

Analisis Regresi Poisson untuk Menduga Hubungan Kelimpahan Makrobenthos dengan Parameter Kualitas Perairan

Dwi Haryo Ismunarti*, Subagyo, Ria Azizah TN

Jurusan Ilmu Kelautan Fakultas Perikanan dan Ilmu Kelautan Kampus Tembalang,
Universitas Diponegoro, Semarang, Indonesia

Abstrak

Fungsi peluang poisson merupakan standar model untuk variabel cacah. Sebagai contoh banyaknya makrobenthos dan faktor-faktor oseanografi yang mempengaruhinya dapat didekati dengan model regresi poisson. Model regresi poisson dapat diduga menggunakan fungsi GLM (generalized linear models) dari program S-PLUS. Faktor-faktor oseanografi yang berpengaruh terhadap jumlah makrobenthos adalah salinitas, kecerahan, kecepatan arus, DO dan suhu.

Kata kunci : peluang poisson, variabel cacah, makrobenthos

Abstract

The poisson probability provides the standard models for count variable. As an example, the number of makrobenthos and oceanography factors could be approach a poisson regression model. The poisson regression can be fitted using the function GLM, which fits Generalized Linear Models in S-PLUS program. Oceanography factors as having an effect on the number of makrobenthos are salinity, kecerahan, kecepatan arus, DO dan suhu.

Key words : poisson probability, count variable, makrobenthos

Pendahuluan

Pengalaman membangun sumberdaya pesisir dan laut selama kurun PJP I selain menghasilkan keberhasilan juga menimbulkan permasalahan ekologis dan sosial ekonomis. Secara ekologis banyak kawasan pesisir dan laut, antara lain Pantai Utara Jawa terancam kapasitas keberlanjutannya karena pencemaran, degradasi fisik habitat, over eksploitasi sumberdaya alam dan konflik penggunaan lahan (Bengen, 2002).

Timbulnya pencemaran pada suatu perairan dapat berakibat menurunnya kualitas air yaitu berubahnya sifat fisika dan kimia air. Kualitas perairan sungai dapat diketahui dengan metode kuantitatif menggunakan kelimpahan spesies sebagai indikatornya. Salah satu pendekatan dapat digunakan makrobenthos melalui monitoring jumlah populasi, komposisi komunitas maupun fungsi ekosistemnya. Organisme benthos merupakan jenis hewan yang hidup melekat atau relatif tidak bergerak dengan pola penyebaran yang khas. Struktur komunitasnya akan berubah karena perubahan lingkungan yang antara lain disebabkan oleh pencemaran perairan.

Keberadaan suatu individu E di luasan tertentu dapat dipandang sebagai kejadian 'ada' dan 'tidak ada'. Peluang E 'ada' adalah $p(E) = p$ sedangkan peluang E 'tidak ada' adalah komplement E' yaitu $q = p(E') = 1 - p$. Kelimpahan adalah jumlah individu E ada di setiap sampling unit (n_e). Jika keberadaan individu-individu di suatu luasan tertentu dapat dianggap independen dan peluang keberadaan individu $p(E)$ kecil maka fungsi peluang dari kelimpahan $p(x = n_e)$ yang merupakan variabel cacah akan mengikuti fungsi peluang kejadian Poisson (Casella & Berger, 1990, Bain & Engelhardt, 1992). Model regresi yang tepat untuk jenis data ini adalah model regresi poisson.

Jika peubah y adalah jumlah makrobentos pada pengamatan ke- i pada luasan A maka y_i merupakan nilai tengah dari y_i . Kemudian y_i dimodelkan sebagai fungsi dari k peubah bebas sehingga diperoleh $\mu_i = A \mu(x_i; \beta); i = 1, 2, \dots, k$ dengan i menyatakan pengamatan ke- i .

Fungsi dapat dipilih fungsi yang tidak negatif misalkan $e^{x/\beta} > 0$. Model regresi

Poisson dengan $e^{x|\beta} > 0$ adalah $\hat{\mu} = Ae^{x|\beta}$.

Nilai $\hat{\mu}$ diduga dengan menggunakan metode kemungkinan maksimum sebagai berikut :

$$L(y, \beta) = \prod_{i=1}^k p(y_i; \beta)$$

$$= \prod_{i=1}^k \frac{[A\mu(x_i, \beta)]^{y_i} e^{-[A\mu(x_i, \beta)]}}{y_i!}$$

$$= \frac{\prod_{i=1}^k [A\mu(x_i, \beta)]^{y_i} e^{-\sum_{i=1}^k A\mu(x_i, \beta)}}{\prod_{i=1}^k y_i!}$$

dan
$$\ln L(y, \beta) = \sum_{i=1}^k y_i \ln A\mu(x_i, \beta) - \sum_{i=1}^k A\mu(x_i, \beta) - \sum_{i=1}^k \ln y_i!$$

kemudian
$$\frac{\partial \ln L(y, \beta)}{\partial \beta} = 0$$

sehingga diperoleh hasil

$$\sum_{i=1}^k \left[\frac{y_i}{\mu(x_i, \hat{\beta})} - A \right] \left[\frac{\partial \mu(x_i, \hat{\beta})}{\partial \hat{\beta}} \right] = 0$$

Nilai penduga $\hat{\mu}$ diperoleh dari persamaan di atas dengan menggunakan metode *Poisson Iteratively Reweighted Least Squares (PIRLS)* (Peterson, 1997).

Pengujian terhadap penduga $\hat{\mu}$ digunakan statistik χ^2 dengan hipotesis

$H_0 : \mu_j = 0$ dan $H_1 : \mu_j \neq 0$;

$$\chi^2_{hitung} = \frac{b_j^2}{c_{jj}} \quad j = 0, 1, \dots, p$$

dengan p jumlah peubah bebas dan b_j penduga $\hat{\mu}$ untuk peubah bebas ke-j

$$c_{jj} = \frac{-\partial^2 \ln L(\hat{\beta})}{\partial \beta_j^2}, j = 0, 1, \dots, p$$

Materi dan Metode

Materi yang digunakan adalah hasil penelitian terhadap kualitas perairan Sungai Banjir Kanal Barat pada bulan Juli 2003. Lokasi penelitian terdapat pada lampiran 1. Variabel yang dipergunakan adalah variabel terikat : kelimpahan makrobenthos yang

ditemukan per stasiun per waktu pengamatan sedangkan variabel bebas : suhu, salinitas, kecerahan, kedalaman, kecepatan arus, nitrat dan DO (Ismunarti at all, 2003)

Model regresi poisson diturunkan dengan fungsi GLM (*Generalize Linear Models*) dalam program S-PLUS 2000 dengan mendefinisikan formula modelnya, fungsi peluang dan fungsi linknya.

Pengambilan keputusan dalam pemodelan statistika bergantung pada kevalidan dari model. Kehati-hatian dalam mengevaluasi data dengan model merupakan bagian terpenting dalam pemodelan data statistika. Empat hal yang memungkinkan ketidak-tepatan suatu pemodelan adalah: tidak adanya fungsi peluang yang spesifik dari variabel tak bebas y, tidak ada fungsi link yang spesifik, tidak ada bagian sistematis yang spesifik dari model dan adanya penyimpangan data observasi.

Evaluasi terhadap model digunakan sisaan (residual). Sisaan didefinisikan sebagai selisih aratan (data) dan nilai ramalan (dugaan) yang diperoleh dari

model
$$e_i = y_i - \hat{y}_i \quad i = 1, 2, \dots, n.$$
 Pemeriksaan

sisaan bermanfaat dan berlaku pada keadaan yang melibatkan pendugaan model seperti regresi linear maupun nonlinear dan model analisis ragam (Draper & Smith, 1992). Pemeriksaan terhadap sisaan melalui grafik dengan lebih seksama kemungkinan akan membawa ke arah model yang lebih baik atau metode pendugaan yang lebih sesuai. Keefektifan analisis data tergantung pada kegigihan dalam mendapatkan model yang relevan dan menyingkap hal-hal yang masih tersembunyi dalam sisaan.

Sisaan Pearson dapat digunakan untuk menguji kesesuaian model regresi poisson. Sisaan ini mendekati simpangan baku respon dengan nilai tengah nol dan varian mendekati satu (McCulagh &

Nelder, 1989). Sisaan Pearson adalah :
$$r_p = \frac{y - \mu}{\sqrt{V(\mu)}}$$

Jumlah kuadrat dari sisaan pearson adalah statistik

Chi-squared yaitu
$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$$

Dalam analisis statistika untuk mengevaluasi adanya penyimpangan data pengamatan ke-i sehingga berpengaruh terhadap pendugaan parameter digunakan beda antara pendugaan dengan

melibatkan pengamatan ke- i dan pendugaan dengan tanpa melibatkan pengamatan ke- i . Statistik tersebut

adalah *Cook's Distance* yaitu $D_i = \frac{h_i r_{pi}^2}{d(1-h_i)}$ h_i

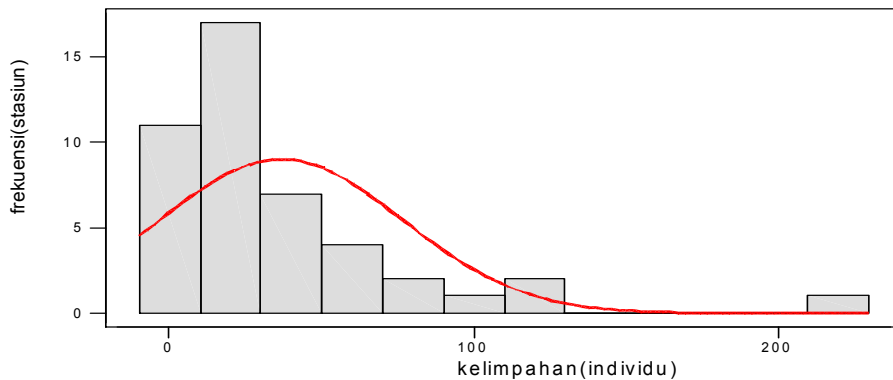
adalah diagonal *matriks hat*, r_{pi} sisaan pearson terbakukan dan d banyaknya variabel penjelas. Selanjutnya nilai D_i dibandingkan dengan nilai tabel $F(d, n-d, 1-a)$ untuk a yang ditentukan. D_i yang besar menandakan pengamatan ke- i sangat berpengaruh

(Draper & Smith, 1992). Plot D_i dengan observasi akan memudahkan dalam analisis.

Hasil dan Pembahasan

Hasil pengamatan menunjukkan kelimpahan populasi makrobenthos di Banjir Kanal Barat rendah dengan kisaran nilai 0 sampai 145 individu, rata-rata kelimpahan 24,56 individu dengan simpangan baku 37.59 individu. Nilai simpangan baku yang lebih besar dari rata-rata menunjukkan terjadi pengelompokan dari pola distribusi spasialnya.

Gambar 1. Histogram Kelimpahan makrobenthos



yang berarti ada beberapa stasiun pengamatan tidak ditemukan sama sekali makrobenthos. Populasi makrobenthos agak tinggi (28-108 individu) di sekitar muara kemudian menurun dan terendah (0-14 individu) di daerah sekitar jembatan Jalan Arteri. Fenomena yang menarik adalah populasi makrobenthos agak tinggi (17-118 individu) di sekitar daerah kawasan industri (Kelurahan Ngemplak Simongan).

Histogram merupakan salah satu metode untuk menduga pola sebaran. Gambar 1 adalah histogram kelimpahan makrobenthos pada setiap pengamatan menunjukkan ciri khas dari variabel cacah dengan kemencengan (*skewed*) negatif dan *tailed* yang panjang. Pola sebaran ini merupakan *typical* dari fungsi peluang Poisson (Anonim,1993).

Pendefinisian variabel dan model regresi Poisson dalam S-PLUS

```
> benthos<-kualitas[,"V13"]
> suhu<-kualitas[,"V2"]
```

```
> salinitas<-kualitas[,"V3"]
> kecerahan<-kualitas[,"V4"]
> kedalaman<-kualitas[,"V5"]
> arus<-kualitas[,"V6"]
> nitrat<-kualitas[,"V8"]
> DO<-kualitas[,"V10"]
> benthos.fit<glm(benthos~suhu+salinitas
+kecerahan+kedalaman+arus+nitrat+DO, family
=poisson (link=log), data=kualitas)
> summary(benthos.fit)
```

Fungsi *summary* akan menginformasikan model yang diperoleh, parameter yang diduga dan pengujian hipotesis terhadap parameter. Hipotesis yang diujikan adalah :

$H_0 : \beta_i = 0, i = 1, 2 \dots 6$ dengan hipotesis alternatif $H_1 : \beta_i \neq 0$. Statistik uji adalah t value. H_0 ditolak jika nilai t value lebih besar dari t tabel pada nilai \pm tertentu.

Tabel 1. Pendugaan Model 1

(Dispersion Parameter for Poisson family taken to be 1)
Null Deviance: 1475.6 pada 44 derajat begon
Residual Deviance: 871.8557 pada 37
Number of Fisher Scoring Iterations: 4

Nilai t value dari Tabel 1 menunjukkan bahwa variabel kedalaman dan nitrat tidak berperan besar dalam menentukan keberadaan makrobenthos di perairan. Faktor-faktor yang berpengaruh nyata terhadap peluang keberadaannya berturut-turut adalah salinitas, kecerahan, kecepatan arus, DO dan suhu. Faktor-faktor tersebut berpengaruh secara bersama-sama. Selanjutnya akan diturunkan model baru dengan tidak menyertakan kedua variabel tersebut. Untuk memperoleh model dugaan yang terbaik selanjutnya akan dilakukan pengujian parsial secara beruntun dengan menambahkan satu per satu variabel-variabel suhu, salinitas, kecerahan, arus dan DO ke dalam model. Pengujian dilakukan dengan fungsi anova.

```
> benthos.fit<glm(benthos~suhu+salinitas+
kecerahan+arus+DO, family=poisson(link=log),
data=kualitas)
> anova(benthos.fit, test="Chi")
> summary(benthos.fit)
```

Nilai-nilai statistik chi-square dari tabel 2 yang hampir mendekati nol menunjukkan penambahan variabel suhu sangat signifikan menurunkan sebesar 31.9523 keragaman variabel benthos. Hal ini berarti variabel suhu sangat diperlukan dalam model. Meskipun demikian model dengan hanya melibatkan satu variabel suhu belum merupakan model yang terbaik. Penambahan variabel salinitas, kecerahan, arus dan DO secara berturut-turut ke dalam model ternyata sangat signifikan dalam menurunkan simpangan/sisaan dengan demikian variabel-variabel tersebut sangat diperlukan dalam model dan variabel tersebut secara bersama-sama menentukan peluang keberadaan makrobenthos di perairan.

Tabel 2. Analisis Devian Poisson model]

Response: benthos
Terms added sequentially (first to last)

Tabel 3. Pendugaan Model 2

(Dispersion Parameter for Poisson family taken to be 1)
 Null Deviance: 1475.6 on 44 degrees of freedom
 Residual Deviance: 879.705 on 39 degrees of freedom
 Number of Fisher Scoring Iterations: 4

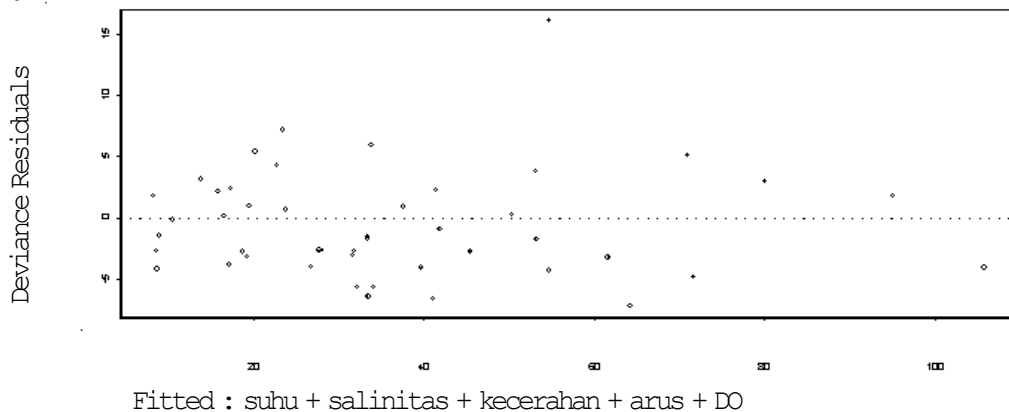
Tabel 3 menunjukkan penduga koefisien regresi dari model. Tanda negatif pada koefisien arus berarti semakin tinggi kecepatan arus semakin kecil peluang keberadaan makrobenthos ditemukan. Sedangkan tanda positif dari variabel suhu, salinitas, kecerahan dan DO menunjukkan semakin tinggi variabel-variabel tersebut semakin besar peluang keberadaan makrobenthos ditemukan pada kondisi perairan yang seperti ini.

Pemeriksaan terhadap sisaan (residual) sangat bermanfaat untuk mengevaluasi kebaikan model penduga. Sisaan yang akan digunakan dalam model regresi poisson adalah sisaan Pearson. Sisaan ini mendekati simpangan baku respon dengan nilai tengah nol dan varian mendekati satu (McCullagh & Nelder, 1989). Jumlah kuadrat dari sisaan pearson adalah statistik Chi-squared. Sisaan Pearson diperoleh dengan menggunakan fungsi **residuals** dalam S-PLUS. Grafik diperoleh menggunakan fungsi

win.graph dan **plot**.

```
> psisaan<-residuals(benthos.fit, type="pearson")
> win.graph
function(width = 10.5, height = 8,
  pointsize = 16, density,
  units.per.inch = 1)
val <- graphsheet (width = width,
  height = height, pointsize
  = pointsize,
  units.per.inch =
  units.per.inch)
par(pch = 16)
invisible(val)
> par(mfrow=c(4,1))
> plot(benthos.fit)
```

Gambar 2a plot sisaan dengan penduga dan 2b simpangan baku sisaan dengan penduga tidak menampakkan adanya pola kenaikan keragaman dengan demikian tidak perlu dilakukan transformasi terhadap variabel. Gambar sisaan terhadap normal baku terkadang sangat bermanfaat terutama pada model regresi linear yang mengasumsikan sebaran sisaannya normal. Tetapi dalam model regresi poisson karena tidak adanya asumsi sebaran normal maka plot ini kurang bisa diinterpretasikan.

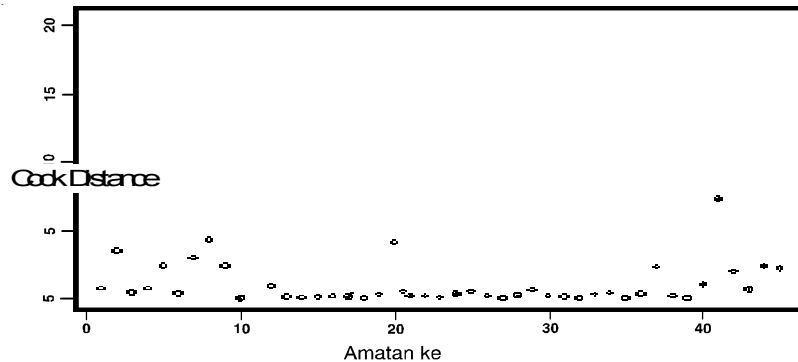


Pengaruh besar terhadap pendugaan parameter model, untuk itu perlu dilakukan evaluasi apakah pengamatan tersebut perlu dikeluarkan ataukah tidak sehingga diperoleh model penduga yang terbaik.

Statistik Cook's Distance dihitung dengan memanfaatkan fungsi **lm.influence**

```
> Z s i s a a n <- p s i s a a n / s q r t ( 1
  lm.influence (benthos.fit) $hat)
```

```
> benthos.hat<-lm.influence (benthos.fit) $hat
> D<-benthos.hat*Zsisaan*Zsisaan/ (3*(1-
  benthos.hat))
> plot (1:45,D,xlab="sampling ke",ylab="Cook's
  Distance")
> title ("Plot Cook's Distance")
> D
```



Gambar 3. Plot Statistik Cook's Distance

Plot gambar 3 menunjukkan terdapat 2 pengamatan yang ganjil yang jauh dari tipikal data lainnya. Pengamatan tersebut dengan nilai $D_{11} = 20.4916$ dan nilai $D_{41} = 7.293822$. Nilai-nilai tersebut lebih besar dari $F(5, 40, 95\%) = 4.46$. Tidak adanya

hal yang khusus dari kedua pengamatan untuk selanjutnya akan dilakukan pemodelan kembali dengan tidak menyertakan kedua amatan tersebut.

Hasil pada tabel di bawah menunjukkan sisaan Deviance $527.1748/37=14.287$ dari $879.705/39=22.559$

Tabel. 4 Analisis Model 3

(Dispersion Parameter for Poisson family taken to be 1)
 Null Deviance: 926.2559 on 42 degrees of freedom
 Residual Deviance: 527.1748 on 37 degrees of freedom
 Number of Fisher Scoring Iterations: 4

Tabel 5. Analysis of Deviance Table Poisson model Terms added sequentially (first to last)

Kesimpulan dan Saran

Hasil pengamatan menunjukkan kelimpahan makrobenthos di Banjir Kanal Barat rendah dengan keragaman tinggi. Faktor-faktor yang berpengaruh nyata terhadap peluang keberadaannya berturut-turut adalah salinitas, kecerahan, kecepatan arus, DO dan suhu. Faktor-faktor tersebut berpengaruh secara bersama-sama. Model terbaik diperoleh $\log(m) = -14.33 + 0.46 \text{ Suhu} + 0.038 \text{ salinitas} + 0.053 \text{ kecerahan} - 0.215 \text{ kec. arus} + 0.52 \text{ DO}$

Hal yang paling mendasar dalam pendugaan model adalah menelusuri fungsi peluang yang spesifik dari variabel tak bebas y. Tidak terpenuhinya asumsi-asumsi pada metode pendugaan optimum model akan mengakibatkan ketidaktepatan pendugaan dan model yang diperoleh tidak dapat diandalkan. Selain itu keefektifan analisis data juga tergantung pada kegigihan dalam mendapatkan model yang relevan dan menyingkap hal-hal yang masih tersembunyi dalam sisaan. Penelusuran pola sisaan dan cook's distance bermanfaat untuk mendapatkan model terbaik.

Ucapan Terima kasih

Ucapan terima kasih kepada Proyek Peningkatan Penelitian Pendidikan Tinggi Direktorat Jendral Pendidikan Tinggi, Departemen Pendidikan Nasional, Sesuai dengan Surat Perjanjian Pelaksanaan Penelitian Dosen Muda Nomor : 028/P4T/DPPM/PDM/III/2003 tanggal 28 Maret 2003 yang telah membiayai penelitian ini. Kepada mahasiswa Jur. IK FPIK UNDIP sdr. Widodo, Rudiono, Rahmat dan Edy Yusuf Hamzah atas kerja samanya.

Daftar Pustaka

Aitkin M., Anderson D., Francis B. and Hinde J. 1990. *Statistical Modelling in GIM*. Oxford Science Publishing, New York

Anonim. 1993. *S-PLUS Guide to Statistical and Mathematical Analysis*. Stat Sci Adivision Of

MathSoft.Inc. Washington.

Aunuddin. 1989. *Analisis Data*. PAU Ilmu Hayat Institut Pertanian Bogor, Bogor.

Bain, L.J. and M.Engelhardt. 1992. *Introduction to Probability and Mathematical Statistics*. Duxbury Press Belmont, Calofomia.

Bengen, D.G. 2002. *Konsep Pengelolaan Wilayah Pesisir Secara Terpadu dan Berkelanjutan*. Makalah Pelatihan Pengelolaan Sumberdaya Wilayah Pesisir dan Lautan Semarang, 4-8 Maret 2002

Casella, G and R.L.Berger. 1990. *Statistical Inference*. Wadsworth & Brooks/Cole Publishing Company, California

Draper N. and Smith H.1992. *Analisis Regresi Terapan*. Ed. Kedua Terjemahan Bambang Sumantri. PT Gramedia, Jakarta.

Everitt B. 1994. *A Handbook of Statistical Analyses using S-PLUS*. Chapman & Hall , London.

McCullagh, P. and J.A.Nelder. 1989. *Generalized Linear Model*. Chapman & Hall. London

Myers, R.H. 1990. *Clasical and Modern Regression with Application*. PWS-Kent Publishing Comp. Boston

Nasution, A.H. dan A.Rambe. 1984. *Teori Statistika*. Ed. Ke-2. Ehratara, Jakarta.

Peterson, L.E. 1997. *PIRLS: Poisson Iteratively Reweighted Least Squares Computer Program for Aditive, Multiplicative, Power, and Non-Linear Models*. Center for Cancer Control Research Baylor College of Medicine Houston, Texas.

Purnomo, K. 1989. *Struktur Komunitas Makrozoobentos dalam Kaitan Pemantauan Dampak Aktifitas Manusia di Daerah Sungai Cikao Kab. Purwakarta Jawa Barat*. Tesis. Pasca Sarjana Institut Pertanian Bogor, Bogor.

Retnaningdyah, C. 1997. *Kepekaan Makroinver-*

tebrata Bentos Terhadap tingkat Pencemaran Deterjen di Kali Mas Surabaya. Tesis. Program Pascasarjana Universitas Indonesia, Jakarta.

Rosenberg, D.M. and Resh V.H. 1993. *Fresh Water Bionitoring and Benthic Macroinvertebrates.* Chapman & Hall Inc. New York.