PENDEKATAN METODE PEMULUSAN KERNEL PADA PENDUGAAN AREA KECIL (*SMALL AREA ESTIMATION*)

Indahwati[1], Kusman Sadik[1], Ratih Nurmasari[2]

[1]Dosen Departemen Statistika FMIPA IPB
[2]Mahasiswa S1 Departemen Statistika FMIPA IPB

ABSTRAK

Pendugaan area kecil merupakan pendugaan parameter suatu area yang lebih kecil dengan memanfaatkan informasi dari luar area, dari dalam area itu sendiri, dan dari luar survei. Berdasarkan peubah penjelas yang digunakan, terdapat dua model area kecil, yaitu *basic area level model* dan *basic unit level model*, dimana kedua model tersebut mengasumsikan bahwa penduga langsung memiliki hubungan yang linier dengan peubah penjelas. Ada kalanya asumsi tersebut tidak dapat dipenuhi dan salah satu solusinya adalah dengan menggunakan pendekatan nonparametrik, seperti pemulusan Kernel. Simulasi yang telah dilakukan menunjukkan bahwa pemulusan Kernel dapat mereduksi bias pendugaan pada pola hubungan yang tidak linier dengan berbagai jumlah area. Nilai *Mean Square Error* (MSE) pendugaan area kecil dengan menggunakan pemulusan Kernel pada pola hubungan yang tidak linier relatif lebih kecil dibandingkan metode parametrik yang menggunakan model Fay-Herriot. MSE pada pemulusan Kernel memiliki kecenderungan semakin kecil jika jumlah area semakin banyak.

Kata Kunci : nonparametrik, pendugaan area kecil, pemulusan Kernel

PENDAHULUAN

Latar Belakang

Metode pendugaan area kecil dapat dilakukan dengan cara memodelkan penduga langsung dengan peubah penjelas. Berdasarkan peubah penjelas yang tersedia, terdapat dua model area kecil, yaitu *basic area level model* dan *basic unit level model*. Kedua model tersebut mengasumsikan bahwa penduga langsung memiliki hubungan yang linier dengan peubah penjelas (Mukhopadhyay & Maiti 2004). Salah satu solusi untuk mengatasi masalah asumsi tersebut adalah dengan menggunakan pendekatan nonparametrik.

Beberapa peneliti telah mencoba mengkaji pendekatan nonparametrik pada pendugaan area kecil, diantaranya Pushpal K. Mukhopadhyay dan Tapabrata Maiti yang telah mengkaji metode Kernel pada tahun 2004 serta *local polynomial* *regression* pada tahun 2006, Opsomer *et al* juga telah melakukan pengkajian mengenai metode *penalized spline* pada pendugaan area kecil pada tahun 2004.

Metode nonparametrik yang digunakan pada penelitian ini adalah metode pemulusan Kernel. Mukhopadhyay dan Maiti (2004) telah mengkaji metode ini untuk area sebanyak 100, sedangkan penelitian ini juga akan mengkaji penerapan metode Kernel pada jumlah area yang berbeda-beda.

Tujuan

1. Mengkaji penerapan salah satu metode nonparametrik, yaitu pemulusan Kernel, pada pendugaan area kecil.
2. Mengetahui keefektifan pendugaan area kecil dengan metode pemulusan Kernel pada pola hubungan yang linier dan tak linier antara penduga langsung dengan peubah penjelas.
3. Mengetahui keefektifan pendugaan area kecil dengan metode pemulusan Kernel pada jumlah area yang berbeda-beda.

TINJAUAN PUSTAKA

Pendugaan Area Kecil

Area kecil atau *small area* diartikan sebagai bagian dari populasi, baik berdasarkan area geografi maupun sosial-demografi. Suatu daerah disebut area kecil jika di dalam daerah tersebut, contoh yang terambil kurang banyak untuk mendapatkan nilai penduga langsung dengan presisi yang memadai. Pendugaan area kecil merupakan pendugaan parameter suatu area yang lebih kecil dengan memanfaatkan informasi dari luar area, dari dalam area itu sendiri, dan dari luar survei (Rao 2003).

Dua jenis model eksplisit pada pendugaan area kecil adalah *basic area level model* dan *basic unit level model* (Rao 2003).

1. *Basic area level model* mengasumsikan bahwa peubah penjelas yang tersedia hanya ada untuk level area tertentu. Misalkan tersedia vektor peubah penjelas $\boldsymbol{x_i} = (\boldsymbol{x_{1i}}, \boldsymbol{x_{2i}}, \ldots, \boldsymbol{x_{pi}})^T$, dan parameter θ_i yang akan diduga diasumsikan memiliki hubungan dengan x_i. Peubah penjelas tersebut dimodelkan:

$\theta_i = \mathbf{x_i}^T \boldsymbol{\beta} + u_i$ ....(1)

dengan i = 1, 2, ... , m dan u_i ~ N(0, σ_u^2), dimana m merupakan banyaknya area kecil, $\boldsymbol{\beta}$ merupakan vektor koefisien regresi, dan u_i merupakan pengaruh acak pada area ke-i. Parameter θ_i dapat diketahui dengan mengasumsikan bahwa penduga langsung pada area ke-i (y_i) telah tersedia, yaitu:

$y_i = \theta_i + \varepsilon_i$ ....(2)

dimana i = 1, 2, ... , m, ε_i merupakan *sampling error* dan diasumsikan ε_i ~N(0,D_i). Model gabungan dari persamaan (1) dan (2) adalah:

$y_i = \mathbf{x_i}^T \boldsymbol{\beta} + u_i + \varepsilon_i$ ....(3)

dimana i = 1, 2, ... , m.

2. *Basic unit level model* merupakan model yang data-data pendukungnya bersesuaian secara individu dengan data respon, misal $\boldsymbol{x_{ij}} = (\boldsymbol{x_{ij1}}, \boldsymbol{x_{ij2}}, \ldots, \boldsymbol{x_{ijp}})^T$, sehingga dapat dibuat suatu model regresi tersarang:

   $y_{ij} = \mathbf{x_{ij}}^T \boldsymbol{\beta} + u_i + \varepsilon_i$ ....(4)

   dimana i = 1, 2, ..., m , j = 1, ..., n_i, dengan n_i merupakan banyaknya contoh yang tersurvei pada area ke-i, u_i~N(0, σ_u^2), dan ε_i ~ N (0,D_i).

Pemulusan Kernel

Secara sederhana, model regresi nonparametrik adalah:

$y = m(x) + \varepsilon$ ....(5)

dimana y merupakan peubah respon yang diamati, $m(x)$ merupakan fungsi regresi yang ingin diduga dan tidak dapat didekati dengan model parametrik, serta ε merupakan eror pengamatan yang tidak dapat dijelaskan oleh fungsi regresi $m(x)$.

Beberapa metode pemulusan yang biasa digunakan untuk menduga $m(x)$ pada persamaan (5) adalah *local polynomial smoothers*, *regression splines*, *smoothing splines*, *penalized splines*, dan *Kernel smoothers*.

Salah satu ide untuk menduga fungsi *m(x)* adalah dengan menggunakan *local averaging procedure* atau rata-rata lokal terboboti sebagai fungsi pemulus, yaitu:

$\hat{m}_h(x_i) = \frac{1}{m}\sum_{i=1}^{m} \hat{W}_{hi}(x) y_i$ ....(6)

dimana $\hat{m}_h(x_i)$ merupakan dugaan dari fungsi regresi pada titik pengamatan ke-i, *m* merupakan banyaknya pengamatan, y_i merupakan peubah respon pada pengamatan ke-i, dan $\hat{W}_{hi}(x)$ merupakan fungsi pembobot pada daerah di sekitar x_i dengan lebar jendela *h* (Hardle 1994).

Lebar jendela (*h*) merupakan parameter pemulusan yang menentukan kemulusan kurva yang dihasilkan. Beberapa cara yang dapat digunakan untuk menentukan *h* antara lain *least square cross validation*, *likelihood cross validation*, dan *generalized cross validation*. Nilai *h* juga bisa ditentukan secara subjektif atau berdasarkan penelitian yang telah dilakukan sebelumnya (Silverman 1986). Menurut Hardle (1994), lebar jendela optimum yang menghasilkan *asymptotic mean square error* (AMSE) minimum adalah $h \cong m^{-1/5}$.

Fungsi $\hat{W}_{hi}(x)$ yang digunakan pada pemulusan Kernel (*Kernel smoothers*):

$$\hat{W}_{hi}(x) = \frac{K_h(x - x_i)}{\hat{f}_h(x)} \quad \text{(7)}$$

dimana $\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i)$ dan K(.) merupakan fungsi Kernel yang dapat dinyatakan sebagai $K_h(u) = \frac{1}{h}K(\frac{u}{h})$. Fungsi K(.) memiliki sifat simetris, kontinu dan terhingga, serta $\int K(x)dx = 1$. Beberapa fungsi Kernel yang umum digunakan adalah *Box, Parzen, Triangle,* dan *Gaussian* (Normal).

Pemulusan Kernel pada Pendugaan Area Kecil

Jika hubungan antara penduga langsung dengan variabel penjelas tidak linier, maka persamaan (1) dapat didekati dengan sebuah fungsi yang lebih umum, yaitu:

$$\theta_i = m(x_i) + u_i \quad \text{(8)}$$

dimana:

i = 1, 2, ..., m, sedangkan m merupakan banyaknya area,

$m(x_i)$ = fungsi pemulusan yang menggambarkan hubungan yang sesungguhnya antara x dan y pada area ke-i,

θ_i = parameter pada area ke-i,

u_i = pengaruh acak dari area ke-i dan u_i~N(0, σ_u^2).

Fungsi $m(x_i)$ diduga dengan persamaan Nadaraya-Watson Kernel, yaitu:

$$\hat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{i=1}^{n} K_h(x - x_i)} \quad \text{...(9)}$$

Dugaan parameter pada area kecil ke-i, $\hat{\theta}_i$, adalah:

$$\hat{\theta}_i = E(\theta_i \mid y_i) = \gamma_i y_i + (1 - \gamma_i) \hat{m}_h(x_i) \quad \text{..(10)}$$

dimana $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2 + D_i}$, dan σ_u^2 dapat diduga dengan metode momen, yaitu:

$\hat{\sigma}_u^2 = \max(0, \frac{1}{m-1} \sum_{i=1}^{m} W_{hi}(x) \{ y_i - \hat{m}(x) \}^2 - D)$ dan D_i diasumsikan konstan untuk semua area.

Pendugaan Area Kecil dengan Model Fay-Herriot

Model yang digunakan oleh Fay dan Herriot (1979) adalah sebagai berikut:

$$y_i = \theta_i + \varepsilon_i = \mathbf{x_i}^T \boldsymbol{\beta} + u_i + \varepsilon_i \quad \text{...(11)}$$

dengan i = 1, 2, ... , m dan u_i ~ N(0, σ_u^2), dimana m merupakan banyaknya area kecil, y_i merupakan penduga langsung pada area ke-i, θ_i merupakan parameter area kecil yang menjadi perhatian dan akan diduga, $\boldsymbol{\beta}$ merupakan vektor koefisien regresi, u_i merupakan pengaruh acak pada area ke-i, ε_i merupakan *sampling error* dan diasumsikan ε_i ~ N(0,D_i). Nilai $\boldsymbol{\beta}$ dapat diduga dengan menggunakan *weighted least square*, sedangkan σ_u^2 diduga dengan *maximum likelihood* atau *restricted maximum likelihood*. $\hat{\theta}_i$ diperoleh dengan cara menghitung rata-rata terboboti antara penduga langsung, y_i, dan penduga sintetik, $x_i^T \hat{\beta}$, yaitu:

$$\hat{\theta}_i = \gamma_i y_i + (1 - \gamma_i) x^T \hat{\beta} \quad \text{...(12)}$$

dimana $\gamma_i = \frac{\sigma_u^2}{\sigma_u^2 + D_i}$.

DATA DAN METODE

Data

Peubah penjelas diasumsikan hanya tersedia untuk level area tertentu sehingga digunakan *basic area level model*. Data yang digunakan adalah data simulasi untuk beberapa area kecil dengan satu peubah penjelas. Jumlah area kecil (m) yang dibangkitkan adalah m=100, m=50, dan m=25. Pola data yang dibangkitkan juga berbeda-beda, yaitu linier dan tidak linier (kubik dan logaritma natural).

Metode

1. Membangkitkan data untuk setiap pola hubungan dan banyaknya area (m) tertentu dengan langkah-langkah sebagai berikut:
   a. membangkitkan peubah penjelas (x_i) sebanyak m, dimana x_i ~ uniform(0,1)
   b. memetakan x_i melalui fungsi matematis tertentu untuk memperoleh $m(x_i)$
   c. membangkitkan pengaruh acak area (u_i) sebanyak m, dimana u_i ~N(0,0.04)
   d. menghitung parameter area kecil θ_i, dimana $\theta_i = m(x_i)+u_i$
   e. membangkitkan eror (ε_i) sebanyak m, dimana sepertiga area pertama ε_i ~N(0,0.01), sepertiga area kedua ε_i ~N(0,0.04), dan sepertiga area terakhir ε_i ~N(0,0.09)
   f. menghitung penduga langsung (y_i), dimana $y_i = \theta_i + \varepsilon_i$.

   Langkah a-b dilakukan sebanyak satu kali karena x_i diasumsikan tetap (*fixed*). Langkah c-f diulang sebanyak R kali, dimana R yang digunakan adalah 100.
2. Menduga parameter area kecil dengan menggunakan dua pendekatan, yaitu pendekatan nonparametrik dengan menggunakan pemulusan Kernel dan pendekatan parametrik yang menggunakan model Fay-Herriot.
3. Mengukur performa simulasi pendugaan yang dilakukan dengan menghitung:

- *Absolute Relative Bias* (ARB): $ARB(\hat{\theta}_i) = \frac{1}{R}\left|\frac{\sum_{j=1}^{R}(\hat{\theta}_{ij}-\theta_{ij})}{\theta_{ij}}\right|$ ....(13)
- *Mean Square Error* (MSE): $MSE(\hat{\theta}_i) = \frac{1}{R}\sum_{j=1}^{R}(\hat{\theta}_{ij}-\theta_{ij})^2$ ....(14)

HASIL DAN PEMBAHASAN

Pembangkitan Data dan Pendugaan Parameter Area Kecil

Fungsi $m(x_i)$ yang digunakan untuk memetakan x_i adalah sebagai berikut:

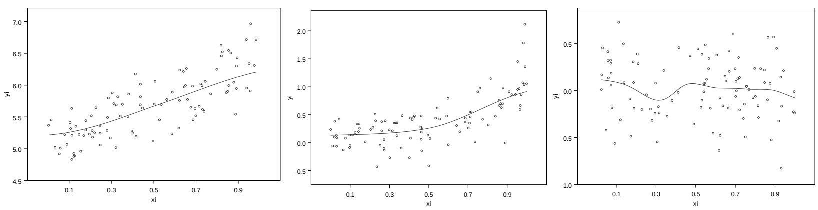
1. Linier: $m(x_i) = 5 + 1.5x_i$
2. Kubik: $m(x_i) = 0.1 + 0.2x_i + 0.3x_i^2 + 0.4x_i^3$

3. Logaritma natural: $m(x_i) = -0.005\ln(x_i)$

Fungsi pemulus yang digunakan adalah fungsi Kernel *Gaussian* karena fungsi tersebut memanfaatkan semua titik pengamatan dengan bobot yang berbeda-beda untuk memperoleh penduga $m(x_i)$. Bobot yang diberikan fungsi Kernel *Gaussian* semakin kecil jika titik pengamatan semakin jauh dari titik yang akan diduga. Persamaan matematis fungsi Kernel *Gaussian* adalah sebagai berikut:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2), -\infty < x < \infty \qquad (15)$$

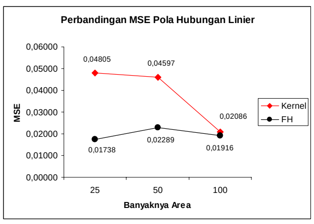
Lebar jendela yang digunakan pada penelitian ini mengikuti formula dari Hardle (1994), yaitu sebesar $m^{-1/5}$. Untuk jumlah area sebanyak 100, 50, dan 25, lebar jendela yang digunakan masing-masing adalah 0.40, 0.46, dan 0.53.

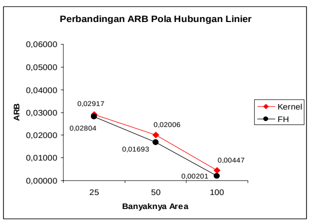


Linier Kubik Logaritma Natural

Proporsi keragaman area kecil terhadap keragaman total (γ_i) adalah sebesar 0.8 untuk sepertiga area pertama, 0.5 untuk sepertiga area kedua, dan 0.308 untuk sepertiga area terakhir. Nilai γ_i menunjukkan besarnya kontribusi yang diberikan oleh penduga langsung terhadap dugaan parameter area kecil, sedangkan 1-γ_i menunjukkan besarnya kontribusi penduga sintetik. Penduga sintetik dari pemulusan Kernel berupa dugaan dari pemulusan Kernel ($\hat{m}(x_i)$) dari persamaan Nadaraya-Watson Kernel. Sedangkan pada model Fay-Herriot, penduga sintetik diperoleh dari $x^T\hat{\beta}$ dimana β diduga dengan *weighted least square*.

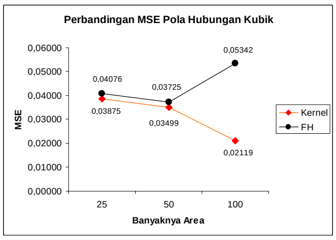
Perbandingan Metode Kernel dengan Metode Parametrik pada Pola Linier

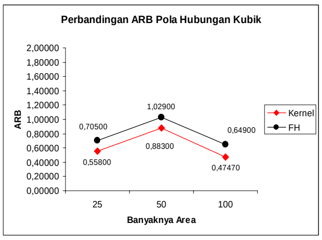




Pemulusan Kernel menghasilkan MSE dan ARB yang lebih besar daripada model Fay-Herriot. Selisih MSE pada m=100 sangat kecil, mengindikasikan bahwa pemulusan Kernel pada pola hubungan yang linier dengan jumlah area yang besar akan menghasilkan pendugaan yang hampir sama baiknya dengan model Fay-Herriot. Pada jumlah area yang semakin sedikit, model Fay-Herriot menghasilkan pendugaan yang jauh lebih baik dibandingkan pemulusan Kernel. Pendugaan dengan metode pemulusan Kernel menghasilkan MSE yang semakin kecil jika jumlah area semakin banyak.

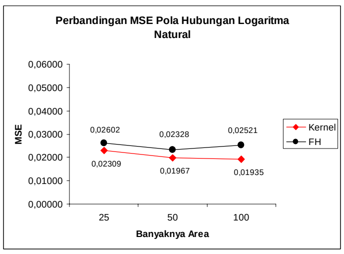
Perbandingan Metode Kernel dengan Metode Parametrik pada Pola Kubik

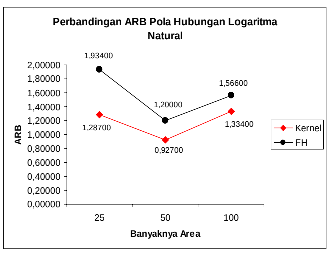




MSE dan ARB pemulusan Kernel lebih kecil daripada model Fay-Herriot. Pada jumlah area sebesar 25 dan 50 selisih MSE tidak terlalu jauh, namun pada saat jumlah area sebesar 100 selisih MSE keduanya menjadi sangat besar. Hal ini mungkin disebabkan karena pola hubungan kubik hampir menyerupai pola hubungan linier pada saat jumlah area relatif sedikit. Secara visual, pola hubungan kubik pada saat m=50 dan m=25 dapat dlihat pada Lampiran 1. Nilai MSE dari pemulusan Kernel semakin kecil jika jumlah area semakin besar.

Perbandingan Metode Kernel dengan Metode Parametrik pada Pola Logaritma Natural





Pemulusan Kernel menghasilkan MSE dan ARB yang lebih kecil dibandingkan model Fay-Herriot. MSE dari metode Kernel semakin kecil jika jumlah area semakin besar. Selisih MSE dari kedua metode hampir sama dan relatif cukup kecil pada semua jumlah area. Selisih ARB antara pemulusan Kernel dan model Fay-Herriot cukup besar pada berbagai jumlah area.

KESIMPULAN

Pemulusan Kernel sebagai salah satu pendekatan nonparametrik dapat diterapkan pada pendugaan area kecil. Pendugaan area kecil dengan menggunakan pemulusan Kernel lebih baik dibandingkan metode parametrik pada pola hubungan yang tidak linier, sedangkan pada pola hubungan yang linier, metode parametrik tetap lebih baik dibandingkan metode Kernel. MSE dari metode pemulusan Kernel cenderung semakin kecil jika jumlah area semakin banyak.

SARAN

Beberapa hal yang dapat dikaji lebih lanjut antara lain:

1. Peubah penjelas yang digunakan lebih dari satu.
2. Lebar jendela pada pemulusan Kernel dipilih dengan menggunakan metode-metode tertentu.
3. Menggunakan proporsi keragaman area kecil yang lebih beragam.
4. Ragam pengaruh acak area kecil (σ_u^2) dan ragam *sampling error* (D_i) diduga dari data.

DAFTAR PUSTAKA

Fay RE, Herriot RA. 1979. *Estimation of Income for Small Places: An Application of James-Stein Procedures to Census Data*. Journal of the American Statistical Associations: 269-277.

Hardle W. 1994. *Applied Nonparametric Regression*. http://www.quantlet.com. [25 April 2008].

Mukhopadhyay P, Maiti T. 2004. *Two Stage Non-Parametric Approach for Small Areas Estimation*. Proceedings of ASA Section on Survey Research Methods: 4058-4065.

Mukhopadhyay P, Maiti T. 2006. *Local Polynomial Regression for Small Area Estimation*. Proceedings of ASA Section on Survey Research Methods: 3447-3452.

Opsomer et al. 2004. *Nonparametric Small Area Estimation Using Penalized Spline Regression.* Proceedings of ASA Section on Survey Research Methods: 1- 8.

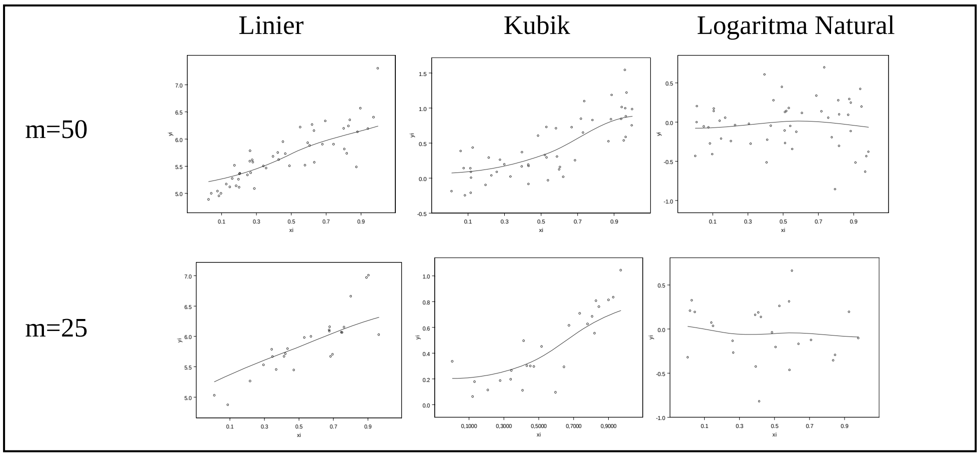
Rao JNK. 2003. *Small Area Estimation*. New Jersey: John Willey &Sons, Inc.

Silverman BW. 1986**.** *Density Estimation For Statistics and Data Analysis.* London: Chapman and Hall.

Wu H, Zhang JT. 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*. New Jersey: John Wiley & Sons, Inc.

LAMPIRAN

Lampiran 1   Kurva hasil pemulusan Kernel.



Lampiran 2   Tabel perbandingan MSE.

Banyaknya Area	Statistik	Linier		Kubik		Logaritma Natural	
		Kernel	FH	Kernel	FH	Kernel	FH
25	Rata-rata	0,04805	0,01738	0,03875	0,04076	0,02309	0,02602
	St. deviasi	0,03249	0,01371	0,00869	0,03353	0,00489	0,00652
50	Rata-rata	0,04597	0,02289	0,03499	0,03725	0,01967	0,02328
	St. deviasi	0,04805	0,01112	0,03119	0,03002	0,00933	0,01143
100	Rata-rata	0,02086	0,01916	0,02119	0,05342	0,01935	0,02521
	St. deviasi	0,01056	0,00867	0,01160	0,02684	0,00928	0,00990

Lampiran 3   Tabel perbandingan ARB.

Banyaknya Area	Statistik	Linier		Kubik		Logaritma Natural	
		Kernel	FH	Kernel	FH	Kernel	FH
25	Rata-rata	0,02917	0,02804	0,55800	0,70500	1,28700	1,93400
	St. deviasi	0,02168	0,01854	0,76900	0,80600	2,45400	3,82300
50	Rata-rata	0,02006	0,01693	0,88300	1,02900	0,92700	1,20000
	St. deviasi	0,01656	0,02025	1,02800	4,29600	1,32600	2,03000
100	Rata-rata	0,00447	0,00201	0,47470	0,64900	1,33400	1,56600
	St. deviasi	0,00415	0,00171	0,71100	1,13600	3,80900	1,91500