
IDENTIFIKASI MODEL $GSTARX-(p,q)$ DENGAN PENGOLAHAN DATA

MENGGUNAKAN *SOFTWARE R*

(Kajian Beban Pencemaran *BOD* di tiga Lokasi Titik Pantau DAS Citarum)

Soekardi Hadi Prabowo

Prodi Matematika, Fakultas Sains dan Teknologi

Universitas Islam As-syafi'iyah, Jakarta

s.hadip@yahoo.co.id

ABSTRACT. *This paper examines the identification phase of the Generalized Space time Autoregressive with exogenous X-Variable ($GSTARX-(p,q)$) model by processing data using R software. The identification process includes stationary time series data, time and autoregressive time lag. The $GSTARX-(p, q)$ model is a special multivariate time series model of vector autoregressive (VAR) which involves an exogenous (X) variable and a location weighting matrix. The results of stationary test were visually using software R, based on the actual data plot of the level of Biological Oxygen Demand (BOD) pollution load - each location of the monitoring point location was thought to show the time series vector distribution pattern simultaneously meeting stationary assumptions. This is confirmed visually through the autocorrelation function (ACF) plot which shows a significantly fast and formally decreasing pattern using the ADF test, for estimating the error value of the model using Ordinary least Square (OLS), indicating there is no root unit. Whereas in the autoregressive order determination visually for each location through a Partial ACF plot, the data from the three locations shows a cut off on the lag time 1, while the location lag order is chosen 1, because the location is in the Citarum watershed same. Then concluded the selected temporary model is $GSTARX-(1,1)$.*

Keywords: *Model Identification, $GSTARX-(p, q)$, ACF, PACF, ADF*

ABSTRAK. *Paper ini mengkaji tahap identifikasi model Generalized Space time Autoregressive with exogenous X-Variable ($GSTARX-(p,q)$) dengan pengolahan data menggunakan software R. Proses identifikasi meliputi kestasioneran data time series, dari waktu kewaktu dan penentuan orde lag time autoregressive. Model $GSTARX-(p,q)$ merupakan model time series multivariate khusus dari vector autoregressive (VAR) yang melibatkan variabel eksogen (X) dan matriks bobot lokasi. Hasil uji kestasioneran secara visual menggunakan software R, berdasarkan plot data aktual kadar beban pencemaran Biological Oxygen Demand (BOD) masing - masing lokasi lokasi titik pantau diduga menunjukkan pola sebaran vector data time series secara simultan memenuhi asumsi stasioner. Hal ini, dipertegas secara visual melalui plot autocorrelation function (ACF) yang menunjukkan pola menurun cepat secara signifikan dan secara formal dengan menggunakan uji ADF, untuk nilai error model hasil estimasi menggunakan Ordinary least Square (OLS), menunjukkan tidak terdapat unit root. Sedangkan pada penentuan orde autoregressive secara visual untuk masing-masing lokasi melalui plot Partial ACF, dari data ketiga lokasi terlihat pola grafik terputus (cut off) pada lag time 1, Sedangkan orde lag lokasi dipilih 1, karena lokasi penelitian berada pada bagian DAS Citarum yang sama. Maka disimpulkan model sementara yang terpilih adalah $GSTARX-(1,1)$.*

Kata kunci: *Identifikasi Model, $GSTARX-(p,q)$, ACF, PACF, ADF*

1. PENDAHULUAN

Identifikasi model merupakan langkah pertama yang perlu dilakukan dalam membangun model *space time*, salah satu tahap dari empat tahapan prosedur *ARIMA Box-jenkins* yang terdiri atas Identifikasi, Estimasi, Cek diagnosa dan Prediksi atau peramalan pada pemeriksaan perilaku data *time series univariat* yang diadopsi ke dalam data *space time*. (Box, G.E.P., et.al, 2008). *Space Time* adalah salah satu model yang menggabungkan unsur dependensi waktu dan lokasi pada suatu data *time series* multivariat. Pertamakali diperkenalkan oleh Pfeifer dan Deutsch (1980) untuk karakteristik lokasi homogen dalam bentuk model *Space Time Autoregressive orde p (STAR(p))*. Kemudian dikembangkan oleh Ruchjana B.N., dkk (2002) menjadi *Generalized STAR* yang dapat diterapkan pada data yang memiliki karakteristik lokasi heterogen. perluasan dari model $GSTAR(p,q)$ dengan memasukkan variabel eksogen X , yang diharapkan dapat meningkatkan akurasi pemodelan dan peramalan disebut model $GSTARX(p,q)$.

Penelitian menggunakan model $GSTARX$ dalam berbagai kasus telah dilakukan antara lain oleh Kurnia, J.D (2015), Ditago (2015), Mubarak (2015) dan Astuti, D (2016). Fokus kajian pada makalah adalah penelitian identifikasi model $GSTARX(p,q)$, dengan. Bentuk umum dituliskan sebagai berikut :

$$Z(t) = \sum_{k=1}^p \left(\sum_{l=0}^{\lambda_l} \Phi W^{(l)} Z(t-k) + \tau X(t) + e(t) \right) \quad (1)$$

(Astuti, D, 2016 dalam P. Hadi Soekardi, 2017) dengan

$Z(t), Z(t-k)$: Vektor acak berukuran $(N \times (T-1) \times 1)$

Φ dan τ : Matriks parameter dari model $GSTARX$, berukuran $(N \times N)$

$e(t)$: Vektor *error* \sim iid $N(0, \sigma^2)$. dan berukuran $(N \times (T-1) \times 1)$

$X(t)$: Vektor variabel eksogen berukuran $N(T-1) \times N(T-1)$

$W^{(l)}$: Matriks bobot spasial, berukuran $N(T-1) \times N(T-1)$ untuk orde spasial p dengan elemen diagonal bernilai nol dan jumlah dari tiap elemen pada tiap baris sama dengan satu.

Penelitian menggunakan model *GSTARX* dalam berbagai kasus telah dilakukan antara lain oleh Kurnia, J.D. (2015), Ditago (2015), Mubarak (2015) dan Astuti, D (2016).

2. METODE PENELITIAN

Data yang digunakan sebagai dasar pembentukan model pada persamaan (1) merupakan data sekunder kadar beban cemaran *Biological Oxygen Demannnd* (BOD), ditambah *Chemical Oxygen Demannnd* (COD), diambil dari bulan Januari 2010 sampai dengan Juni 2014 di tiga lokasi titik pantau Inlet dan Outlet Jatilur serta Rengasdengklok, semuanya berada di bagian hilir DAS Citarum. (www.perumjasatirta2@go.id).

Analisis data untuk identifikasi model pada persamaan (1), terdiri atas pemeriksaan asumsi kestasioneran data *time series* dan penentuan orde model *lag time p* serta *lag spasial q*.

2.1 Tahap Pendeteksian Kestasioneran Data *Time Series*,

Menurut Wei (2006), tahap pendeteksian data *time series* dilakukan dengan langkah-langkah sebagai berikut:

Langkah pertama, menganalisis grafik secara visual dengan membuat plot antara nilai observasi dengan waktu dan transformasi yang sesuai. Pola data stasioner merupakan syarat yang harus dipenuhi pada pemodelan data *time series*, karena data yang memiliki pola stasioner pada suatu model *space time* dapat memberikan hasil prediksi lebih akurat. Data dikatakan stasioner bila pola data menunjukkan keteraturan terhadap *mean* dan *variance* atau diperkirakan memiliki *mean* konstan dan *variance* dalam plot menunjukkan fluktuasi data yang konstan dari waktu ke waktu.

Langkah Kedua, menggunakan uji *Correlogram*, melalui visual pola plot *Autocorrelation Function* (ACF) berdasarkan indikator nilai koefisien ACF secara bertahap satu per satu untuk data di masing-masing lokasi titik pantau. Hal ini dapat dilakukan karena model *GSTARX* sebagai perluasan dari model *GSTAR* dan pengembangan model STAR. Sedangkan model STAR merupakan

bentuk khusus dari model *time series* multivariat *vector Autoregressive* (VAR) yang diperkenalkan oleh Sims, C.A. pada tahun 1980. (Lüthkepol, H., 2007). Karena identifikasi pada model VAR dapat dilakukan baik secara simultan maupun satu persatu untuk data di setiap lokasi titik pantau. (Rosadi Dedi, 2012 dalam P Hadi Soekardi, 2017).

Jika Plot *ACF* turun signifikan secara cepat atau menyusut secara eksponensial (*tail off*) mengindikasikan data stasioner dalam *mean* untuk data di tiap lokasi titik pantau, maka untuk vektor data *time series* secara simultan disimpulkan stasioner. Namun Jika Plot *ACF* turun secara lambat mengindikasikan data tidak stasioner.

Langkah Ketiga, secara formal dapat dideteksi dengan menggunakan uji akar unit *Augmented Dickey-Fuler* (ADF) berdasarkan hipotesis berikut ini,

H_0 : Data dari variabel model tidak stasioner (terdapat unit root)

H_1 : Data dari variabel model stasioner (tidak terdapat unit root)

Dengan statistik uji yang digunakan untuk menguji parameter model *GSTARX*($p,1$). untuk *lag time* p yang terpilih pada plot *ACF*, sedangkan *lag spasial* (lokasi) q diasumsikan sama dengan 1, karena data berasal dari ketiga lokasi yang berada dalam bagian Daerah Aliran Sungai yang sama, secara simultan adalah,

$$F_{hitung} = \frac{\frac{SS_e - SS_u}{r}}{\frac{SS_e}{n - s}} \quad (2)$$

Dalam hal ini,

SS_e : Jumlah kuadrat error dari model yang dibatasi

SS_u : Jumlah kuadrat error dari model yang tanpa dibatasi

n : Banyak observasi

s : Banyak parameter yang diestimasi dalam model tanpa dibatasi

Sedangkan estimasi parameter model *GSTARX*($p,1$) dilakukan dengan menggunakan metode *ordinary least square* (OLS). untuk kriteria pengujian dari ADF dilakukan melalui perbandingan konsultasi antara nilai F_{hitung} pada persamaan (2) dengan nilai kritis F_{Tabel} untuk derajat kebebasan pembilang = r

dan derajat kebebasan penyebut = $n - k$ serta taraf signifikan α (biasanya dipilih 0,05), maka keputusan diambil berdasarkan ketentuan, bila ternyata:

°) $F_{Hitung} < F_{Tabel}$ terima H_o , disimpulkan data variabel model tidak stasioner

°) $F_{Hitung} \geq F_{Tabel}$ tolak H_o .disimpulkan data variabel model stasioner

Untuk mengatasi data yang tidak stasioner dalam mea dapat dilakukan proses pemeddaan (*differencing*) terhadap time series aktual, $Z_a(t) = Z(t) - Z(t - 1)$ dan yang tidak stasioner dalam *variance* maka dilakukan tramsformasi *Box-Cox* pada data

dengan formula $Z_{Transf} = \frac{Z^\lambda - 1}{\lambda}$ aktual, pada umumnya penentuan nilai λ menggunakan lon (*logaritma Natural*) atau akar kuadrat pada data aktual. (Juanda Bambang, 2012).

2.2 Tahap Penentuan Orde Model

Sebagai dasar penentuan orde model dapat digunakan plot fungsi *Autokorelasi Parsial (PACF)*, yang juga dilakukan secara individual untuk masing-masing data di tiap lokasi titik pantau. Dengan ketentuan jika pada plot *PACF* untuk lag p terlihat garis yang terputus dengan lainnya (*cut off*), hal ini menunjukkan bahwa *orde Autoregressive (AR)* model untuk tiap lokasi yang sesuai adalah p . sedangkan untuk lag spasial dipilih 1, karena ketiga data *time series* berad dalam bagian Daerah Aliran Sungai Citarum yang sama. (P. Hadi Soekarrdi, 2018).

Seluruh proses pengolahan data baik plot dan uji *ADF* kestasioneran data, proses *differencing*, transformasi *Box-Cox* dan plot *PACF* semuanya menggunakan *software R* yang bersifat gratis, efisien dan praktis serta tidak memerlukan lisensi bagi pengguna, sehingga kita terhindar dari perbuatan tidak baik, dengan melakukan pembajakan. Suatu tindakan yang melanggar etika profesi (Rosadi Dedi, 2015 dalam P. Hadi Soekarrdi dan Hardya, 2018)

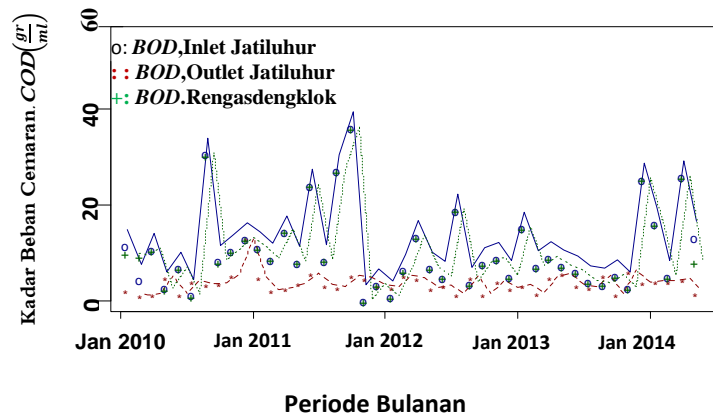
3. HASIL DAN PEMBAHASAN

Berdasarkan hasil pengolahan data plot data aktual menggunakan *software R* melalui *script* sebagai berikut :

```
# Input data kadar beban pencemaran BOD di tiga lokasi #
```

```
BOD.InletJatiluhur<-c(22.56,9.43,... ,10.67,2.23,18.68)
BOD.OutletJatiluhur<-c(31.56,19.13,... ,10.67,2.23,17.88)
BOD.RengasDengklok<-c(9.56,9.18,... ,10.67,2.23,19.78)
Data<-cbind(BOD.InletJatiluhur, BOD.OutletJatiluhur,
BOD.RengasDengklok)
Kadar.BOD<-ts(Data,start=c(2010,1),frequency=12)
Kadar.BOD
plot(Kadar.BOD)
```

Hasil pengolahan disajikan dalam Gambar.1 sebagai berikut :

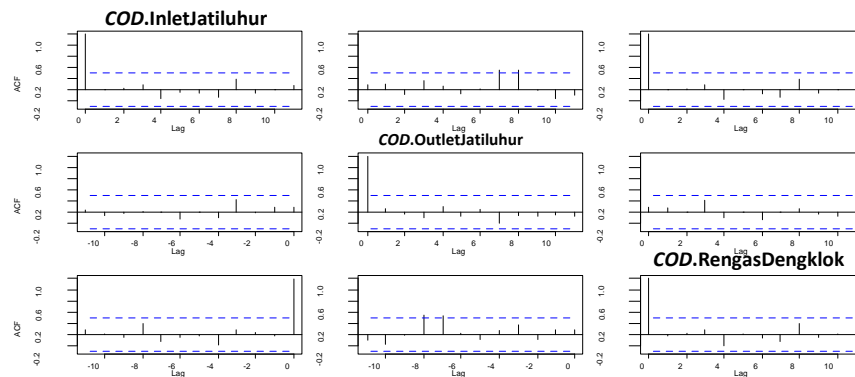


:

Sumber : Hasil pengolahan *software R*

Gambar 1. Plot Data Aktual Kadar Beban Pencemaran $BOD \left(\frac{\text{gram}}{\text{liter}} \right)$ di Tiga Titik Pantau DAS Sungai Citarum Periode Januari 2010-Agustus 2014

Berdasarkan Gambar 1 terlihat plot data aktual kadar beban pencemaran BOD pada tiga titik pantau di Sungai Citarum, menunjukkan kecenderungan fluktuasi data dengan *mean* dan *variance* konstan dari waktu ke waktu, hal ini mengindikasikan data *time series* multivariat beban pencemaran BOD secara simultan stasioner. Selanjutnya, dari hasil pengolahan dengan *software R* untuk plot *ACF* melalui *script* $\rightarrow \text{acf}(\text{Data})$, hasilnya disajikan dalam Gambar.2 sebagai berikut :



Sumber : Hasil pengolahan *softwar R*

Gambar 2. Plot *ACF* Kadar Beban Pencemaran COD ($\frac{\text{gram}}{\text{liter}}$) di Titik Pantau DAS Sungai Citarum Periode Januari 2010-Agustus 2014

Berdasarkan Gambar 2 terlihat plot *ACF* data asli untuk ketiga titik pantau di lokasi Inlet dan Outlet Jatiluhur serta Rengasdengklok, secara berturut-turut pada arah diagonal utama dari kiri ke kanan, menunjukkan pola menurun secara signifikan (*tail off*), hal ini mempertegas kestasioneran data *time series* multivariat kadar beban cemaran *BOD*. Selanjutna Dipertegas pula secara formal mengenai kestasioneran data dalam *mean* dan *variance* secara simultan, dengan menggunakan statistik *Augmented Dickey Fuller (ADF)* pada data *error* dari estimator *OLS* untuk uji *unit root*. Hasil perhitungan dengan *Software R*, melalui *script* sebagai berikut :

```

Input Data in smple sebanyak 44 datum (80 % dari total data)
Z1.1<-BOD.InletJatiluhur[-1]Z1.43<-BOD.InletJatiluhur[-44]
Z2.1<-BOD.OutletJatiluhur[-1]Z2.43<-BOD.OutletJatiluhur[-44]
Z3.1<-BOD.Rengasdengklok[-1]Z3.43<-BOD.Rengasdengklok[-44]
V1.43<-0.5*(Z2.43+Z3.43)X1.1<-COD.InletJatiluhur[-1]
V2.43<-0.5*(Z1.43+Z3.43)X2.1<-COD.OutletJatiluhur[-1]
V3.43<-0.5*(Z1.43+Z2.43)X3.1<-COD.Rengasdengklok[-1]
Y<-matrix(c(Z1.1,Z2.1,Z3.1),ncol=1)
XS<-
matrix(c(Z1.43,NOL.43,NOL.43,NOL.43,Z2.43,NOL.43,NOL.43,NOL.43,Z3.
43,V1.43,NOL.43,NOL.43,NOL.43,V2.43,NOL.43,NOL.43,NOL.43,V3.43,X1.
1,NOL.43,NOL.43,NOL.43,X2.1,NOL.43,NOL.43,NOL.43,X3.1),ncol=9)
XR<-
matrix(c(Z1.43,NOL.43,NOL.43,NOL.43,Z2.43,NOL.43,NOL.43,NOL.43,Z3.

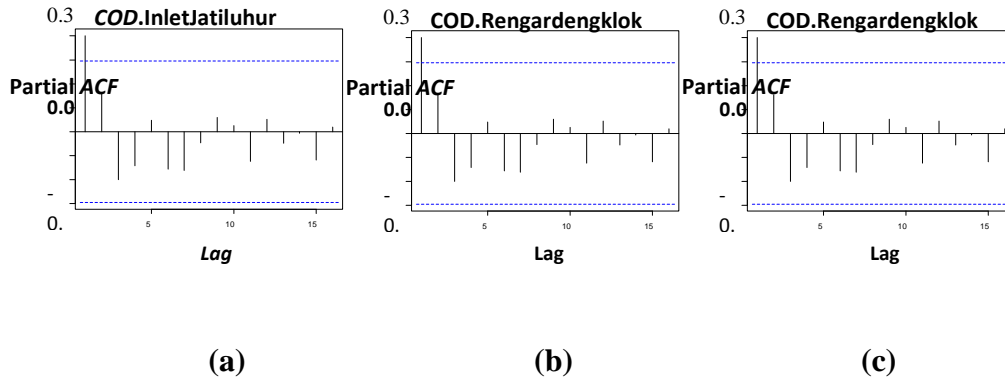
```

```

43,V1.43,NOL.43,NOL.43,NOL.43,V2.43,NOL.43,NOL.43,NOL.43,V3.43,X1.
1,NOL.43,NOL.43,NOL.43,X2.1,NOL.43,NOL.43,NOL.43,X3.1),ncol=9)
KompS1<-t(XS)%*%XS
KompS2<-t(XS)%*%Y
estgstrOLSu<-solve(KompS1)%*%KompS2
estgstrOLSu
KompR1<-t(XR)%*%XR
KompR2<-t(XR)%*%Y
estgstrOLSe<-solve(KompR1)%*%KompR2
estgstrOLSe
n<-length(Y)r<-6s<-9
Z.estu<-
0.15582*Z1.43+0.12423*Z2.43+0.40208*Z3.43+0.635598*V1.43+0.17213*V2.
43+0.37893*V3.43++0.04516*X1.1+1.0082*X2.1+0.04846*X3.1
etu<-Y-Z.estuSSu<-t(etu)%*%etu
Z.este<-
0.75863*V1.43+0.30266*V2.43+0.68859*V3.43+0.05195*X1.1+1.99902*X2.1+
0.06923*X3.1
ete<-Y-Z.esteSSe<-t(ete)%*%ete
PMB<-(SSe-SSu)/r
PMB
[1,] 8042.194
PNYBT<-SSe/(n-s)
PNYBT
[1,] 1026.630
Fh<-PMB/PNYBT
Fh [1,] 7.833582 #FHitung#
F.Tabel<-qf(0.95,6,34)
F.Tabel
[1] 2.380313

```

Berdasarkan hasil pengolahan di atas, diperoleh nilai F_{Hitung} sebesar 7,8335. Ternyata nilai F_{Hitung} tersebut lebih besar bila dibandingkan atau dikonsultasikan dengan nilai $F_{Tabel} = 2,380$. Dengan demikian disimpulkan tidak ada *unit root* dalam model *GSTARX*(*p*,1), sehingga asumsi kestasioneran data secara simultan dipenuhi. Kemudian penentuan orde *lag time* *p* dilakukan dengan menggunakan plot *PACF*, melalui *script* pada *software R* `pacf(ZBOD.lokasi-i)` untuk masing-masing lokasi 1-Inlet Jatiluhur, 2-Outlet Jatiluhur dan 3-Rengasdengklok. Hasilnya disajikan pada Gambar.3 sebagai berikut :



Sumber : Hasil pengolahan *softwar R*

Gambar 3. Plot PACF di Tiga Lokasi Titik Pantau inlet Jatiluhur (a) Outlet Jatiluhur(b) dan Rengasdengklok (c)

Berdasarkan Gambar 3(a), 3(b), dan 3(c), terlihat karakteristik plot *PACF* terputus (*cut off*) hanya pada *lag* waktu $p = 1$, hal ini menunjukkan orde *lag* waktu model $GSTARX(p, l)$ yang terpilih adalah 1. Selanjutnya tidak perlu dilakukan pemilihan orde *lag* waktu yang optimal berdasarkan kriteria minimum *Alkaike's Information Criterium (AIC)*. Kemudian karena telah diasumsikan bahwa orde *lag* spasial model $GSTARX$ adalah orde 1, serta data *time series* multivariat yang digunakan dalam penelitian ini memenuhi asumsi stasioner, maka dengan demikian model *space time* yang terbentuk adalah $GSTARX(1, 1)$, dengan bentuk persamaan sebagai berikut :

$$Z(t) = [\Phi_{10} + \Phi_{11}W^{(l)}] Z(t-1) + \gamma X(t) + e(t) \quad (1)$$

Dalam hal ini $Z(t)$, $Z(t-1)$, $X(t)$ dan $e(t)$ merupakan vektor-vektor berukuran $((\times (T - 1)) \times 1)$, sedangkan Φ , W dan γ matriks berukuran (3×3) .

4. KESIMPULAN DAN SARAN

Berdasarkan hasil pembahasan yang telah dilakukan, maka dapat diambil kesimpulan sebagai berikut :

- 1) Asumsi kestasioneran data baik untuk pendeteksian plot data aktual secara visual maupun uji formal semuanya menunjukkan hasil dipenuhi.

- 2) Orde *lag time* yang terpilih adalah 1
- 3) Model *space time* yang terpilih dengan mengasumsikan orde *lag spasial* 1 adalah *GSTAX*(1,1).

DAFTAR PUSTAKA

- Astuti, D, *Penerapan Model Generalized Soace Time Autoregressive with X - Exogen (GSTARX) pada peramalan volume ekspor CPO*, Tesis, Program Magister Statistika, Universitas Padjadjaran Bandung, 2015.
- Juanda B., *Ekonometrika Deret waktu*, IPB Press, 2012.
- Kurnia, J. D., *The Simulation Studies for Generalized Space Time Autoregressive (GSTARX) Model*, SSE FSM SWCU, 2015.
- Lütkepohl, H., *New Introduction to Multiple Time Series Analysis*, Springer, Berlin Heidelberg, New York, 2007.
- Soekardi, H. P., *Pemodelan GSTARX -(1,1) untuk Peramalan Kadar Beban Cemar BOD dengan Variabel Eksogem COD di Daerah Aliran Sungai Citarum*, Tesis, Program Magister Matematika Universitas Padjadjaran Bandung, 2017.
- Soekardi, H. P. dan Hardyaningwati, *Pemodelan Vectyor Autoregressive pada peramalan Kadar Beban Cemar COD dengan Variabel Eksogem COD di Daerah Aliran Sungai Citarum menggunakan Software R*, Jurnal - elektronik Widya Sains, **1** (2018).
- Ruchjana B.N., *Suatu Model Generalized Space Time Autoregressive dan Penerapannya Pada Produksi Minyak Bumi*, Disertasi, Program Doktor Matematika, Institut Teknologi Bandung, 2002.
- Wei, W.W.S., *Time Series Analysis: Univariate and Multivariate method Canada* Addison-Wesley Publishing Co., 2006.