagai jarak tetangga terdekat ke-i.

Jarak *euclidean* merupakan jarak antara titik regresi ke-i dengan lokasi ke-j $(i \neq j)$ yang dinotasikan dengan d_{ij} dan dirumuskan sebagai berikut:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2} \quad \ldots(3.9)$$

dimana :

u_i merupakan *longitude* pada lokasi ke *i*.

u_j merupakan *longitude* pada lokasi ke j.

v_i merupakan *latitude* pada lokasi ke *i*.

v_j merupakan *latitude* pada lokasi ke j.

3.3 Pemilihan *Bandwidth*

Bandwidth merupakan radius suatu lingkaran dengan pusat titik lokasi *i*. Apabila terdapat titik-titik lokasi yang berada pada lingkaran tersebut, maka masih dianggap memiliki pengaruh terhadap penaksiran koefisien regresi pada titik lokasi *i* terebut. Pemilihan *bandwidth* memiliki dampak besar pada hasil yang diperoleh dari GWR. Menurut Nakaya, dkk. (2005), apabila nilai *bandwidth* kecil, hal ini akan menyebabkan variansi menjadi semakin besar dan model yang diperoleh sangat kasar (*undersmoothing*) sehingga tidak mewakili keadaan yang sebenarnya. Sebaliknya, apabila nilai *bandwidth* besar akan menyebabkan variansi yang semakin kecil sehingga model yang diperoleh terlalu halus (*oversmoothing*) sehingga model yang kita dapat tidak berarti, hal ini dikarenakan tidak terdapat perbedaan antar lokasi pengamatan. Apabila model yang diperoleh tidak memiliki perbedaan antar lokasi pengamatan lebih baik menggunakan regresi global. Oleh karena itu diperlukan nilai *bandwidth* yang optimum. Penentuan nilai bandwidth yang optimum dapat dilakukan dengan menggunakan metode *Cross Validation* (CV) yang ditentukan dengan menggunakan rumus sebagai berikut :

$$CV = \sum_{i=1}^{n}[y_i - \hat{y}_{i\neq i}(b)]^2 \quad \ldots(3.10)$$

Dimana $\hat{y}_{i\neq i}(b)$ adalah nilai penaksir y_i dengan radius b. Penentuan nilai *bandwidth* yang optimum diperoleh saat nilai CV minimum. Proses untuk mendapatkan nilai *bandwidth* dapat dilakukan menggunakan metode *golden section search* pada *software* GWR4.

3.4 *Akaike Information Criterion Corrected* (AICc)

AIC merupakan kriteria kesesuaian model untuk menaksir parameter model secara statistik atau untuk mengetahui seberapa dekat parameter yang ditaksir dengan nilai populasi. Perhitungan AIC dapat dilakukan menggunakan rumus sebagai berikut:

$$AIC = 2k - 2\ln(likelihood) \quad \ldots(3.11)$$

dimana :

k merupakan banyak parameter yang akan ditaksir.

ln(*likelihood*) merupakan nilai maksimum *likelihood* model.

Kriteria kesesuaian model untuk menaksir parameter model dengan menggunakan AIC dinilai bias untuk sampel dengan jumlah terbatas. Namun, hal ini dapat dikoreksi dengan AIC *Corrected* (AICc). Penentuan AIC *Corrected* (AICc) dilakukan dengan menggunakan rumus berikut:

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad \ldots(3.12)$$

dimana n merupakan ukuran sampel. Apabila nilai k semakin besar atau peubah yang ditaksirnya semakin banyak, maka untuk penggunaan AICc akan lebih baik daripada dengan nilai AIC. Pemilihan Model Terbaik dari beberapa model dapat dipilih dari model yang memiliki nilai AICc terkecil. Alasan digunakannya AICc adalah berawal dari prinsip parsimony yang menyatakan bahwa model terbaik diharapkan terbentuk dari koefisien/parameter regresi yang tidak banyak tapi mampu menjelaskan model secara keseluruhan (Ningsih, 2016).