SEMINÁRIO

Previsão de Séries Temporais com Modelagem ARIMA (Box-Jenkins)

Fábio Daros de Freitas

Departamento de Engenharia Elétrica/PPGEE – UFES

Terça-feira,1 de abril de 2003 - 16:00h

Local: Auditório do CT - UFES

Informações: Prof. Alberto F. Souza- (27) 3335 2641

Previsão de Séries Temporais com Modelagem ARIMA (Box-Jenkins)

Fábio Daros de Freitas

Departamento de Informática - Departamento de Engenharia Elétrica/PPGEE – UFES

Visão Geral

- Objetivos
- Séries Temporais
- Modelo ARIMA (Box-Jenkins)
- Metodologia Box-Jenkins
- Ferramenta TSTOOL/ARIMA
- Conclusão
- Discussão

Objetivos do Seminário

- Introdução ao modelo ARIMA e à metodologia Box-Jenkins para previsão de séries temporais.
- Apresentar uma implementação de um ambiente para modelagem e previsão ARIMA.
- Modelos sazonais não serão abordados.

Séries Temporais

- Séries Temporais
- Séries Temporais e Processos Estocásticos
- Análise de Séries Temporais

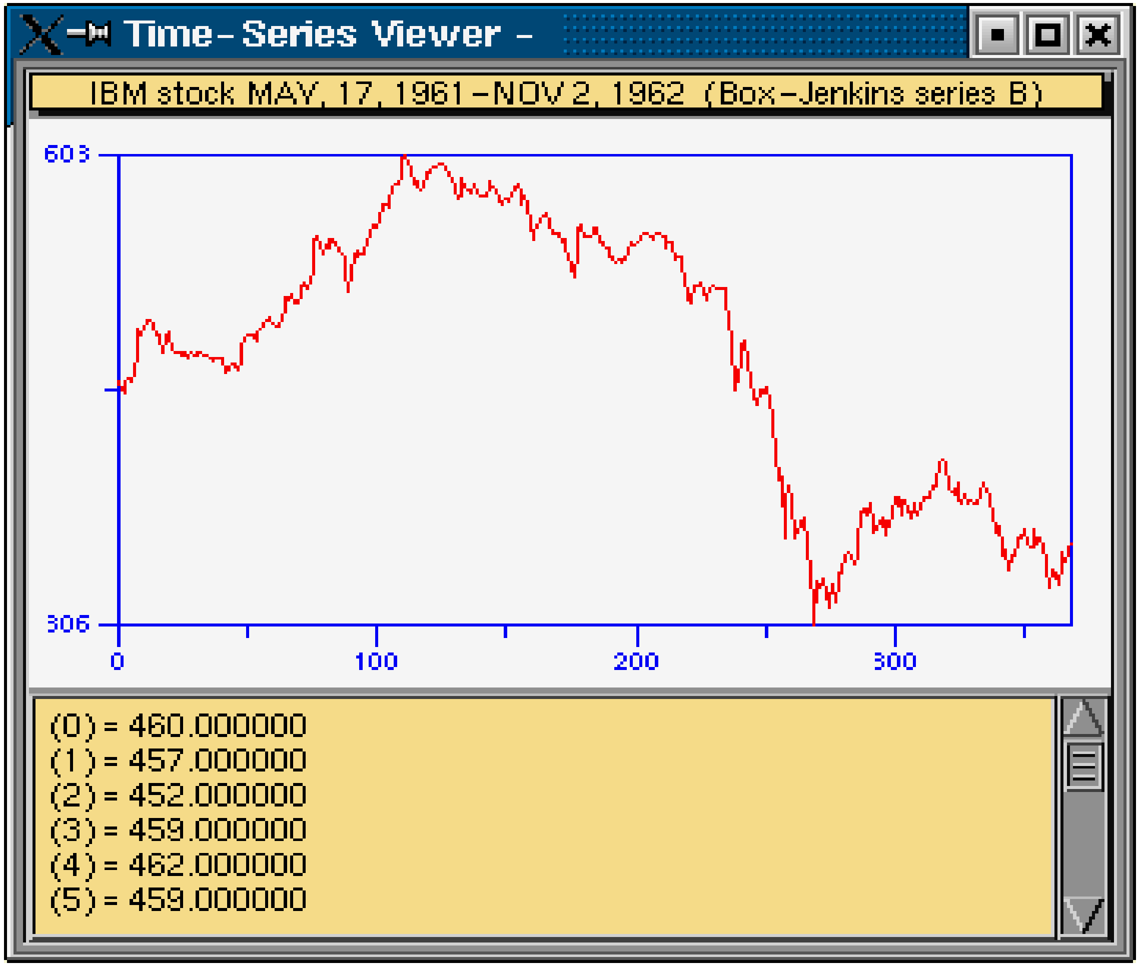
Séries Temporais

- Uma série temporal (ST) é uma seqüência de observações tomadas seqüencialmente no tempo em intervalos regulares.

$$\mathbf{z} = \left(Z_t, Z_{t-1}, Z_{t-2}, \ldots, Z_{t-N} \right)$$

- Exemplo: preços de fechamento de ações na bolsa de valores.

Séries Temporais



369 cotações da ação ordinária da IBM entre 17-MAI-1961 e 2-NOV-1962

Séries Temporais

- Não é uma série temporal:

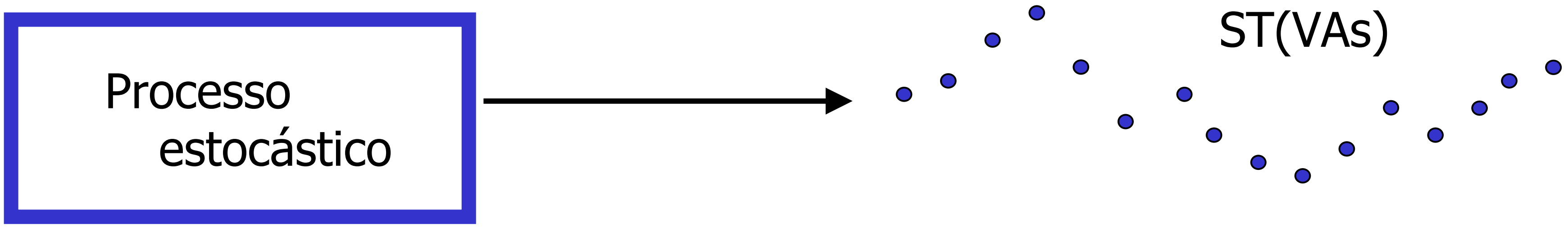
$$\mathbf{z} = \left(Z_t, Z_{t-1}, Z_{t-5}, Z_{t-10}\right)$$
$$\text{amostragem em intervalos irregulare s}$$

- As inferências em ST se dão no mesmo intervalo da amostragem ou, recursivamente, em múltiplos inteiros deste.

$$\mathbf{z} = \left(Z_t, Z_{t-10}, Z_{t-20}, \cdots\right)$$
$$\hat{Z}_{t+20} \rightarrow OK!$$
$$\hat{Z}_{t+5} \rightarrow NÃO!$$

Séries Temporais e Processos Estocásticos

- Um fenômeno estatístico que evolui no tempo segundo as leis da probabilidade é denominado *processo estocástico*.
- A realização de um processo estocástico é um conjunto de variáveis aleatórias (VA) – o registro temporal dos valores é uma série temporal.



Séries Temporais e Processos Estocásticos

- Processos estocásticos ***estacionários*** são aqueles cujas propriedades estatísticas não se alteram com o deslocamento da origem no tempo.
- Algumas propriedades estatísticas:
  - Média: nível sobre o qual o processo oscila

$$\bar{z} = \mu = \frac{1}{N}\sum_{t=1}^{N} z_t$$

  - Variância: espalhamento da oscilação

$$\sigma_z^2 = \frac{1}{N}\sum_{t=1}^{N} \left(z_t - \bar{z}\right)^2$$

Séries Temporais e Processos Estocásticos

- Análise de séries temporais
  - desenvolve um conjunto de métodos para o **tratamento das séries temporais como a realização de processos estocásticos.**

Modelo ARIMA (Box-Jenkins)

- Visão Geral
- Modelos Lineares Estacionários (ARMA).
- Modelos Lineares Não-Estacionários (ARIMA).

Modelo ARIMA – Visão Geral

- Modelo de séries temporais baseado na estrutura de dependência serial (temporal) da série histórica.
  - dependência serial = influência que um valor (evento) recebe dos valores (eventos) anteriores.

$$x_t = \overbrace{f(x_{t-1}, x_{t-2}, \cdots)} + K$$

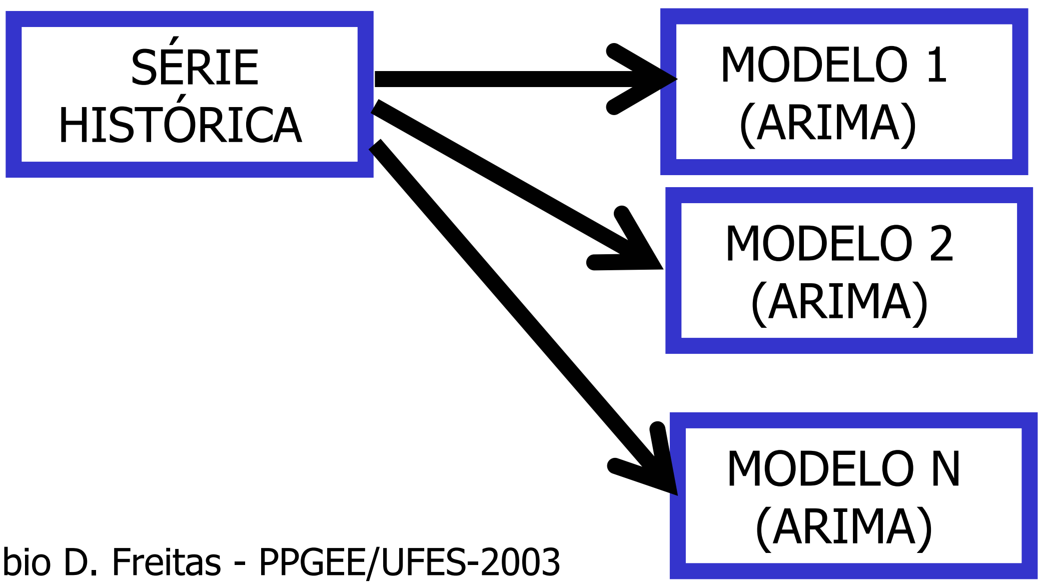
- Nos modelos ARIMA *f* é **linear em *x*** e do tipo:

$$f(x_{t-1}, x_{t-2}, \cdots, x_{t-k}) = b_1 x_{t-1} + b_2 x_{t-2} + \cdots + \cdots b_k x_{t-k} + C$$

onde k é o tamanho (ordem) da dependênci a serial.

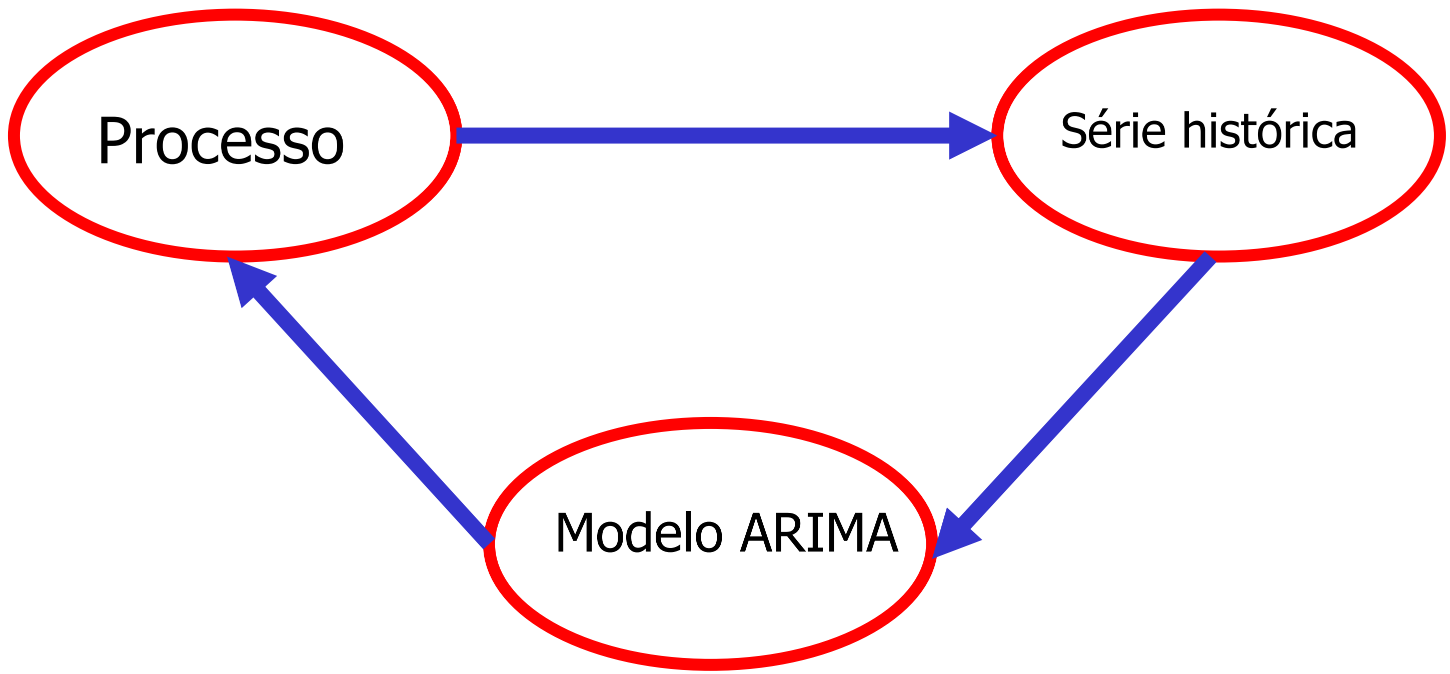
Modelo ARIMA – Visão Geral

- Um modelo é definido pelo conjunto de valores dos parâmetros k, b_1, b_2 , etc...
- "Modelagem" de uma série histórica é seleção de um modelo a partir dos valores da série, que se aproxime do processo real que gerou a série.
- Uma série histórica pode dar origem a vários modelos.



Modelo ARIMA – Visão Geral

- No escopo da modelagem, os termos *Modelo* e *Processo* são equivalentes – o modelo é uma tentativa de "engenharia reversa" do processo.



Modelo ARIMA – Visão Geral

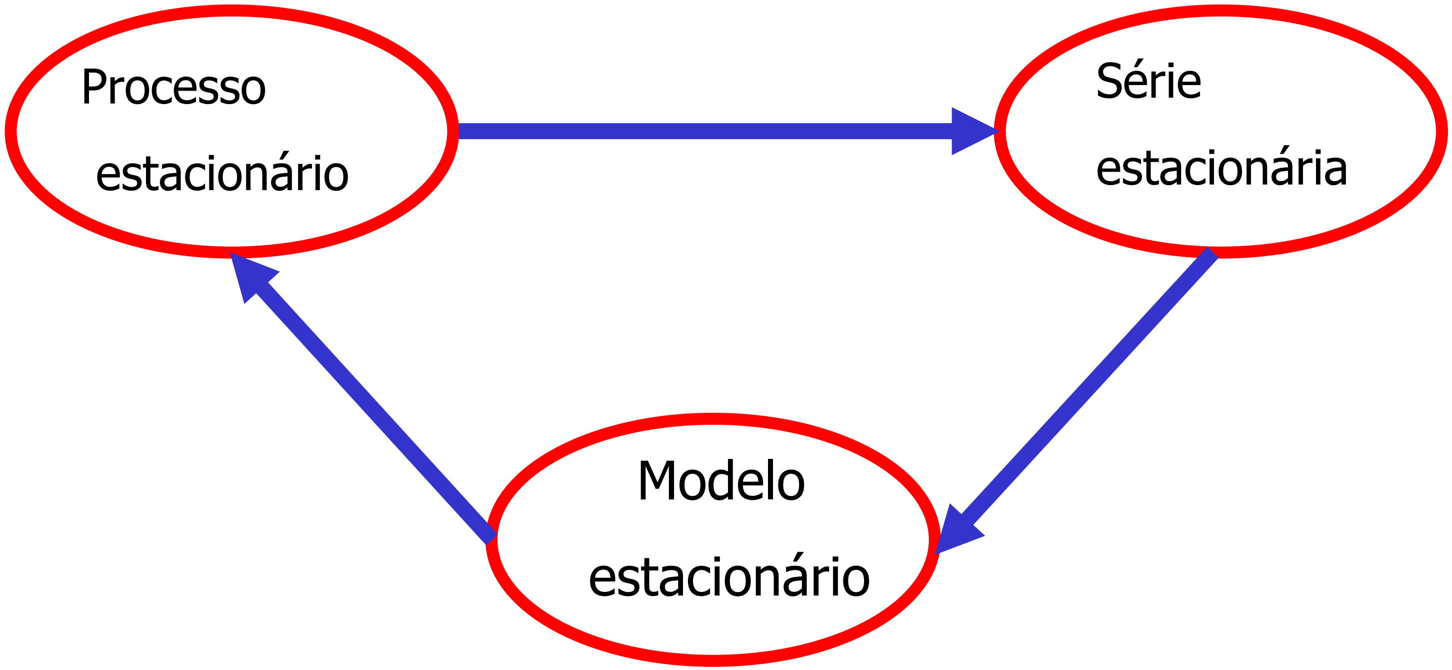
- ARIMA é uma família de **modelos (processos) lineares estacionários**:
  - AR(p) – Autoregressivo *(autoregressive)*
  - MA(q) – Médias móveis *(moving average)*
  - ARMA(p, q) – Autoregressivo-médias móveis *(auroregressive-moving average)*

  e **não estacionários**
  - ARIMA(p, d, q) – Autoregressivo-Integrado-médias móveis *(autoregressive-integrated-moving average)*.

- *p, q* são os tamanhos das dependências seriais, ou *ordem do modelo*.
- *d* é o grau da diferenciação da série histórica que garanta sua estacionariedade ►será visto adiante.

Modelos Lineares Estacionários

- Representam processos estacionários



Modelos Lineares Estacionários

- AR(p) → $\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p} + a_t$
- MA(q) → $\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$

$\tilde{z}_t = z_t - \mu$: desvios do processo em relação à sua média .

$a_t, a_{t-1}, \ldots$: resíduo - (ruído branco) distribuiç ão normal, média zero e variança σ_a^2

- ARMA(p, q) = AR(p) "+" MA(q) →

$$\tilde{z}_t = \underbrace{\phi_1 \tilde{z}_{t-1} + \ldots + \phi_p \tilde{z}_{t-p}}_{\text{AR(p)}} + a_t - \underbrace{\theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}}_{\text{MA(q)}}$$

Modelos Lineares Estacionários

- Operadores em ST:
  - Operador *backward shift* (B)

$$B^j z_t = z_{t-j} \qquad \begin{cases} B^0 z_t = z_t \\ B^1 z_t = z_{t-1} \end{cases}$$

  - Operador *forward shift* (F)

$$F = B^{-1}$$

$$F^j z_t = z_{t+j} \qquad \begin{cases} F^0 z_t = z_t \\ F^1 z_t = z_{t+1} \end{cases}$$

Modelos Lineares Estacionários

- Processo AR(p) – Autoregressivo de ordem p.
  - z_t depende linearmente dos seus valores passados.

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + \ldots + \phi_p \tilde{z}_{t-p} + a_t$$

  - Utilizando o operador B:

$$\phi(B)\tilde{z}_t = a_t$$

onde,

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \ldots - \phi_p B^p$$

$\tilde{z}_t = z_t - \mu$: desvios do processo em relação à sua média.

$a_t, a_{t-1}, \ldots$: resíduo - ruído branco : distribuição normal, média zero e variança σ_a^2

Modelos Lineares Estacionários

- Condição de estacionariedade de AR(p)
  - Raízes de $\phi(B) = 0$
    (equação característica do processo)

    devem estar fora do círculo unitário complexo.

$$\text{AR}(1): |\phi_1| < 1 \quad \text{AR}(2): \begin{cases} |\phi_2| & < & 1 \\ \phi_2 + \phi_1 & < & 1 \\ \phi_2 - \phi_1 & < & 1 \end{cases}$$

- Condição de invertibilidade de AR(p)
  - Invertível para quaisquer φ.

Modelos Lineares Estacionários

- Processo MA(q) – Médias móveis de ordem q.
  - z_t depende linearmente dos valores passados dos erros *a*. O sinal negativo é convenção.

$$\tilde{z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

  - Utilizando o operador B:

$$\tilde{z}_t = \theta(B) a_t$$

$$onde,$$
$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \ldots - \theta_q B^q$$

Modelos Lineares Estacionários

- Condição de estacionariedade de MA(q):
  - Estacionário para quaisquer θ.
- Condição de invertibilidade de MA(q) $\rightarrow$ satisfaz a dualidade dos modelos:
  - Raízes de $\theta(B) = 0$
    (equação característica do processo)
    devem estar fora do círculo unitário.

$$\text{MA}(1): |\theta_1| < 1 \quad \text{MA}(2): \begin{cases} |\theta_2| & < & 1 \\ \theta_2 + \theta_1 & < & 1 \\ \theta_2 - \theta_1 & < & 1 \end{cases}$$

Modelos Lineares Estacionários

- Dualidade entre modelos invertíveis:
  - AR(p) pode ser escrito como MA(∞).

$$\phi(\mathrm{B})\tilde{z}_t = a_t \rightarrow \tilde{z}_t = \phi^{-1}(\mathrm{B})a_t$$

$$\mathrm{Ex}^*.\ \mathrm{AR}(1): \tilde{z}_t = a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \phi_1^3 a_{t-3} + \ldots$$

  - MA(q) pode ser escrito como AR(∞).

$$\tilde{z}_t = \theta(\mathrm{B})a_t \rightarrow \theta^{-1}(\mathrm{B})\tilde{z}_t = a_t$$

$$\mathrm{Ex}^*.\ \mathrm{MA}(1): \tilde{z}_t = a_t - \theta_1 z_{t-1} - \theta_1^2 z_{t-2} - \theta_1^3 z_{t-3} - \ldots$$

(*) Expansão Binomial: $(1+x)^n = 1 + \frac{nx}{1!} + \frac{n(n-1)x^2}{2!} + \frac{n(n-1)(n-2)x^3}{3!} + \cdots$

Modelos Lineares Estacionários - ARMA

- Processo ARMA(p,q) - misto Autoregressivo-Médias Móveis de ordem *p*, *q*.
  - z_t depende linearmente dos seus valores passados e dos erros *a*.

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \ldots + \phi_p \tilde{z}_{t-p} + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

  - Utilizando o operador B:

$$\phi(B)\tilde{z}_t = \theta(B)a_t$$

$\tilde{z}_t = z_t - \mu$: desvios do processo em relação à sua média.

$a_t, a_{t-1}, \ldots$: resíduo - ruído branco : distribuição normal, média zero e variança σ_a^2

Modelos Lineares Estacionários - ARMA

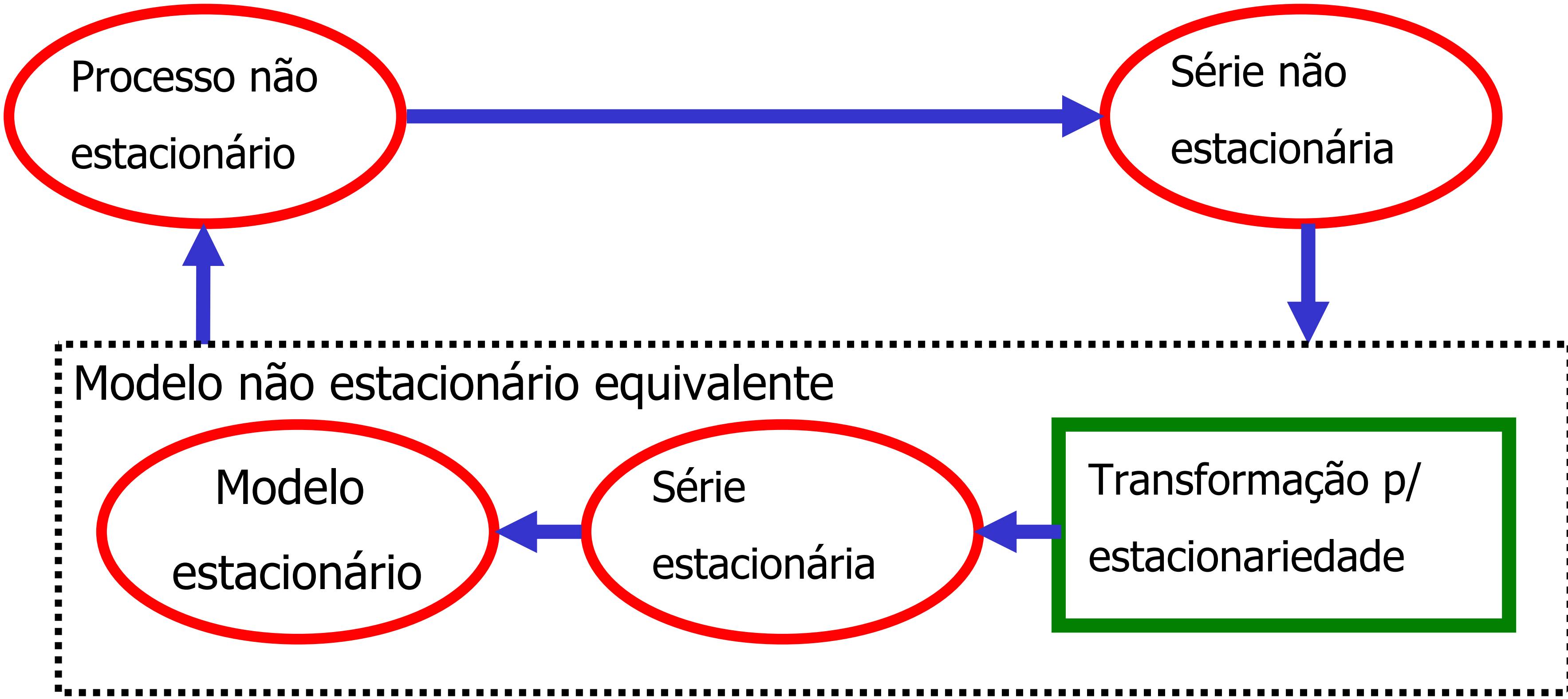
- Importância do modelo ARMA(p,q):
  - Permite modelar uma série estacionária com menos parâmetros (modelo menor) que um AR(p) ou MA(q).

    $\rightarrow (p_{ARMA} + q_{ARMA}) < p_{AR}$, q_{MA}.
- Condições de estacionariedade e invertibilidade necessárias:
  - Raízes de $\phi(B) = 0$ e $\theta(B) = 0$

  fora do círculo unitário complexo.

Modelos Lineares Não Estacionários

- Representam processos não estacionários



Modelos Lineares Não Estacionários

- Operador *d-diferenças* (∇ nabla)

$$\nabla^d = (1-B)^d \quad \begin{cases} \nabla^0 z_t = z_t \\ \nabla^1 z_t = z_t - z_{t-1} \\ \nabla^2 z_t = z_t - z_{t-2} \end{cases}$$

Modelos Lineares Não Estacionários

- Exemplo:
  - Expressão recursiva para o operador de diferenças.

$$\nabla z_t = (1 - B) z_t = z_t - z_{t-1}$$

$$\nabla^2 z_t = (1 - B)^2 z_t = (1 - B)(1 - B) z_t$$

$$= (1 - B)(z_t - z_{t-1}) = (z_t - z_{t-1}) - (z_{t-1} - z_{t-2})$$

.

$$\nabla^d z_t = \nabla^{d-1} z_t - \nabla^{d-1} z_{t-1}$$

Modelos Lineares Não Estacionários

- Operador *soma infinita* (S)

$$Sx_t = \sum_{h=-\infty}^{t} x_h = (1 + B + B^2 + \ldots)x_t$$

$$S^1 = (1 - B)^{-1} = \nabla^{-1}$$

$$S^2 = Sx_t + Sx_{t-1} + Sx_{t-2} + \cdots = (1 + 2B + 3B^2 + \ldots)x_t$$

Modelos Lineares Não-Estacionários

- Exemplo:
  - Tecnicamente, as infinitas somas do operador S não convergem. Portanto, na prática é utilizado o operador finito S_m. Assim, a integração finita utiliza um z_k ($k < t$) como valor inicial.
  - Para d=1 e m=t-k temos:

$$z_t = w_t + w_{t-1} + \dots + w_{k+1} + z_k$$

Modelos Lineares Não Estacionários

- Obtenção da estacionariedade.
  - Se z é uma série não estacionária homogênea*:

$$\mathbf{w} = \nabla^d \mathbf{z} = (\nabla^d z_t, \nabla^d z_{t+1} \ldots)$$

    é uma **série estacionária!**
  - Se **z** tem N observações, **w** terá N – d.
  - Analogamente, a série z é obtida pela integração das suas *d*-diferenças **w**.

$$z_t = S^d w_t$$

- Ainda, temos que:

$$\nabla^d \tilde{z}_t = \nabla^d z_t, \forall d \geq 1$$

(*) apresenta tendência estocástica.

Modelos Lineares Não Estacionários

- ARIMA(p, d, q) = ARI(p, d) "+" IMA(d, q)

$$\nabla^d z_t = \underbrace{\phi_1 \nabla^d z_{t-1} + \ldots + \phi_p \nabla^d z_{t-p}}_{\text{ARI(p, d)}} + \delta + a_t \underbrace{- \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}}_{\text{IMA(d, q)}}$$

$$\phi(B)\nabla^d z_t = \delta + \theta(B) a_t$$

- Ou, em **w**

$$w_t = \phi_1 w_{t-1} + \ldots + \phi_p w_{t-p} + \delta + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

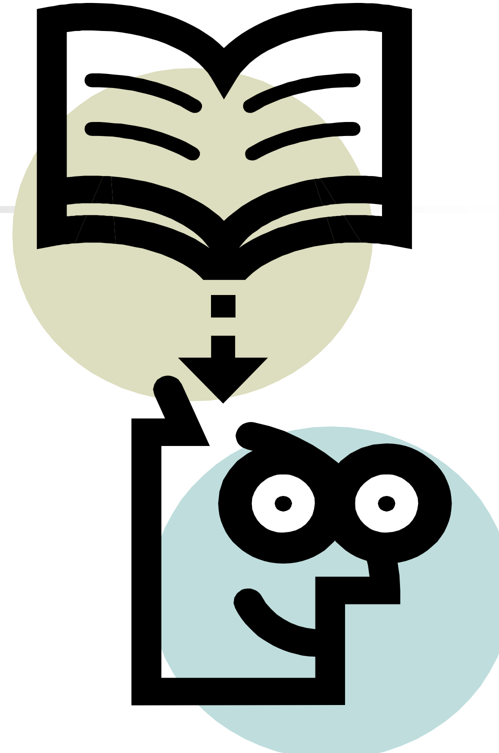
$$\phi(B) w_t = \delta + \theta(B) a_t$$

δ é uma constante de nível do modelo.

Modelos Lineares Não Estacionários

- ARIMA(p, d, q)
  - O termo *Integrado* faz alusão ao fato da série ***z*** ser obtida pela integração (soma finita em *d* vezes) da séries estacionária **w.**

ARIMA - Checkpoint 1



- ARIMA(p, d, q)
  - ARIMA(p, 0, 0) = AR(p)
  - ARIMA(0, 0, q) = MA(q)
  - ARIMA(p, 0, q) = ARMA(p,q)
  - ARIMA(p, d, 0) = ARI(p, d)
  - ARIMA(0, d, q) = IMA(d, q)
- Séries não estacionárias $\rightarrow$ $\mathbf{w} = \nabla^d \mathbf{z}$
- Condições de estacionariedade e invertibilidade:

  raízes de fora do círculo
  unitário. $\phi(B) = 0$ $\theta(B) = 0$

Metodologia Box-Jenkins

- Objetivos
- Etapas da Metodologia
  - Identificação
  - Estimação
  - Diagnóstico
  - *Forecast*

Metodologia Box-Jenkins

- Objetivo: obter parâmetros de modelo que minimizem o erro médio quadrático dos *forecasts*.
- Parâmetros:

$$p = \text{ordem do operador AR não sazonal}$$

$$\phi_1, \phi_2, \ldots, \phi_p = \text{coeficientes AR}$$

$$q = \text{ordem do operador MA não sazonal}$$

$$\theta_1, \theta_2, \ldots, \theta_q = \text{coeficientes MA}$$

$$\delta = \mu_z (1 - \phi_1 - \phi_2 - \ldots - \phi_p)$$

Metodologia Box-Jenkins

- Etapas da Metodologia:
  - Identificação: define a ordem do modelo e uma primeira aproximação do parâmetros a partir da série histórica.
  - Estimação: obtenção de estimativas eficientes dos parâmetros a partir do modelo identificado e da série histórica.
  - Diagnóstico: verificação da qualidade dos parâmetros estimados para o modelo.

Metodologia Box-Jenkins

- Dinâmica da Método:
  - Repetir as etapas 2 e 3 (eventualmente 1, 2, e 3) até que o resultado do modelo seja satisfatório.
    1. Identificação
    2. Estimação
    3. Diagnóstico
    4. *Foreacast*

Identificação

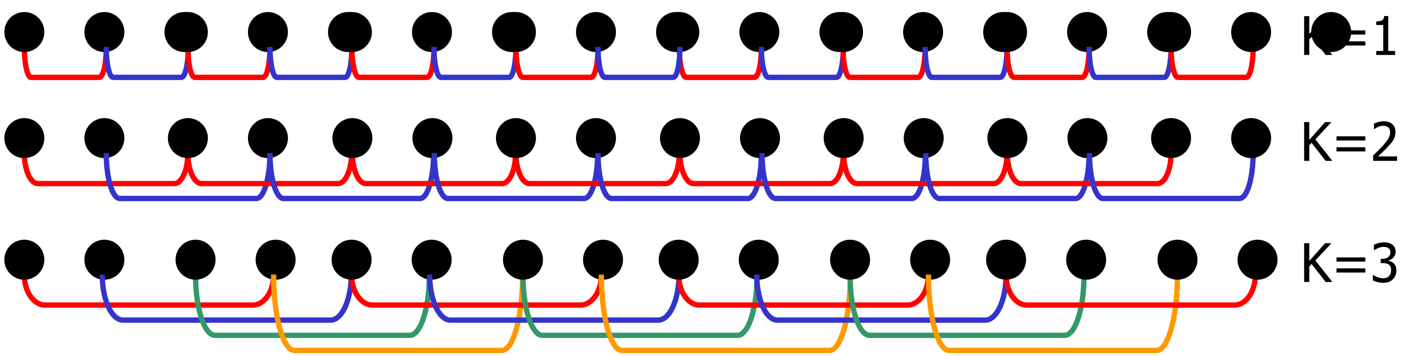
- Objetivos:
  - ? Ordem das *d*-diferenças que torne a série estacionária.
  - ? Ordens p e q dos operadores AR(p) e MA(q).
  - ? Estimativas iniciais dos parâmetros AR e MA.

Identificação

- Ferramentas:
  - Para obter *d* e *q*:
    - Análise da Função Autocorrelação Amostral – ACF *(Sample Autocorrelation Function*).
  - Para obter *p*:
    - Análise da Função Autocorrelação Amostral Parcial – PACF *(Sample Partial Autocorrelation Function*).
  - Estimativas iniciais de AR(p) e MA(q):
    - Método Box-Jenkins

Função Autocorrelação (ACF)

- A ACF mostra como a correlação (dependência serial) entre dois valores z_t e z_{t+k} da série varia com o aumento da distância (lag *k)* entre eles.



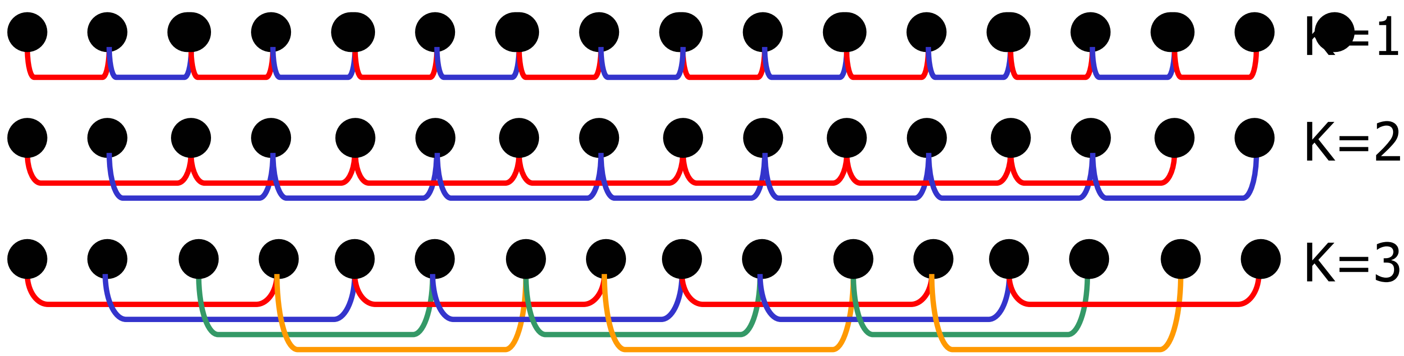
- Fornece uma visão estrutural de dependência serial da série histórica para vários intervalos *k*.
- *Correlograma* é o gráfico da ACF em função de *k*.

Função Autocorrelação (ACF)

- Cálculo dos coeficientes da ACF

$$r_k = \frac{\frac{1}{N}\sum_{t=1}^{N-k}\left(z_t - \bar{z}\right)\left(z_{t+k} - \bar{z}\right)}{\sigma_z^2}$$

$$r_0 = 1$$



Função Autocorrelação (ACF)

- Significância estatística dos valores da ACF
  - Erro Padrão (s*tandard error)* dos coeficientes da ACF - Critério de Bartlett.

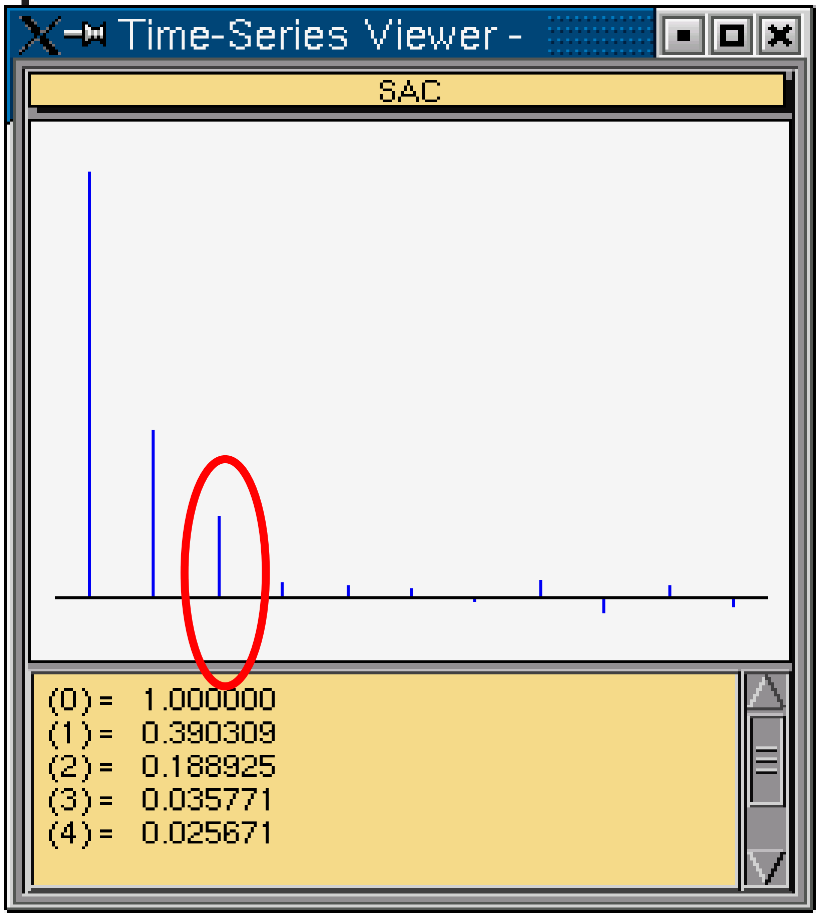
$$se_{r_k} = \sqrt{\sigma^2_{r_k}} \Rightarrow \sigma^2_{r_k} = \frac{1 + 2\sum_{j=1}^{k-1} r_j^2}{N}$$

$$r_k < 1.96\ se_{r_k} \rightarrow r_k \cong 0$$

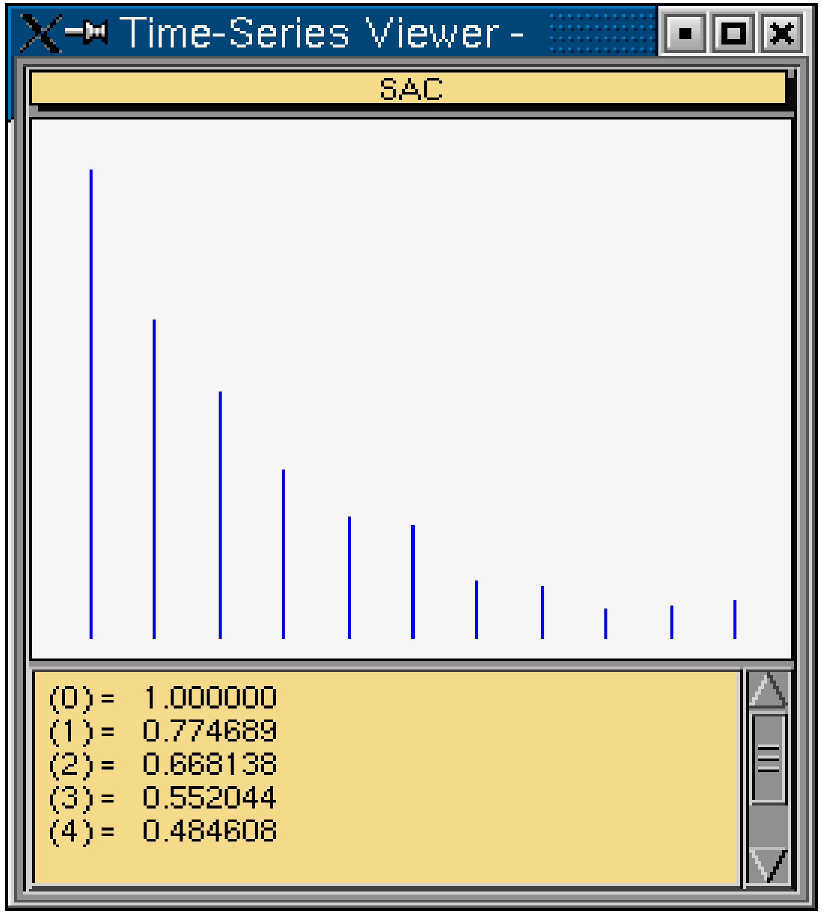
  - i.e. valores menores que o SE são desprezados (estatisticamente = 0).

Função Autocorrelação (ACF)

- Comportamentos típicos da ACF (correlograma)



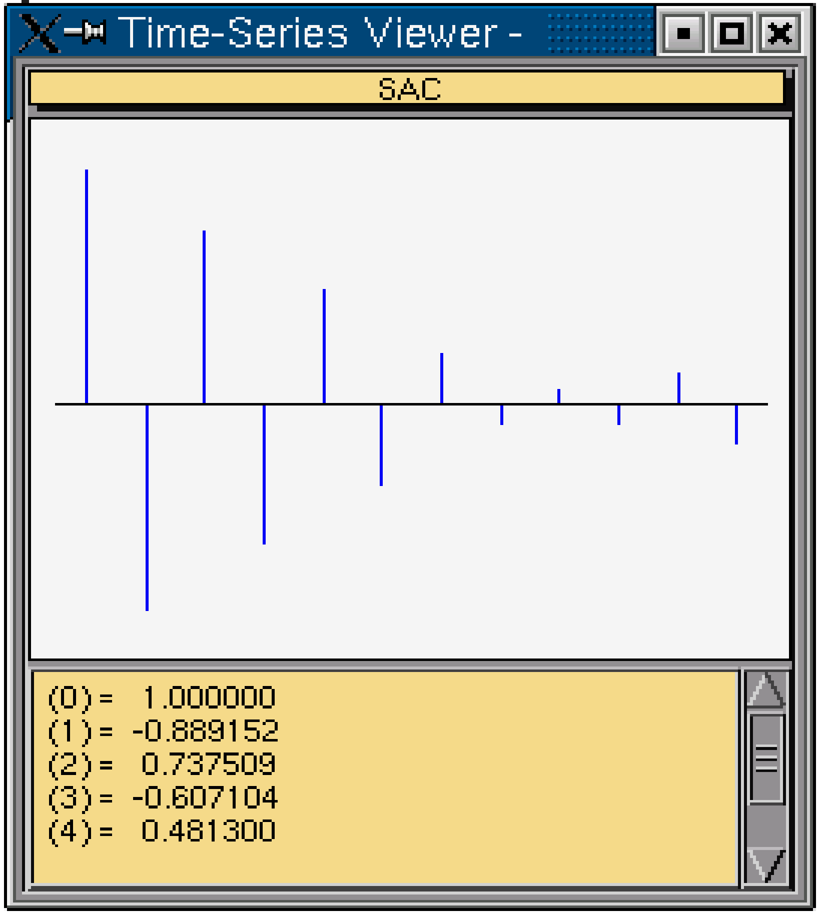
a) com corte no lag k



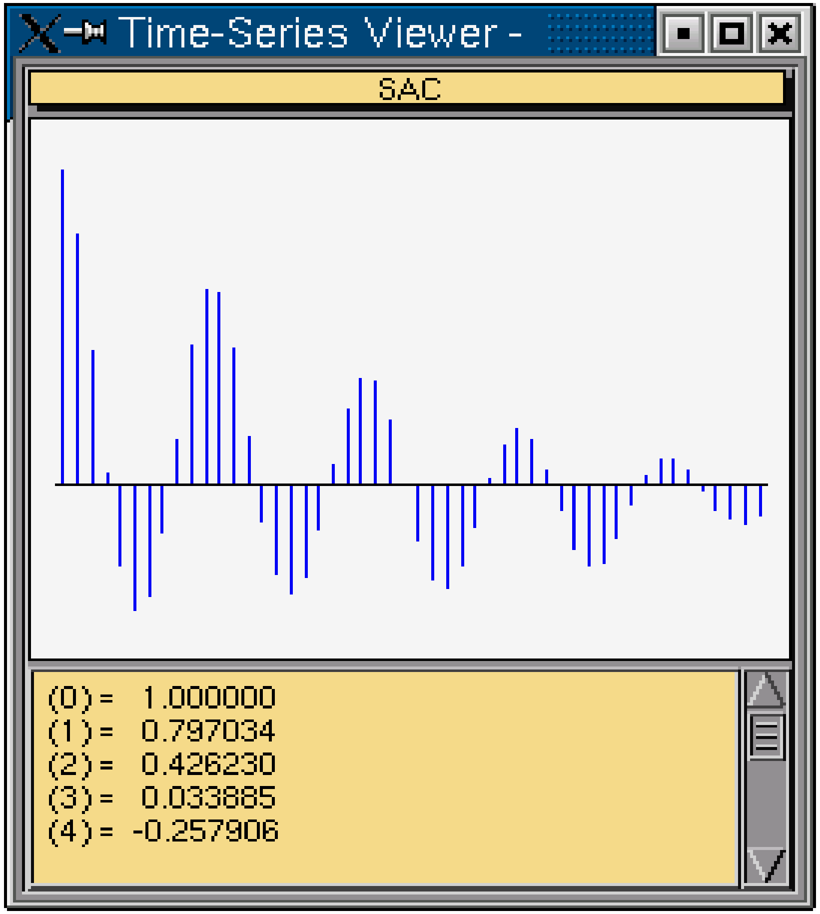
b) amort. exponencial – sem oscilação

Função Autocorrelação (ACF)

- Comportamentos típicos da ACF (correlograma)



c) amort. exponencial – com oscilação



d) amortecimento senoidal

Função Autocorrelação (ACF)

- Estacionariedade da série através do comportamento da ACF:

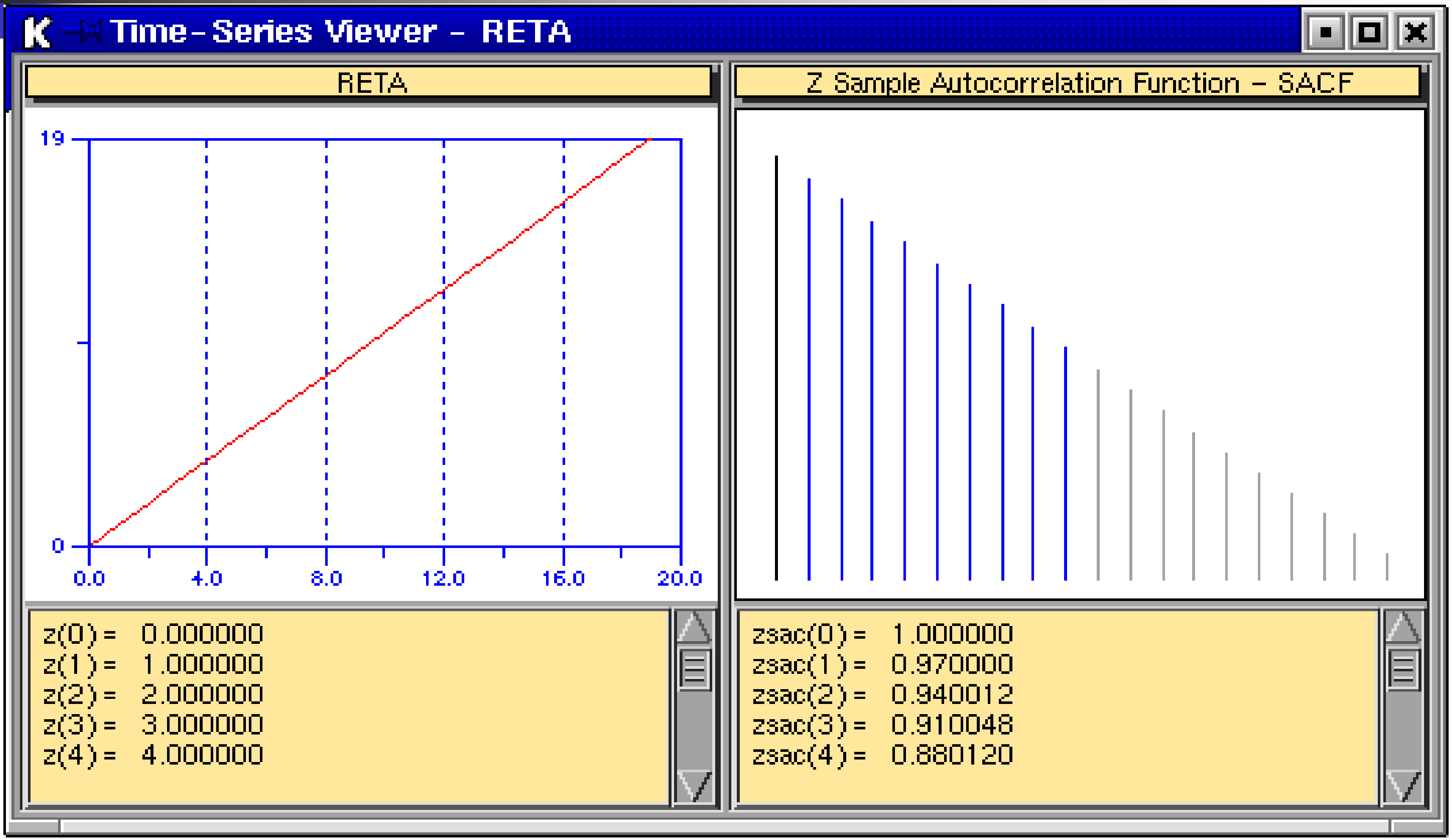
  - Corte ou amortecimento súbito com ou sem oscilações:

    →série estacionária

  - Amortecimento lento com ou sem oscilações (1º valor próximo a 1):

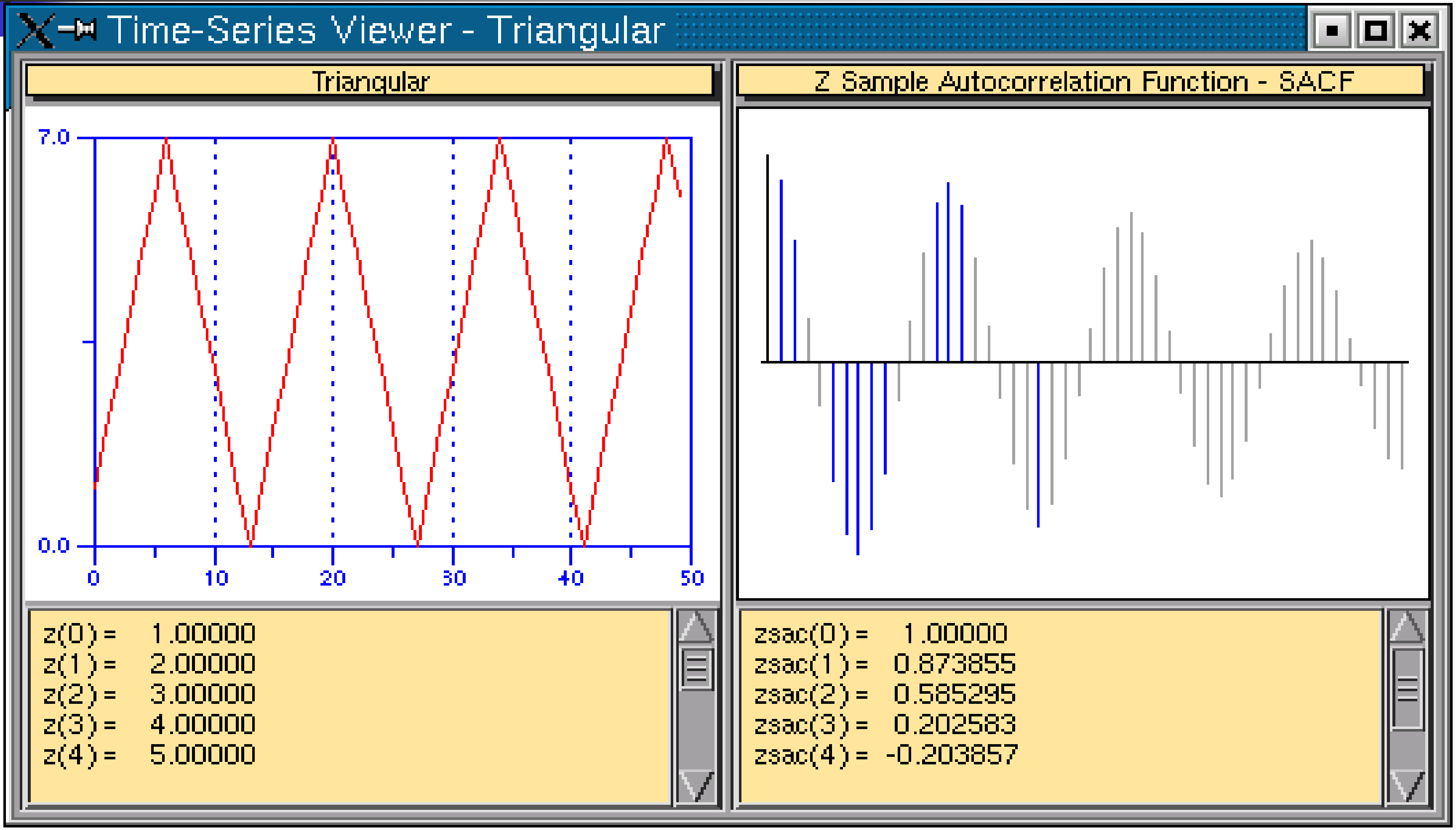
    →série não estacionária

Função Autocorrelação (ACF)



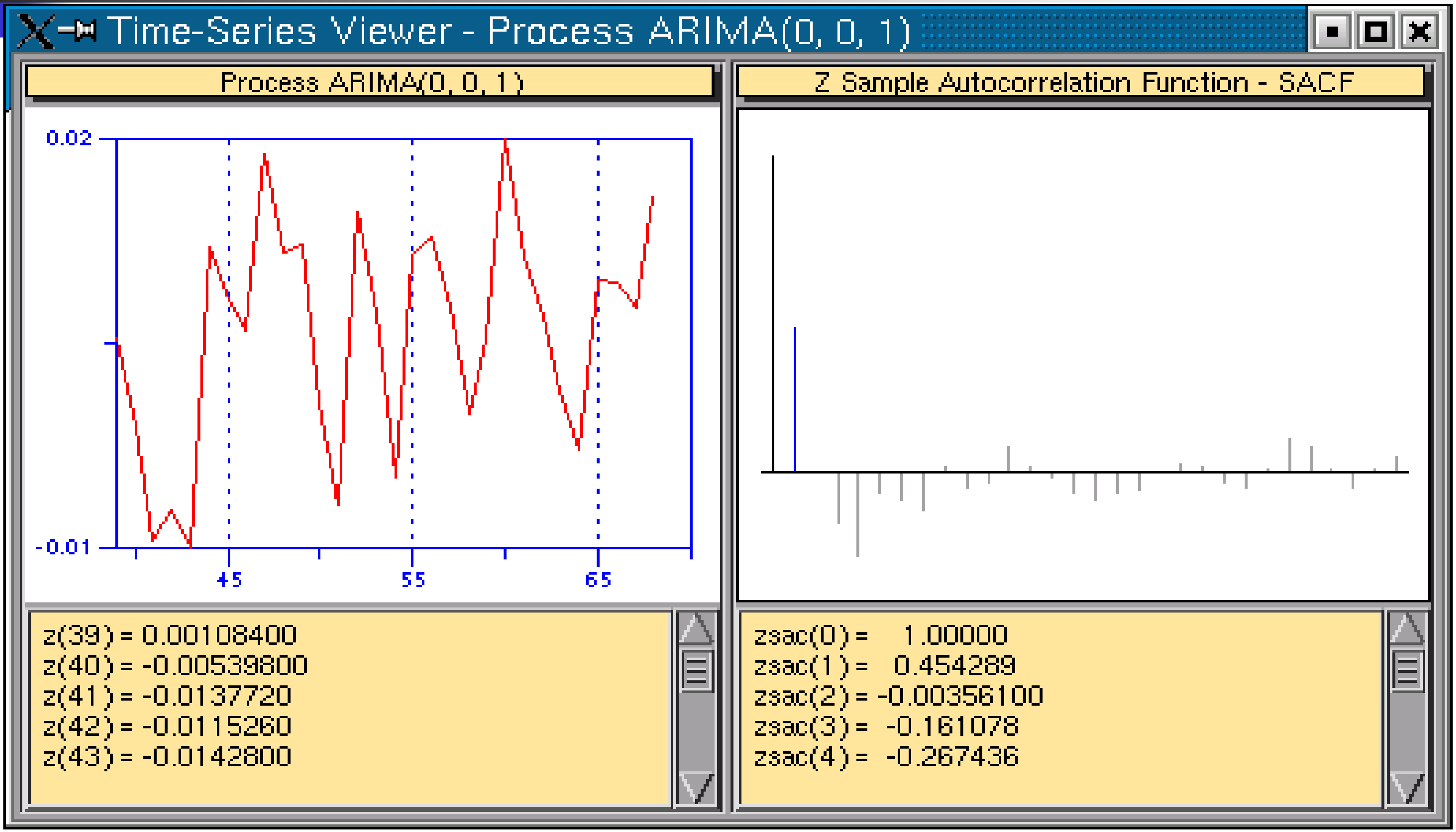
Série Não Estacionária

Função Autocorrelação (ACF)



Série com estacionariedade

Função Autocorrelação (ACF)

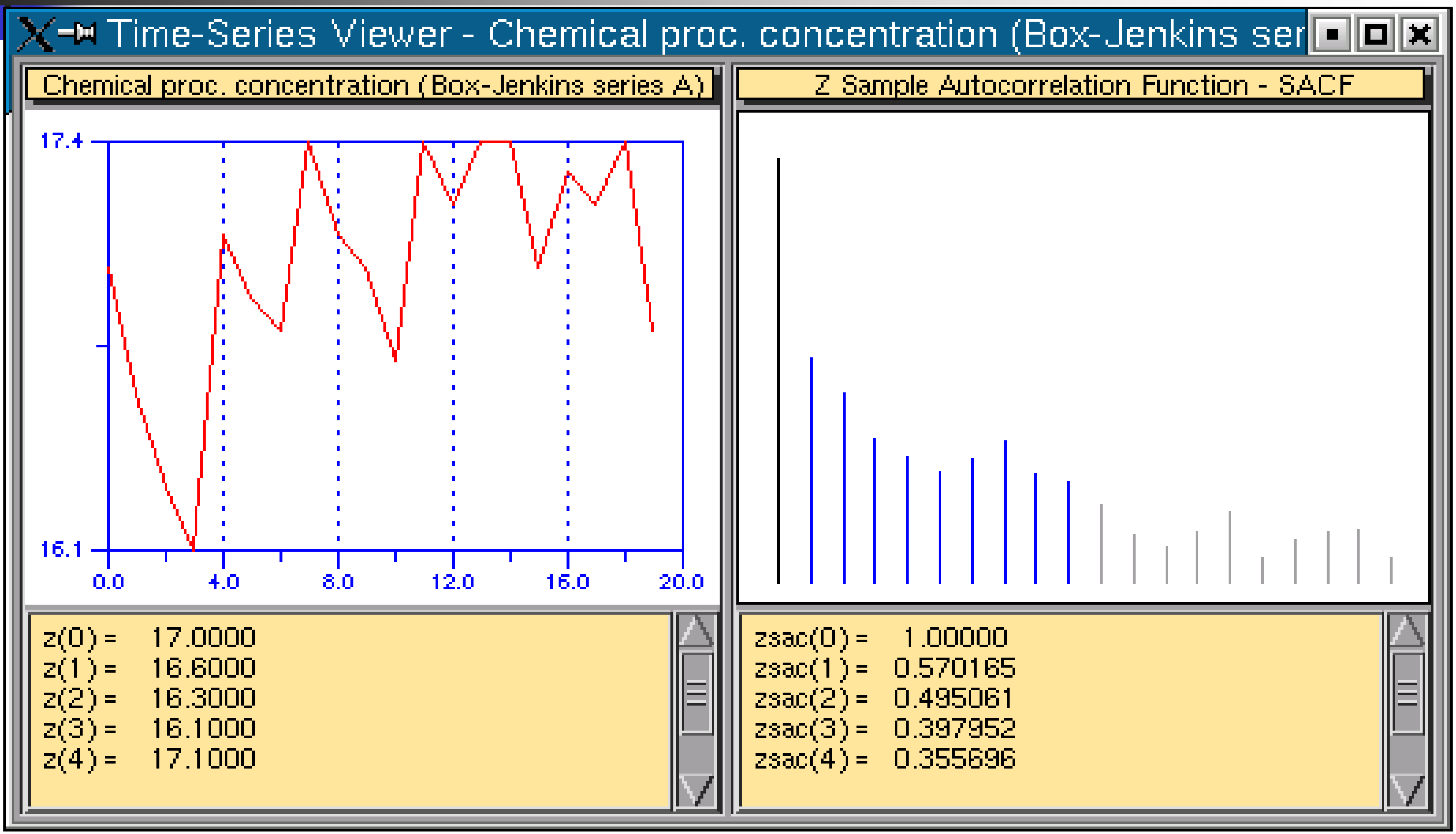


Série MA(1) Φ= -0.8 - Estacionária

Função Autocorrelação (ACF)

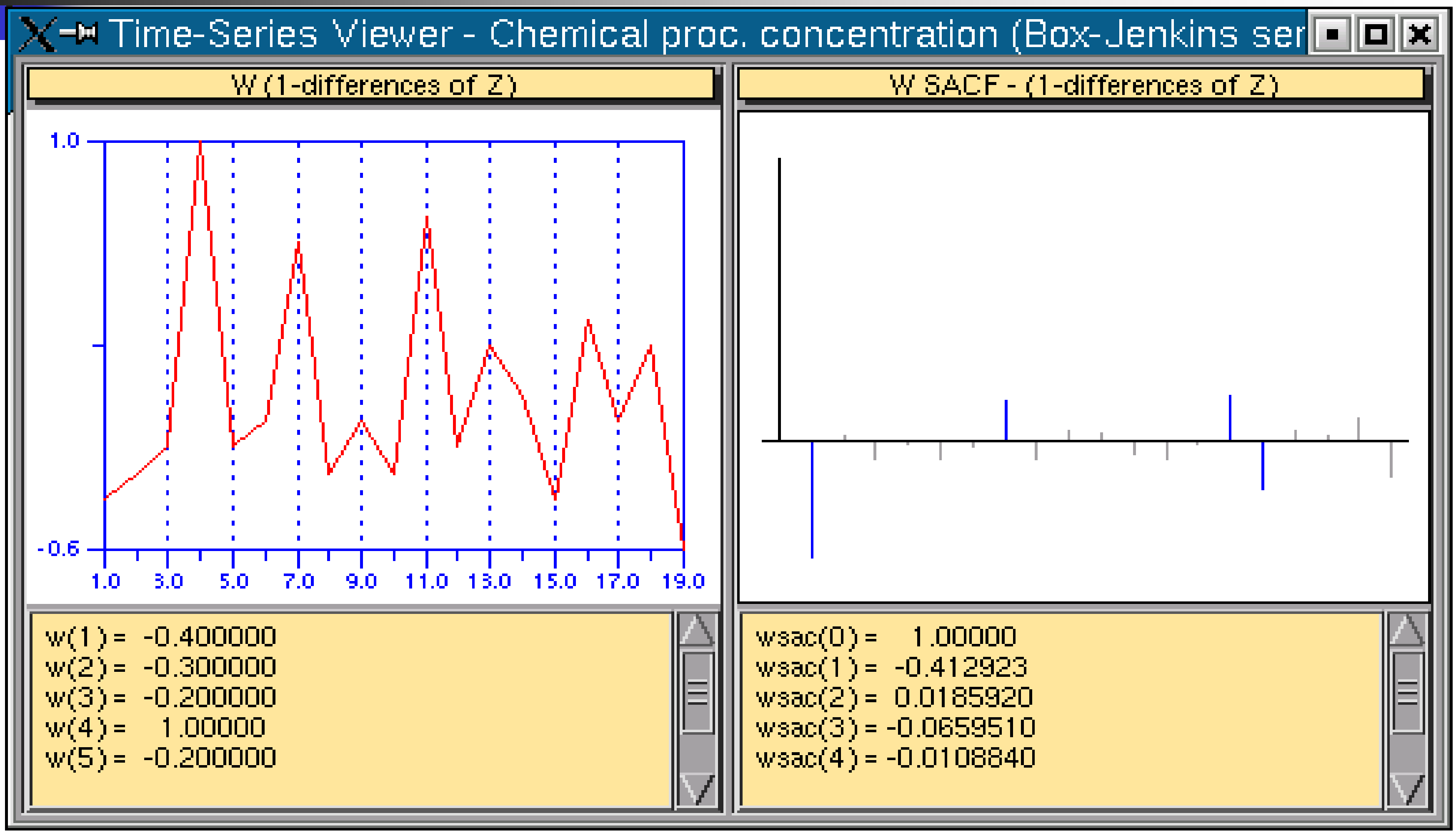
- Série **não** estacionária: aplica-se o método das diferenças (d=1, 2, ...) até que a nova série torne-se estacionária.
  - Tipicamente d=0, 1, ou 2:
    - ***d* = 0** (série original estacionária)
    - ***d* = 1** (série original com tendência constante)
    - ***d* = 2** (série original com tendência variável no tempo).
  - A ordem ótima da diferenciação geralmente apresenta o menor desvio padrão da série.
  - d=0 implica em $\delta \neq 0$ (constante do modelo).
- Usualmente a inspeção das 20 primeiras autocorrelações é suficiente.

Função Autocorrelação (ACF)



Série A (BoxJenkins) - Não Estacionária

Função Autocorrelação (ACF)

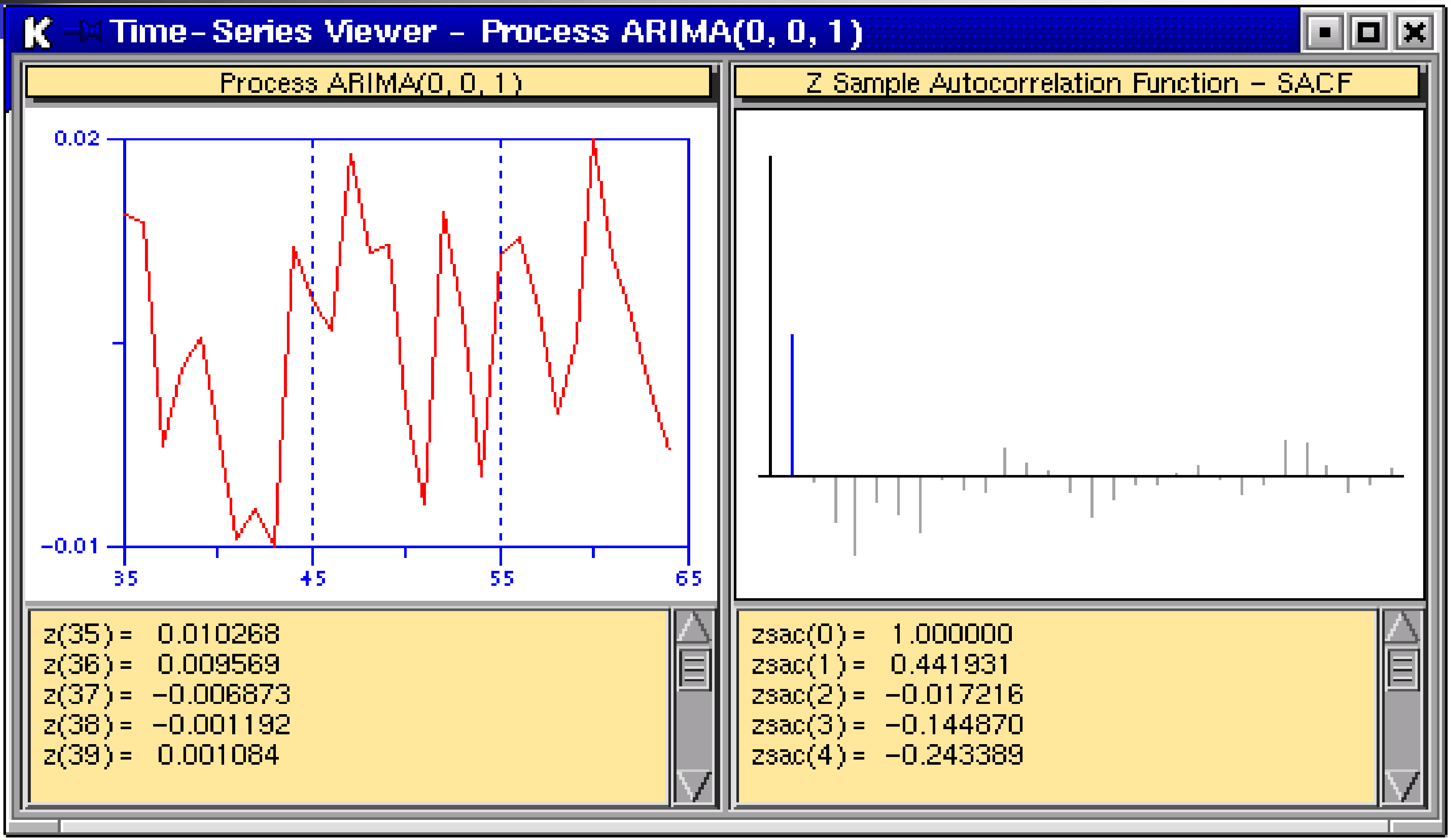


1-diferenças da Série A (BoxJenkins) - Estacionária

Função Autocorrelação (ACF)

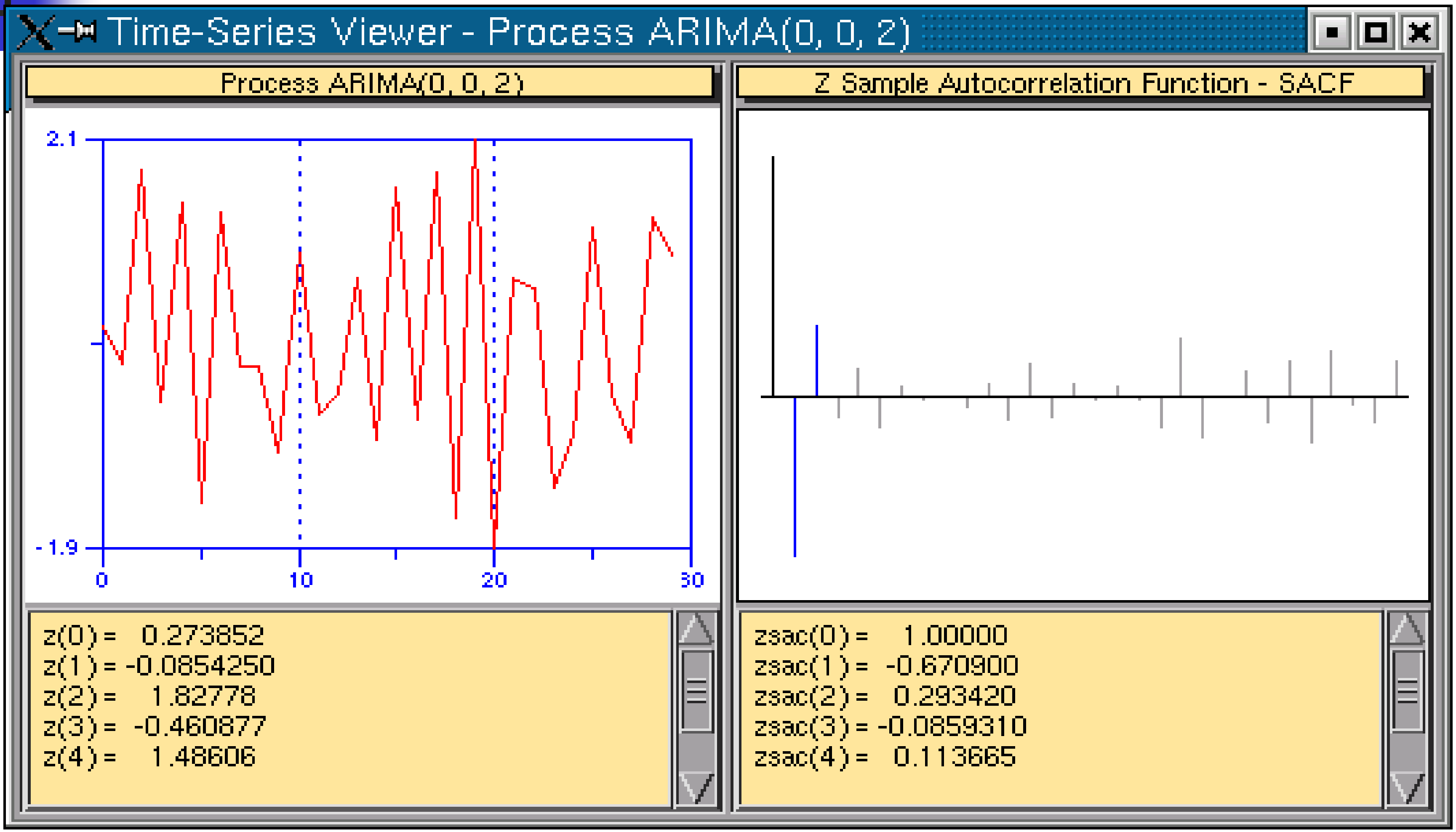
- **A ACF de um processo MA(q) somente terá valores até o *lag p*, com corte abrupto (*cut-off*).**
  - Dependência serial (regressão) com os *q* resíduos passados (*a*).

Função Autocorrelação (ACF)



Série MA(1) Φ= -0.8 - Estacionária

Função Autocorrelação (ACF)



Série MA(2) Φ1= 0.85, Φ2= -0.65 - Estacionária

Função Autocorrelação Parcial (PACF)

- A PACF calcula a autocorrelação entre z_t e z_{t+k} excluindo o efeito dos pontos intermediários.
  - $\exists \ \mathrm{Corr}(z_t, z_{t-1}), \ \exists \ \mathrm{Corr}(z_{t-1}, z_{t-2}) \Rightarrow \exists \ \mathrm{Corr}(z_t, z_{t-2})$

    então a correlação de *lag* 1 se propaga para as correlações de *lag* 2 e superiores.
  - A **correlação parcial** no *lag k* é a correlação descontada das correlações esperadas devido à propagação das correlações de *lag* inferiores.
  - No *lag* 1 a ACF e PACF são iguais.

Função Autocorrelação Parcial (PACF)

- Cálculo dos coeficientes da PACF (Equações de Yule-Walker):

$$r_{kk} = \phi_{kk} = \begin{cases} r_1 & , se\ k = 1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} r_{k-1,j} r_j} & , se\ k = 2,3,\ldots \end{cases}$$

Função Autocorrelação Parcial (PACF)

- Cálculo recursivo dos coeficientes da PACF.

$$\textit{Algoritmo de Durbin - Levinson}$$

$$\text{para } l = 1,2,\cdots,(L-1).$$

$$\text{inicialização}\begin{cases} p_{11} & = & r_1 \\ v_1 & = & 1 - r_1^2 \end{cases}$$

$$\begin{aligned} p_{l+1,l+1} &= (r_{l+1} - p_{l,1}r_l - p_{l,2}r_{l-1} - \cdots - p_{l,l}r_1)/v_l \\ p_{l+1,j} &= p_{l,j} - p_{l+1,l+1}p_{l,l+1-j}, \; j = 1,2,\cdots,l \\ v_{l+1} &= v_l(1 - p_{l+1,l+1})(1 + p_{l+1,l+1}) \end{aligned}$$

Função Autocorrelação Parcial (PACF)

- *Standard Error* dos coeficientes da PACF (Critério de Quenouille)

$$\sigma^2_{r_{kk}} \cong \frac{1}{N} \qquad , \text{para } \; \mathrm{k} \geq \mathrm{p}+1$$

$$se_{r_k} = \frac{1}{\sqrt{N}}$$

$$r_{kk} < 1.96\, se_{r_{kk}} \rightarrow r_{kk} \cong 0$$

Função Autocorrelação Parcial (PACF)

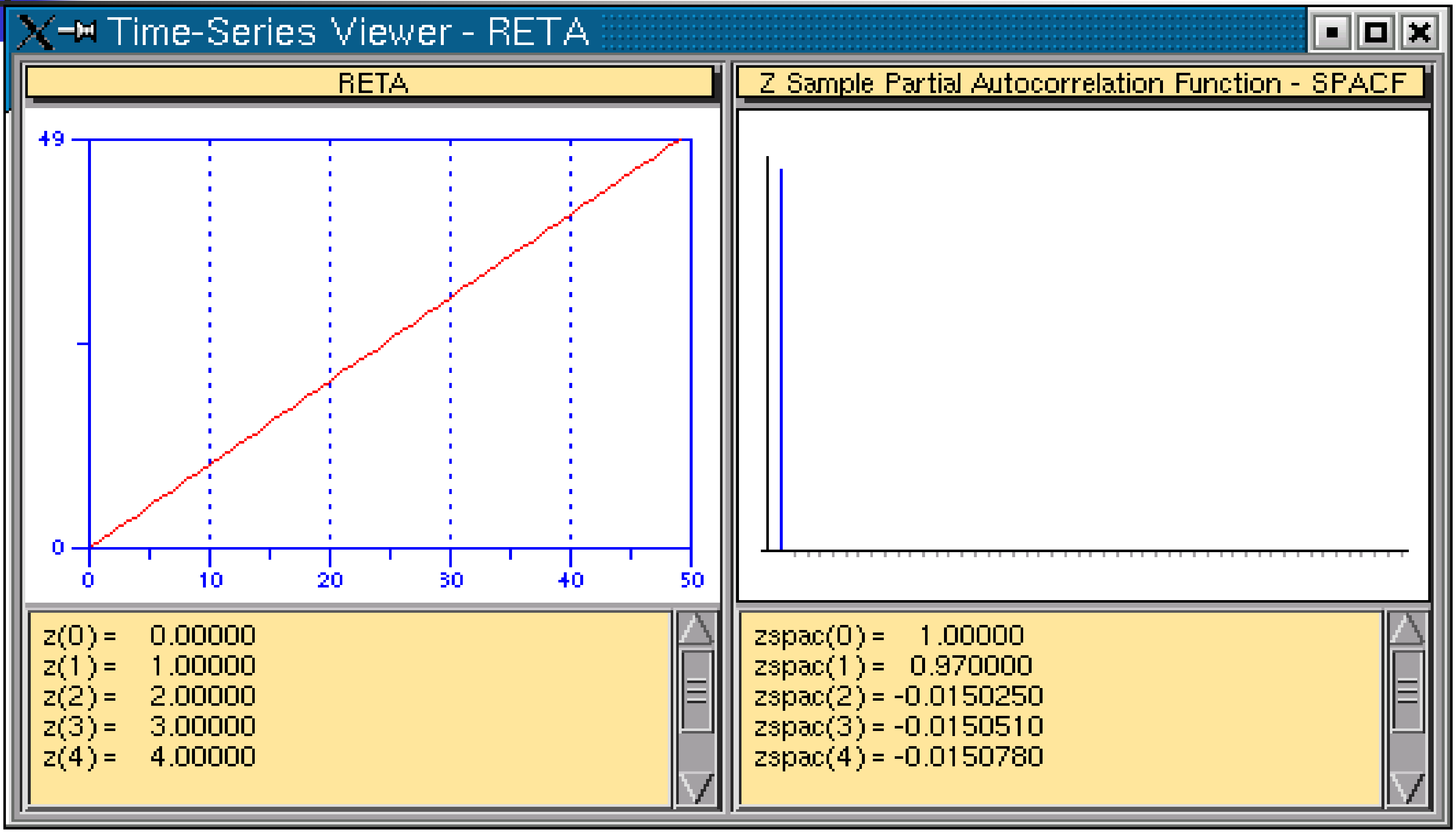
- A PACF mostra quais as correlações são efetivas:
  - Se uma ACF tem correlações até o *lag n* e a PACF tem até o *lag* 1, então as correlações de *lag k* > 1 da ACF são explicadas como propagações da correlação de *lag* 1.
- **A PACF de um processo AR(p) somente terá valores até o *lag p*, com corte abrupto *(cut-off)*.**
  - Dependência serial (regressão) com os *p* valores passados de *z*.

Função Autocorrelação Parcial (PACF)

- Outro significado dos coeficientes da PACF:
  - Os coeficientes r_{kk} da PACF são os coeficientes Φ_{kk} de maior ordem para um modelo AR(k) para a série, ou seja:

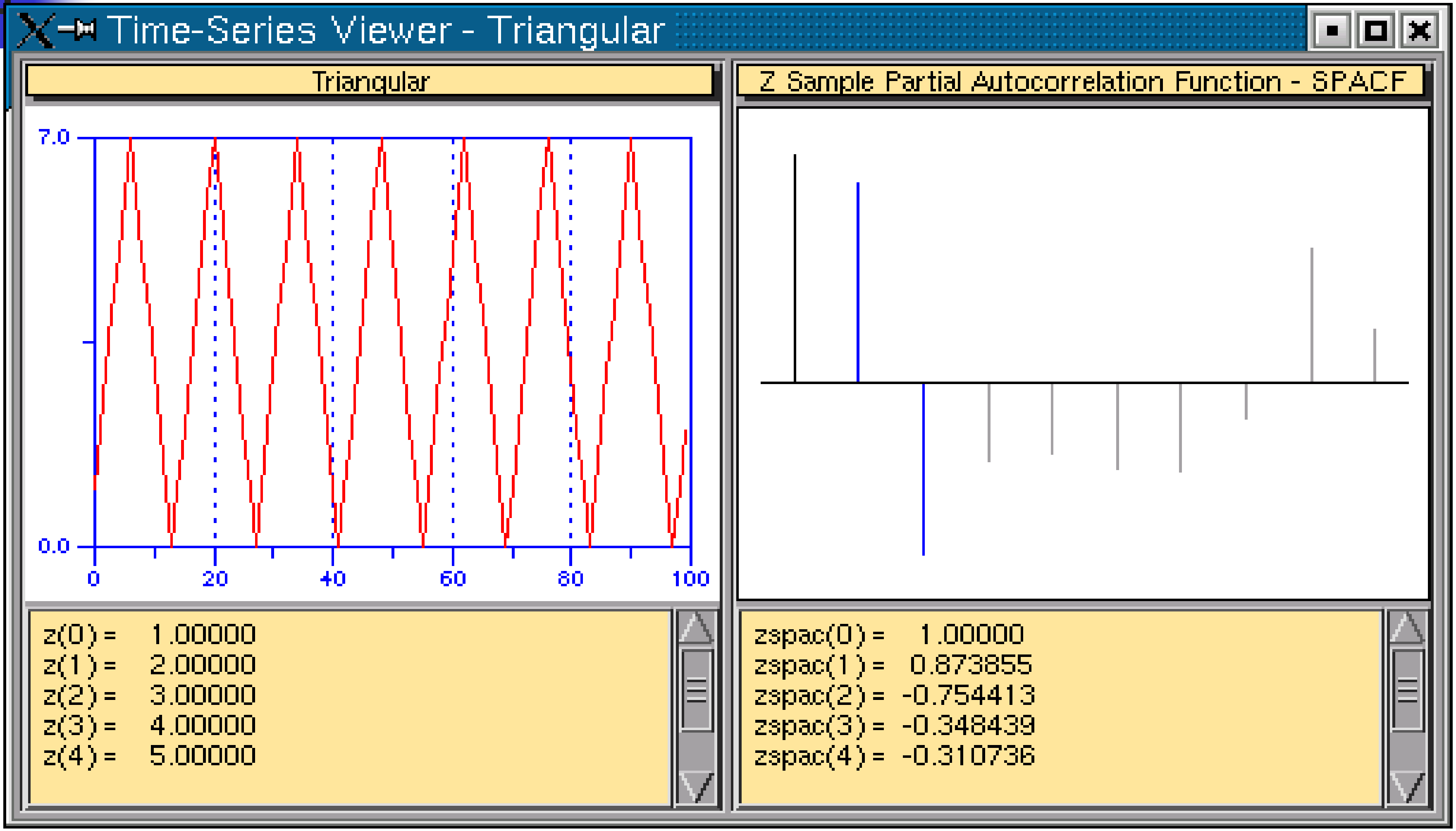
$$k = 1 - \quad \mathrm{AR}(1) : \phi_{11}$$
$$k = 2 \rightarrow \mathrm{AR}(2) : \phi_{21}, \phi_{22}$$
$$k = 3 \rightarrow \mathrm{AR}(3) : \phi_{31}, \phi_{32}, \phi_{33}$$
$$k = 4 \rightarrow \mathrm{AR}(4) : \phi_{41}, \phi_{42}, \phi_{43}\ \phi_{44}$$

Função Autocorrelação Parcial (PACF)



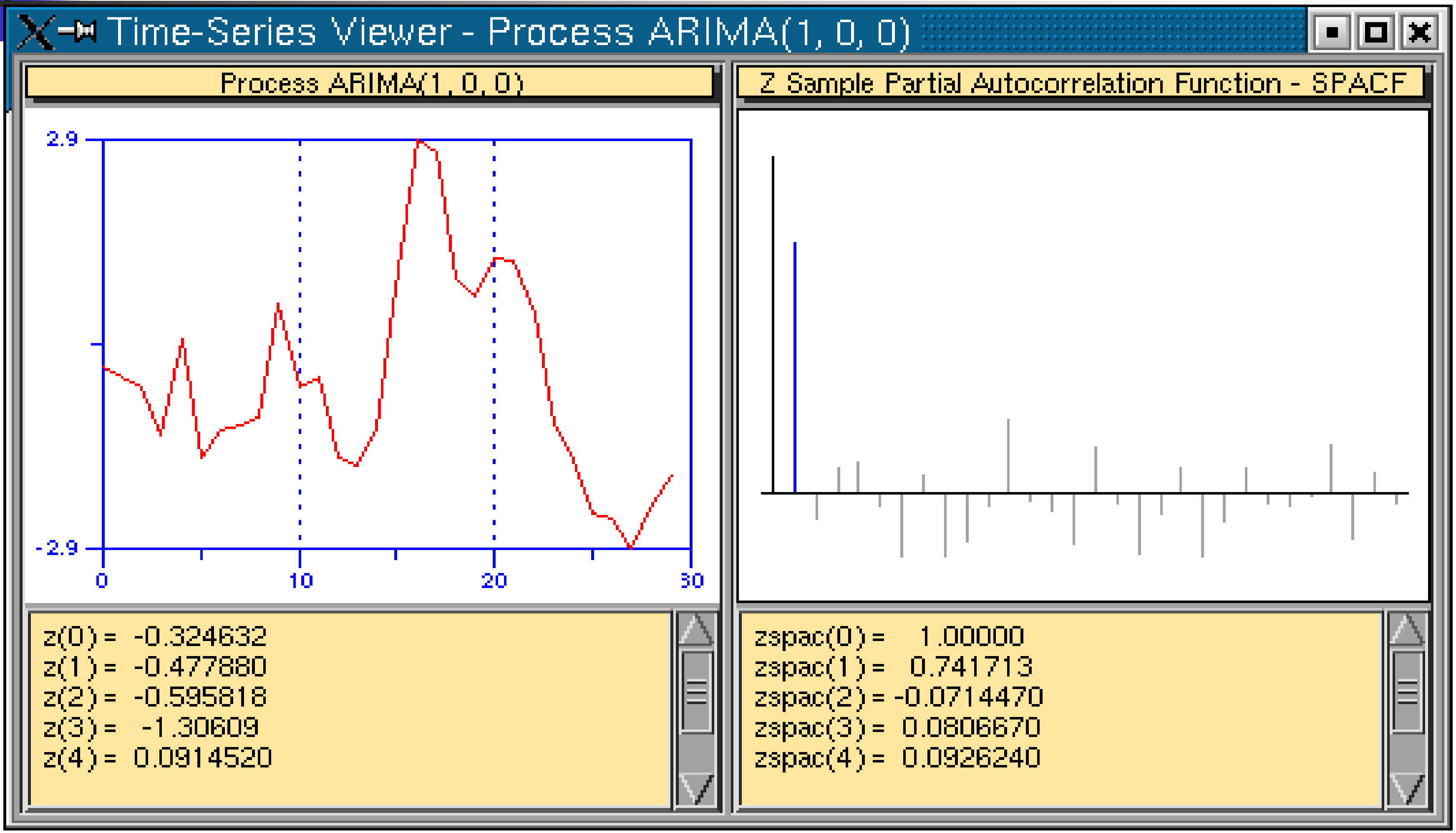
Da série de correlações da reta, apenas a primeira é efetiva

Função Autocorrelação Parcial (PACF)



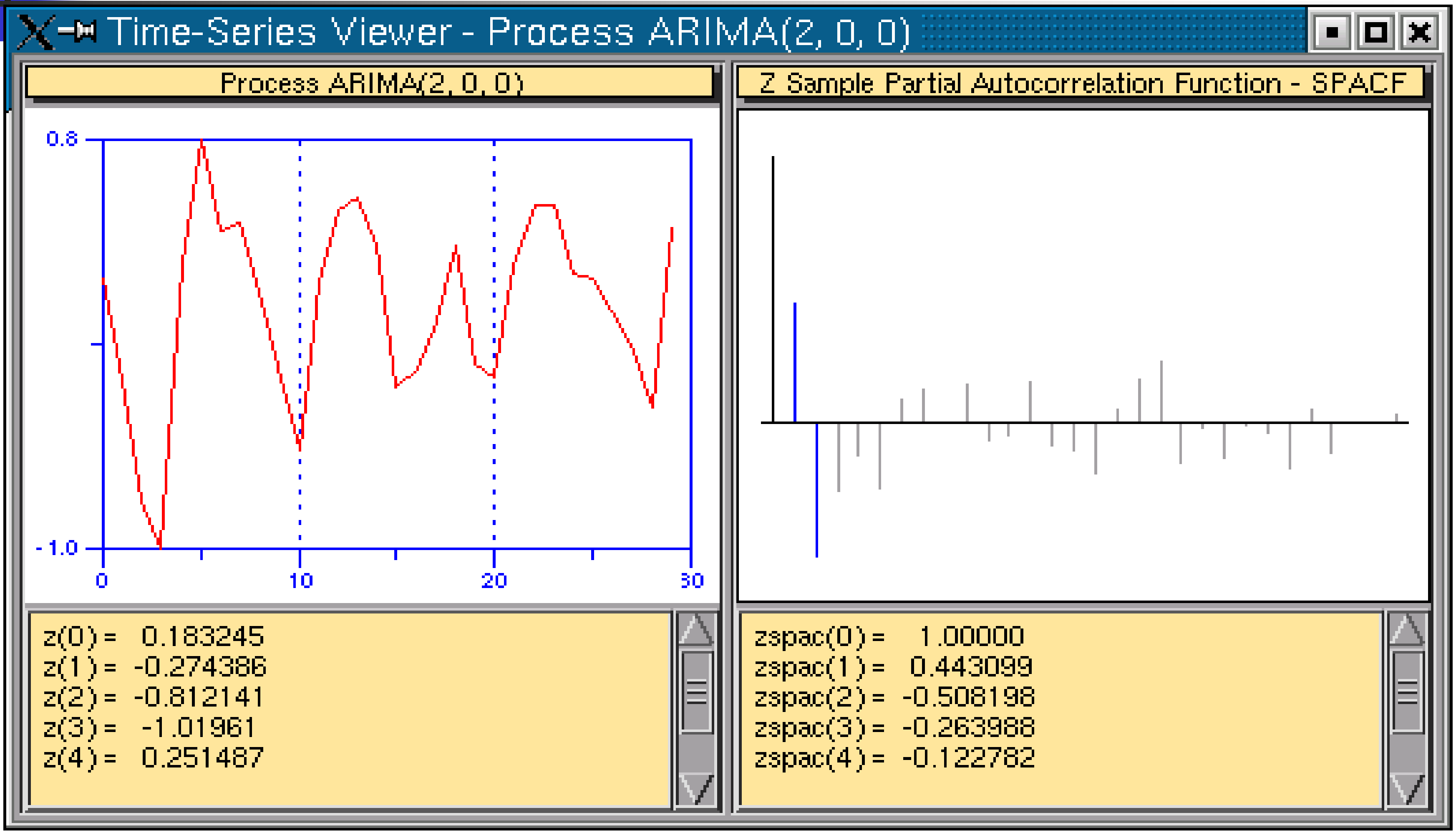
Correlações efetivas

Função Autocorrelação Parcial (PACF)



SAPCF de um processo AR(1)

Função Autocorrelação Parcial (PACF)



SAPCF de um processo AR(2)

Identificação – Estimativas Iniciais

- As estimativas iniciais dos parâmetros do modelo ARIMA:
  - Solução inicial para os algoritmos da etapa de Estimação.
  - Obtidas através das correlações (ACF) e covariâncias da séries.
  - Método Box-Jenkins

Identificação – Estimativas Iniciais

- 1- Estimativas iniciais dos parâmetros AR(p): solução do sistema (c_i são as autocovarianças de *w*):

$$
\begin{array}{ccccccccc}
c_{q+1} & = & \hat{\phi}_1 c_q & + & \hat{\phi}_2 c_{q-1} & + & \ldots & + & \hat{\phi}_p c_{q-p+1} \\
c_{q+2} & = & \hat{\phi}_1 c_{q+1} & + & \hat{\phi}_2 c_q & + & \ldots & + & \hat{\phi}_p c_{q-p+2} \\
\vdots & & \vdots & & \vdots & & & & \vdots \\
c_{q+p} & = & \hat{\phi}_1 c_{q+p-1} & + & \hat{\phi}_2 c_{q+p-2} & + & \ldots & + & \hat{\phi}_p c_q
\end{array}
$$

Identificação – Estimativas Iniciais

- 2- Primeiras q+1 autocovarianças c'_j do processo MA derivado são obtidas:

$$w'_t = w_t - \hat{\phi}_1 w_{t-1} - \ldots - \hat{\phi}_p w_{t-p}$$

$$w'_t = \hat{\phi}(B) w_t \;\therefore\; w'_t = \theta(B) a_t \quad \text{processo MA derivado}$$

$$c'_j = \sum_{i=0}^{p} \phi_i^2 c_j + \sum_{i=1}^{p} (\phi_0 \phi_i + \phi_1 \phi_{i+1} + \ldots + \phi_{p-1} \phi_p)(c_{j+i} + c_{j-i})$$

para $j = 0,1,\ldots,q$

onde $\phi_0 = -1$

Identificação – Estimativas Iniciais

- 3- Estimativas iniciais dos parâmetros MA(q):

processo linearmente convergente da função autocovariança.

$$\gamma_0 = (1 + \theta_1^2 + \ldots + \theta_q^2)\sigma_a^2$$

$$\gamma_k = (-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q)\sigma_a^2 \quad \text{para } k \geq 1$$

computa - se os parâmetros $\sigma_a^2, \theta_q, \theta_{q-1}, \ldots, \theta_1$ nesta ordem.

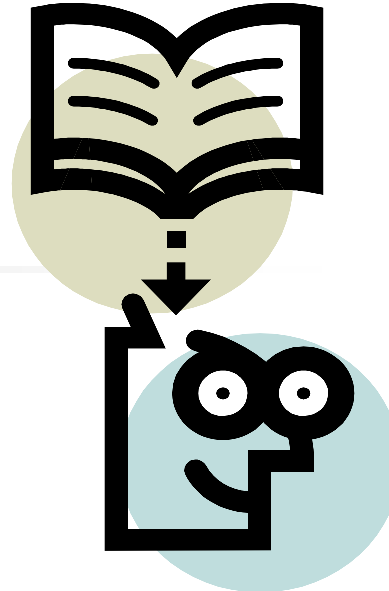
inicialmente, $\theta_1, \theta_2, \ldots, \theta_q = 0$

$$\sigma_a^2 = \frac{c'_0}{1 + \theta_1^2 + \ldots + \theta_q^2}$$

$$\theta_j = -\left(\frac{c'_j}{\sigma_a^2} - \theta_1\theta_{j+1} - \theta_2\theta_{j+2} - \ldots - \theta_{q-j}\theta_q \right)$$

usando - se $\theta_0 = 0$

ARIMA-Checkpoint 2



- Ordem das *d*-diferenças obtidas pela análise da ACF
- Propriedades da ACF e PACF:

Process	ACF	PACF
White-noise	All $\rho_s = 0$	all $\phi_{ss} = 0$
AR(1): $a_1 > 1$	Direct exponential decay(ρ_s=a_1^s	$\phi_{11} = \rho_1$, $\phi_{ss} = 0$, $s \geq 2$
AR(1): $a_1 < 1$	Oscillating decay: $\rho_s = a_1^s$	$\phi_{11} = \rho_1$, $\phi_{ss} = 0$, $s \geq 2$
AR(p)	Decays to 0, May Oscillate	Spikes thru lag p, All $\phi_{ss} = 0$, $s \geq p$
MA(1): $\beta>0$	Positive spike at lag 1, $\rho_s = 0$ for $s \geq 0$	Oscillating decay: $\phi_{11}<0$
MA(1): $\beta<0$	Negative spike at lag 1, $\rho_s = 0$ for $s \geq 0$	Decay: $\phi_{11}< 0$
ARMA(1,1)	Exponential decay beginning at lag 1, Sign ρ_1 = sign($a_1 + \beta$	Exponential decay beginning at lag 1. $\phi_{11} = \rho_1$ sign (ϕ_{ss}) = sign (ϕ_{11})
ARMA(p,q)	Decay(either direct or oscillatory) beginning at lag q	Decay (either direct or oscillatory) beginning at lag p

Estimação

- Objetivos:
  - Obter estimativas eficientes dos parâmetros de modelo, a partir do modelo identificado e da solução inicial.
  - Utilização eficiente da série histórica disponível.

Estimação

- Métodos:
  - Maximização da função de *verossimilhança (Maximum Likehood-ML)* dos parâmetros do modelo para os valores da série histórica.

Estimação

- Maximizar a Função *log-verossimilhança*:

$$l(\phi,\theta,\sigma_a^2) \cong -\frac{S(\phi,\theta)}{2\sigma_a^2} \quad \text{para } n \geq 20$$

- Minimizar a Função Soma dos Quadrados-S (*Sum of Squares)*

$$S(\phi,\theta) = \sum_{t=-\infty}^{N} \left[a_t \middle| \mathbf{w},\phi,\theta\right]^2$$

Estimação

- Outras métodos de estimação bastante utilizados (Box-Jenkins):
  - Bayesiana
  - Kalman Filter

Estimação

- S pode ser minimizada por rotinas de *Quadrados Mínimos (Least Squares).*
  - preenche-se *N+Q* equações

$$a_t = \tilde{w}_t - \phi_1 \tilde{w}_{t-1} - \ldots - \phi_p \tilde{w}_{t-p_t} + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$$

  - os valores dos parâmetros de modelo são obtidos pela minimização de *S(Φ,Θ)*.

$$S(\phi,\theta) = \sum_{t=-\infty}^{N} \left[a_t \middle| \mathbf{w}, \phi, \theta \right]^2$$

Estimação

- Solução para a origem de *t:*
  - *Backforecasts:* valores iniciais para $t < 1$ são obtidos aplicando-se o modelo no sentido inverso do tempo.
  - Na prática S tem a forma:

$$S(\phi,\theta) = \sum_{t=1-Q}^{N} \left[a_t \middle| \mathbf{w}, \mathbf{e}, \phi, \theta\right]^2$$

onde os

a_t, w_t para $t < 1$

são obtidos pelos backforecasts **e**

Estimação

- LS encontra <u>MÍNIMO LOCAL</u> a partir do estado sugerido pela estimativa inicial dos parâmetros.
  - Se a solução não for satisfatória, tenta-se outro estado inicial.

Diagnóstico

- Objetivos:
  - verificar a qualidade dos parâmetros estimados para o modelo.
  - *Benchmark* de diversos modelos.
  - Sugerir alterações nos parâmetros.
- Métodos:
  - *Overfitting*
  - Diagnóstico dos resíduos
  - Critérios de Informação

Diagnóstico

- *Overfitting*:
  - Preencher um modelo mais elaborado visando expor direções de discrepância permitidas pelos parâmetros adicionais.

Diagnóstico dos resíduos a_t

- Resíduos a_t:

$$a_t = \tilde{w}_t - \phi_1 \tilde{w}_{t-1} - \ldots - \phi_p \tilde{w}_{t-p_t} + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$$

  - Inspeção visual do gráfico dos resíduos – *primeiro passo indispensável no diagnóstico*.
  - se o modelo é correto, a_t se aproxima do ruído branco quando N cresce.

Diagnóstico dos resíduos a_t

- Teste *Portmanteau Lack-of-Fit*

$$Q = n\sum_{k=1}^{K} r_k^2(\hat{a}), \quad \text{com} \quad n = N - d$$

é aproximado pela distribuição Chi-Quadrados

$$\chi^2(K - p - q)$$

- Testa K autocorrelações *r* em grupo.
- K ~ 20.
- se o modelo é inadequado, os valores de Q são maiores que os Chi-Quadrados.

Diagnóstico dos resíduos a_t

- Estatística Ljung-Box modificada

$$\tilde{Q} = n(n+2)\sum_{k=1}^{K} \frac{r_k^2(\hat{a})}{(n-k)} \qquad \text{com} \quad n = N - d$$

$$E\left[\tilde{Q}\right] \approx K - p - q$$

da distribuição Chi - Quadrados

$$\chi^2(K - p - q)$$

Diagnóstico dos resíduos a_t

- *Information criteria:* AIC (Akaike) e BIC(Scharwz)
  - Modelos com menor AIC e BIC são preferidos.
  - Modelos escolhidos com BIC terão nunca número de parâmetros maior que os escolhidos com AIC.

$$AIC_{p,q} \cong \ln\left(\hat{\sigma}_a^2\right) + r\frac{2}{n} + \delta$$

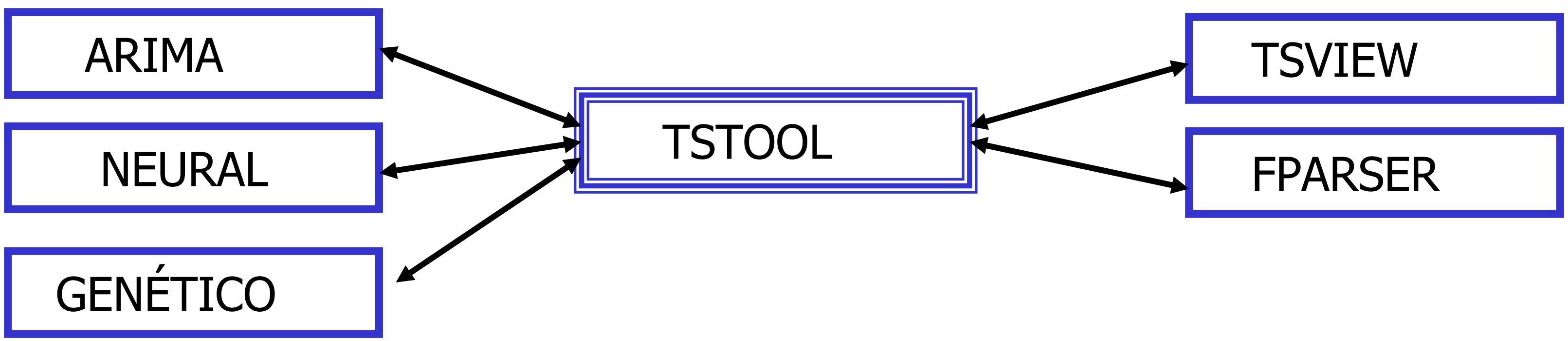
$$BIC_{p,q} = \ln\left(\hat{\sigma}_a^2\right) + r\frac{\ln(n)}{n}$$

$$r = p + q + 1$$

$\hat{\sigma}_a^2$ é a estimativa ML de σ_a^2

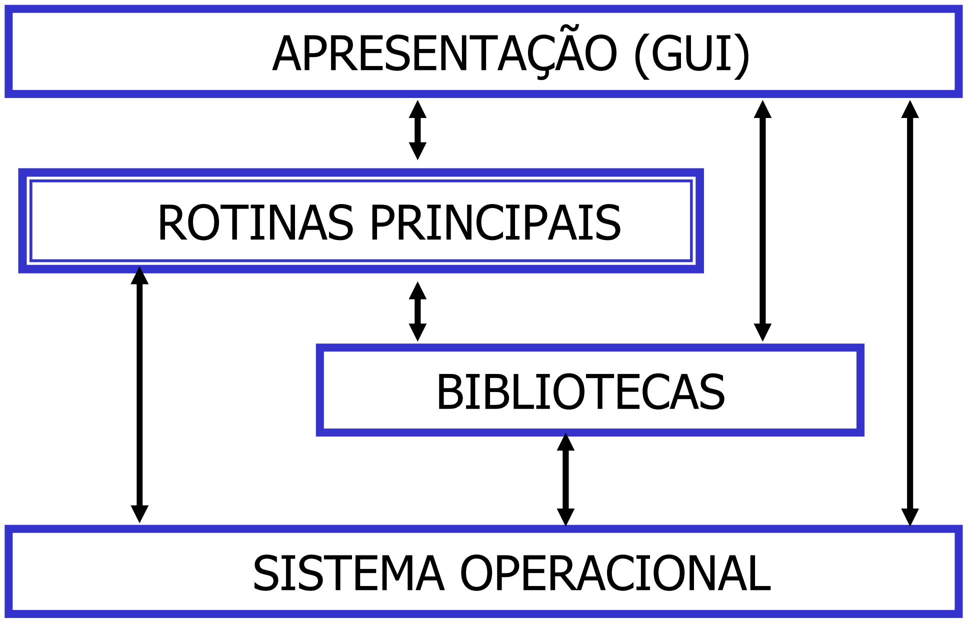
Ferramenta TSTOOL/ARIMA

- Ambiente gráfico modular multi-modelos para previsão de séries temporais
  - Linux
  - Linguagem C
  - Bibliotecas auxiliares:
    - XForms
    - gsl (Gnu Scientific Library).



Ferramenta TSTOOL/ARIMA

- Arquitetura das aplicações
  - Base de código tenta maximizar o isolamento entre camadas para modularidade e portabilidade



Ferramenta TSTOOL/ARIMA

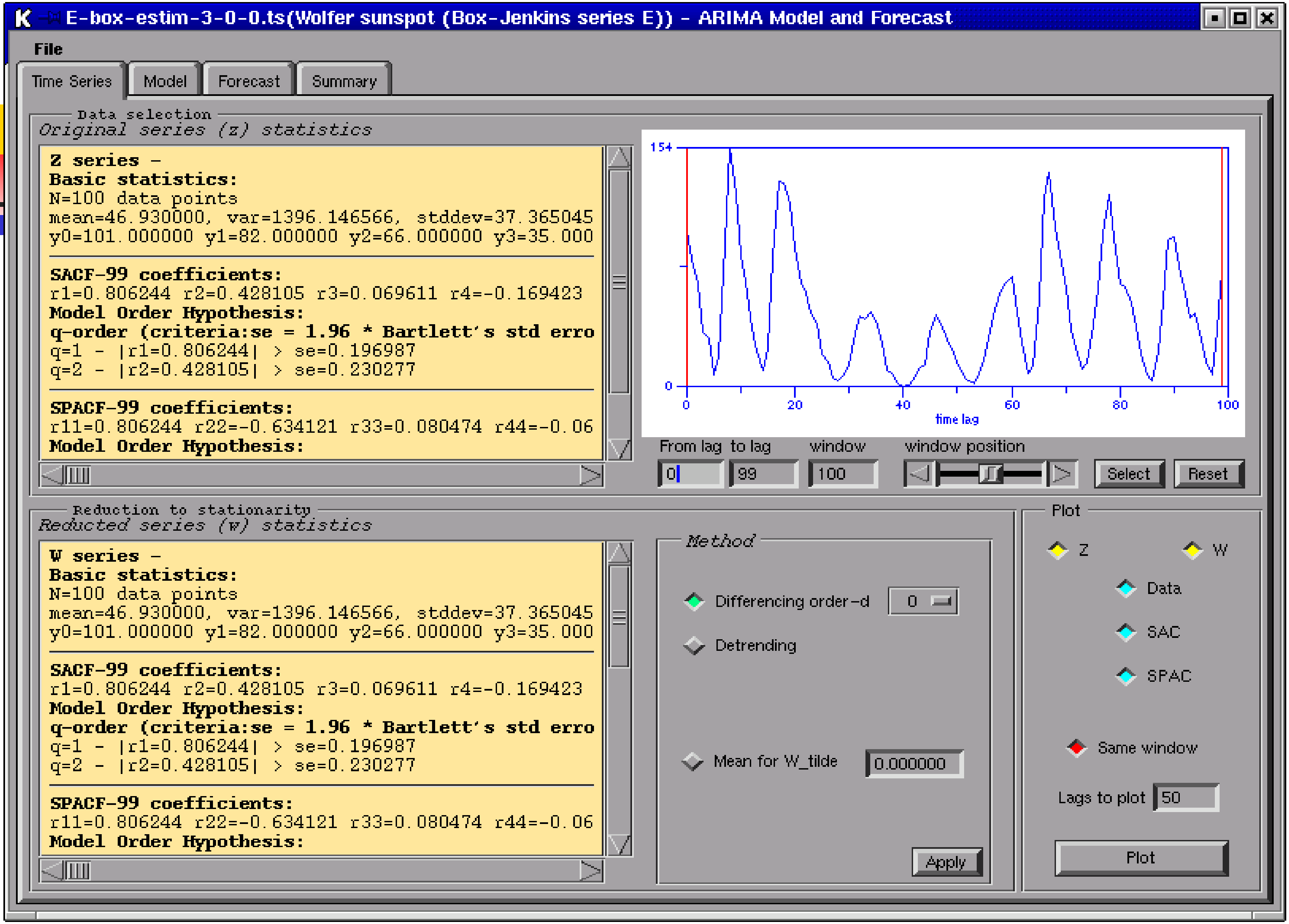
- Módulos do TSTOOL
  - TSTOOL - módulo principal :carrega, salva, exporta, etc...
  - TSVIEW – módulo de gráficos: *plot*, dados, etc...
  - FPARSER– *file parser* com suporte a séries temporais (vetores) para armazenamento e IPC.
  - ARIMA – Box-Jenkins
  - NEURAL* – Rede neural com Aspirin/MIGRAINES
  - GENÉTICO* – Algoritmo genético, neuro-genético.
- ≈ 14.500 LOC** (linhas de código)

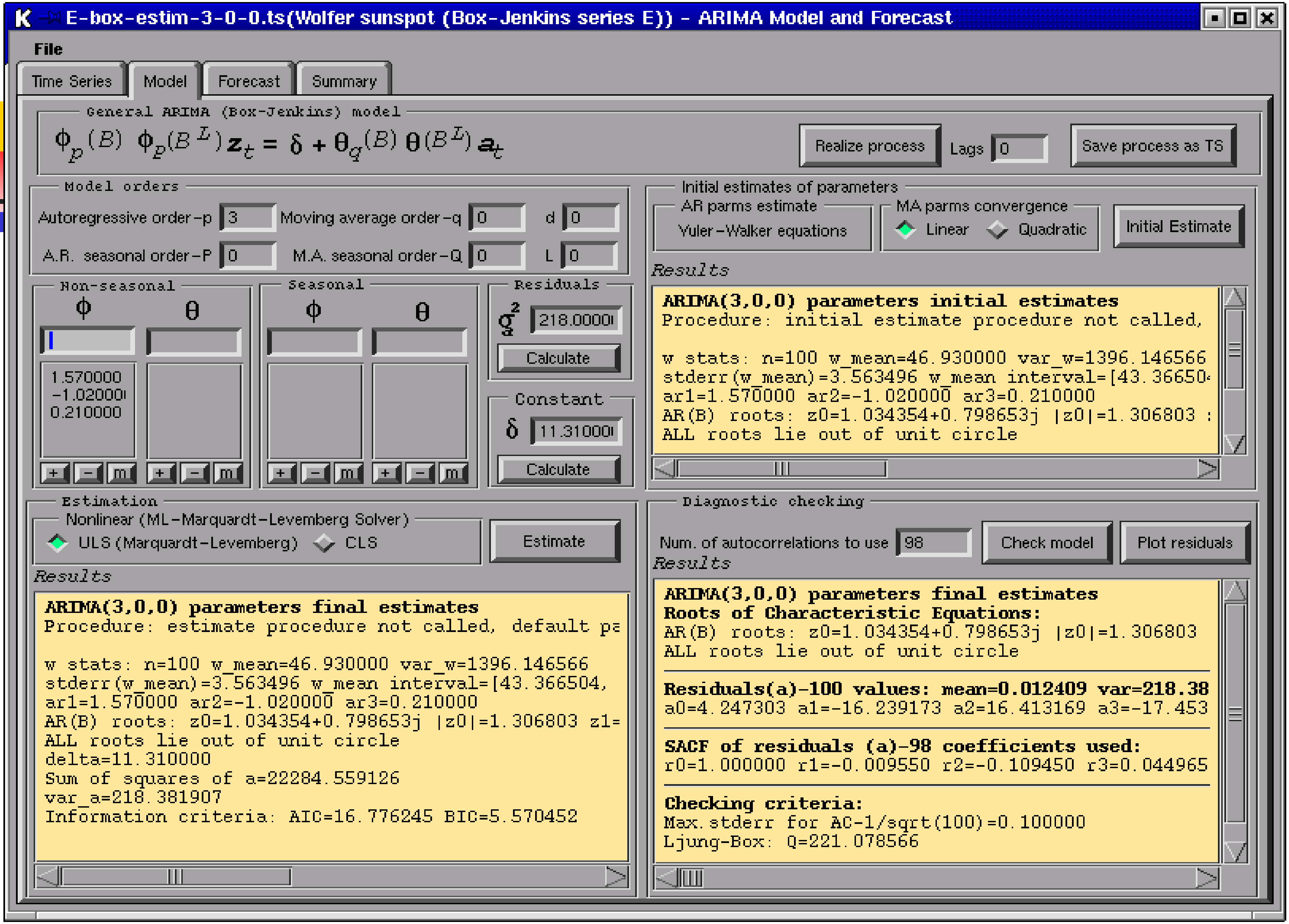
(*) Implementados porém ainda não integrados ao TSTOOL

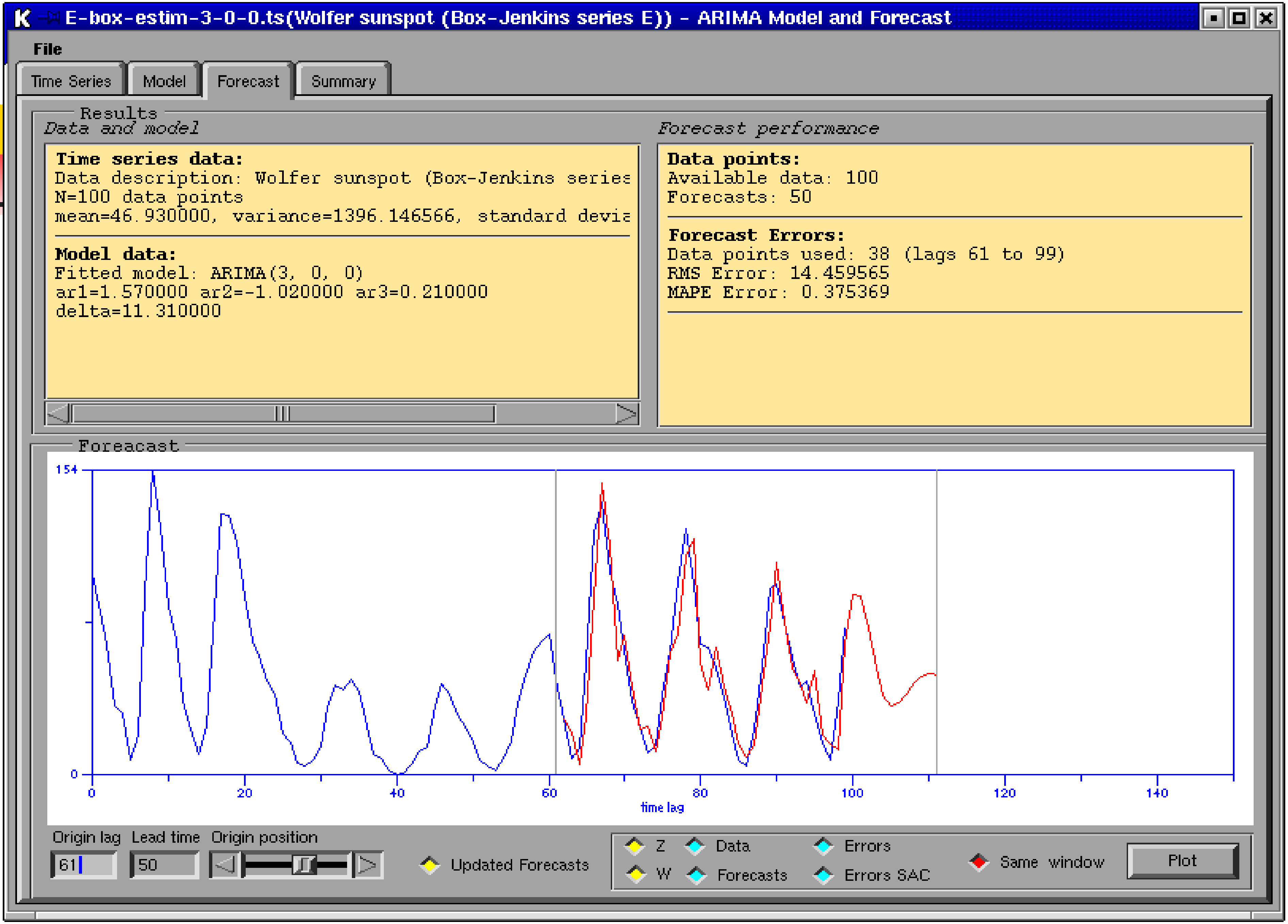
(**) módulos TSTOOL, TRSVIEW, FPARSER, ARIMA

Ferramenta TSTOOL/ARIMA

- Estágio de desenvolvimento – *alpha*.
- Screenshots do módulo TSTOOL/ARIMA
  - Time-series – seleção de *range*, estatísticas, SAC, SPAC.
  - Model – parâmetros, estimativas iniciais, estimativas finais, diagnóstico.
  - Forecast– previsão, erros e resíduos.


E-box-estim-3-0-0.ts(Wolfer sunspot (Box-Jenkins series E)) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Data selection
Original series (z) statistics
Z series -
Basic statistics:
N=100 data points
mean=46.930000, var=1396.146566, stddev=37.365045
y0=101.000000 y1=82.000000 y2=66.000000 y3=35.000
SACF-99 coefficients:
r1=0.806244 r2=0.428105 r3=0.069611 r4=-0.169423
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.806244| > se=0.196987
q=2 - |r2=0.428105| > se=0.230277
SPACF-99 coefficients:
r11=0.806244 r22=-0.634121 r33=0.080474 r44=-0.06
Model Order Hypothesis:
154
0
0
20
40
60
80
100
time lag
From lag
to lag
window
window position
0
99
100
Select
Reset
Reduction to stationarity
Reducted series (w) statistics
W series -
Basic statistics:
N=100 data points
mean=46.930000, var=1396.146566, stddev=37.365045
y0=101.000000 y1=82.000000 y2=66.000000 y3=35.000
SACF-99 coefficients:
r1=0.806244 r2=0.428105 r3=0.069611 r4=-0.169423
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.806244| > se=0.196987
q=2 - |r2=0.428105| > se=0.230277
SPACF-99 coefficients:
r11=0.806244 r22=-0.634121 r33=0.080474 r44=-0.06
Model Order Hypothesis:
Method
Differencing order-d
0
Detrending
Mean for W_tilde
0.000000
Apply
Plot
Z
W
Data
SAC
SPAC
Same window
Lags to plot
50
Plot


E-box-estim-3-0-0.ts(Wolfer sunspot (Box-Jenkins series E)) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
General ARIMA (Box-Jenkins) model
$\phi_p(B)\ \Phi_P(B^L)\ z_t = \delta + \theta_q(B)\ \Theta(B^L)\ a_t$
Realize process
Lags
0
Save process as TS
Model orders
Autoregressive order-p
3
Moving average order-q
0
d
0
A.R. seasonal order-P
0
M.A. seasonal order-Q
0
L
0
Non-seasonal
Seasonal
ϕ
θ
1.570000
-1.020000
0.210000
Residuals
218.0000
Calculate
Constant
δ
11.31000
Calculate
Initial estimates of parameters
AR parms estimate
Yuler-Walker equations
MA parms convergence
Linear
Quadratic
Initial Estimate
Results
ARIMA(3,0,0) parameters initial estimates
Procedure: initial estimate procedure not called,
w stats: n=100 w_mean=46.930000 var_w=1396.146566
stderr(w_mean)=3.563496 w_mean interval=[43.36650
ar1=1.570000 ar2=-1.020000 ar3=0.210000
AR(B) roots: z0=1.034354+0.798653j |z0|=1.306803 :
ALL roots lie out of unit circle
Estimation
Nonlinear (ML-Marquardt-Levemberg Solver)
ULS (Marquardt-Levemberg)
CLS
Estimate
Results
ARIMA(3,0,0) parameters final estimates
Procedure: estimate procedure not called, default pa
w stats: n=100 w_mean=46.930000 var_w=1396.146566
stderr(w_mean)=3.563496 w_mean interval=[43.366504,
ar1=1.570000 ar2=-1.020000 ar3=0.210000
AR(B) roots: z0=1.034354+0.798653j |z0|=1.306803 z1=
ALL roots lie out of unit circle
delta=11.310000
Sum of squares of a=22284.559126
var_a=218.381907
Information criteria: AIC=16.776245 BIC=5.570452
Diagnostic checking
Num. of autocorrelations to use
98
Check model
Plot residuals
Results
ARIMA(3,0,0) parameters final estimates
Roots of Characteristic Equations:
AR(B) roots: z0=1.034354+0.798653j |z0|=1.306803
ALL roots lie out of unit circle
Residuals(a)-100 values: mean=0.012409 var=218.38
a0=4.247303 a1=-16.239173 a2=16.413169 a3=-17.453
SACF of residuals (a)-98 coefficients used:
r0=1.000000 r1=-0.009550 r2=-0.109450 r3=0.044965
Checking criteria:
Max.stderr for AC-1/sqrt(100)=0.100000
Ljung-Box: Q=221.078566


E-box-estim-3-0-0.ts(Wolfer sunspot (Box-Jenkins series E)) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Results
Data and model
Time series data:
Data description: Wolfer sunspot (Box-Jenkins series
N=100 data points
mean=46.930000, variance=1396.146566, standard devia
Model data:
Fitted model: ARIMA(3, 0, 0)
ar1=1.570000 ar2=-1.020000 ar3=0.210000
delta=11.310000
Forecast performance
Data points:
Available data: 100
Forecasts: 50
Forecast Errors:
Data points used: 38 (lags 61 to 99)
RMS Error: 14.459565
MAPE Error: 0.375369
Foreacast
154
0
0
20
40
60
80
100
120
140
time lag
Origin lag
Lead time
Origin position
61
50
Updated Forecasts
Z
W
Data
Forecasts
Errors
Errors SAC
Same window
Plot

TSTOOL/ARIMA- *Benchmak* com os exemplos de Box-Jenkins

Resultados do módulo TSTOOL/ARIMA X Exemplos Box-Jenkins					
Série	***Modelo***	***Estimativas Iniciais***		***Estimativas***	
		Box-Jenkins	TSTOOL/ARIMA	Box-Jenkins	TSTOOL/ARIMA
Chemical proc. concentration (Box-Jenkins series A)	(1,0,1)	**ar1**=0.87 **ma1**=0.48 **delta**=2.45 **var_a**=0.098 **S**=19.270010	**ar1**=0.868278 **ma1**=0.480381 **delta**=2.247506 **var_a**=0.097630 **S**=19.270133	**ar1**=0.92 **ma1**=0.58 **delta**=1.45 **var_a**=0.097 **S**=19.182801	**ar1**=0.916566 **ma1**=0.609902 **delta**=1.423594 **var_a**=0.097067 **S**=19.175424
Chemical proc. concentration (Box-Jenkins series A)	(0,1,1)	**ma1**=0.53 **delta**=0.0 **var_a**=0.107 **S**=20.265666	**ma1**=0.528070 **delta**=0.002041 **var_a**=0.103100 **S**=20.276040	**ma1**=0.70 **delta**=0.0 **var_a**=0.101 **S**=19.719609	**ma1**=0.718569 **delta**=0.002041 **var_a**=0.002041 **S**=19.721280
IBM stock (Box-Jenkins series B)	(0,1,1)	**ma1**=-0.09 **delta**=0.0 **var_a**=52.2 **S**=19192.615141	**ma1**=-0.086212 **delta**=-0.279891 **var_a**=52.152651 **S**=19192.213953	**ma1**=-0.09 **delta**=0.0 **var_a**=52.2 **S**=19192.615141	**ma1**=-0.073683, **delta**=-0.279891 **var_a**=52.160028 **S**=19194.920355

TSTOOL/ARIMA- *Benchmak* com os exemplos de Box-Jenkins

Chemical temperature (Box-Jenkins series C)	(1,1,0)	**ar1**=0.81 **delta**=0.0 **var_a**=0.019 **S**=4.366651	**ar1**=0.805496 **delta**=-0.006743 **var_a**=0.017864 **S**=4.063117	**ar1**=0.82 **delta**=0.0 **var_a**=0.018 **S**=4.060537	**ar1**=0.825973, **delta**=-0.006033 **var_a**=0.017863 **S**=4.060911
Chemical temperature (Box-Jenkins series C)	(0,2,2)	**ma1**=0.09 **ma2**=0.07 **delta**=0.0 **var_a**=0.020 **S**=4.061787	**ma1**=0.085572 **ma2**=0.066229 **delta**=-0.002679 **var_a**=0.019496 **S**=4.368707	**ma1**=0.13 **ma2**=0.12 **delta**=0.0 **var_a**=0.019 **S**=4.355676	**ma1**=0.130983, **ma2**=0.122996, **delta**=-0.002679 **var_a**=0.019423 **S**=4.355705

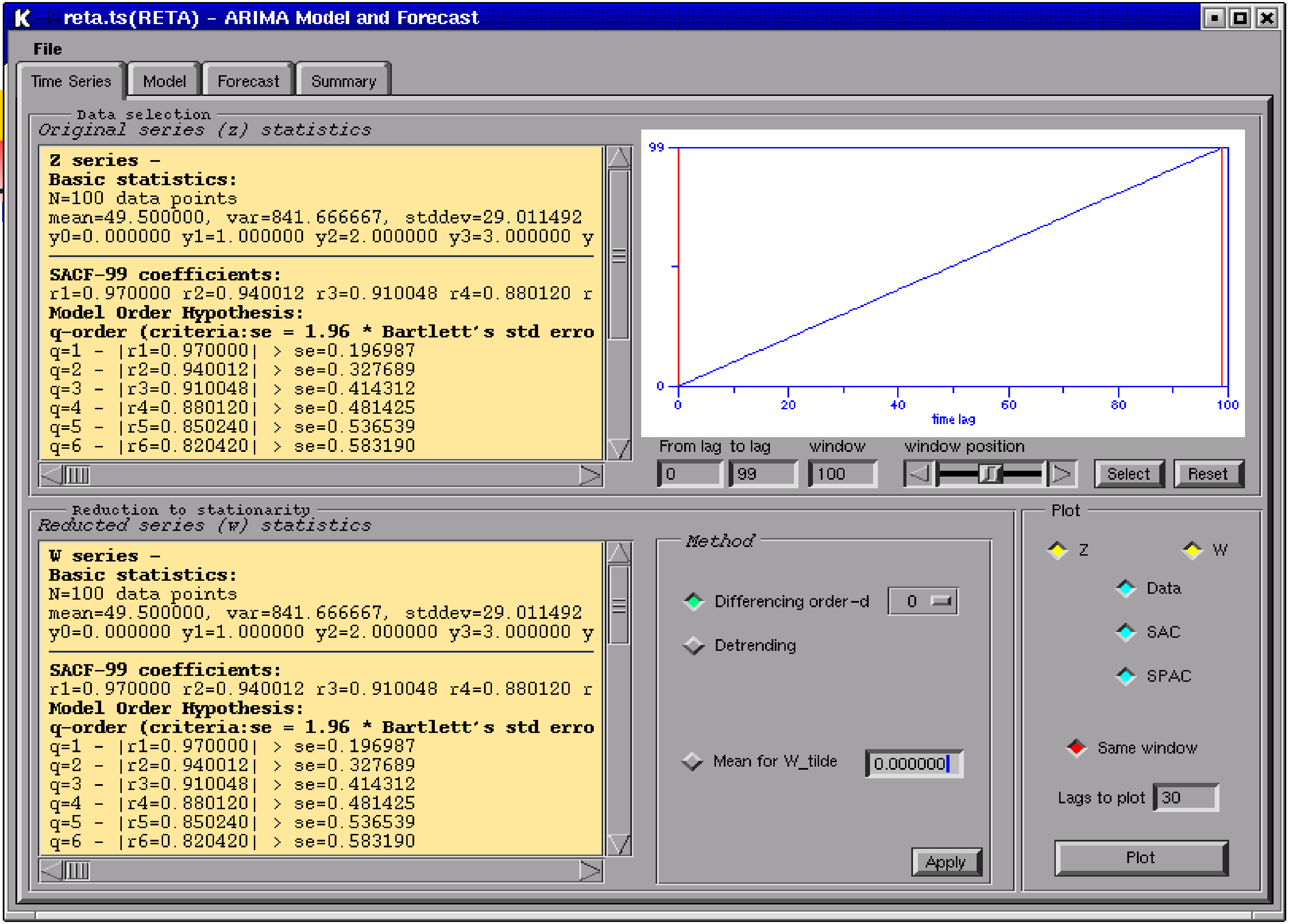
Chemical viscosity (Box-Jenkins series D)	(1,0,0)	**ar1**=0.86 **delta**=1.32 **var_a**=0.093 **S**=27.990700	**ar1**=0.861462 **delta**=1.265212 **var_a**=0.089483 **S**=27.987230	**ar1**=0.87 **delta**=1.17 **var_a**=0.090 **S**=27.976432	**ar1**=0.876410, **delta**=1.128692 **var_a**=0.089510 **S**=27.978955
Chemical viscosity (Box-Jenkins series D)	(0,1,1)	**ma1**=0.05 **delta**=0.0 **var_a**=0.096 **S**=29.718366	**ma1**=0.051262 **delta**=0.003560 **var_a**=0.096174 **S**=29.717809	**ma1**=0.06 **delta**=0.0 **var_a**=0.096 **S**=29.716169	**ma1**=0.059482, **delta**=0.003560 **var_a**=0.096169 **S**=29.716159

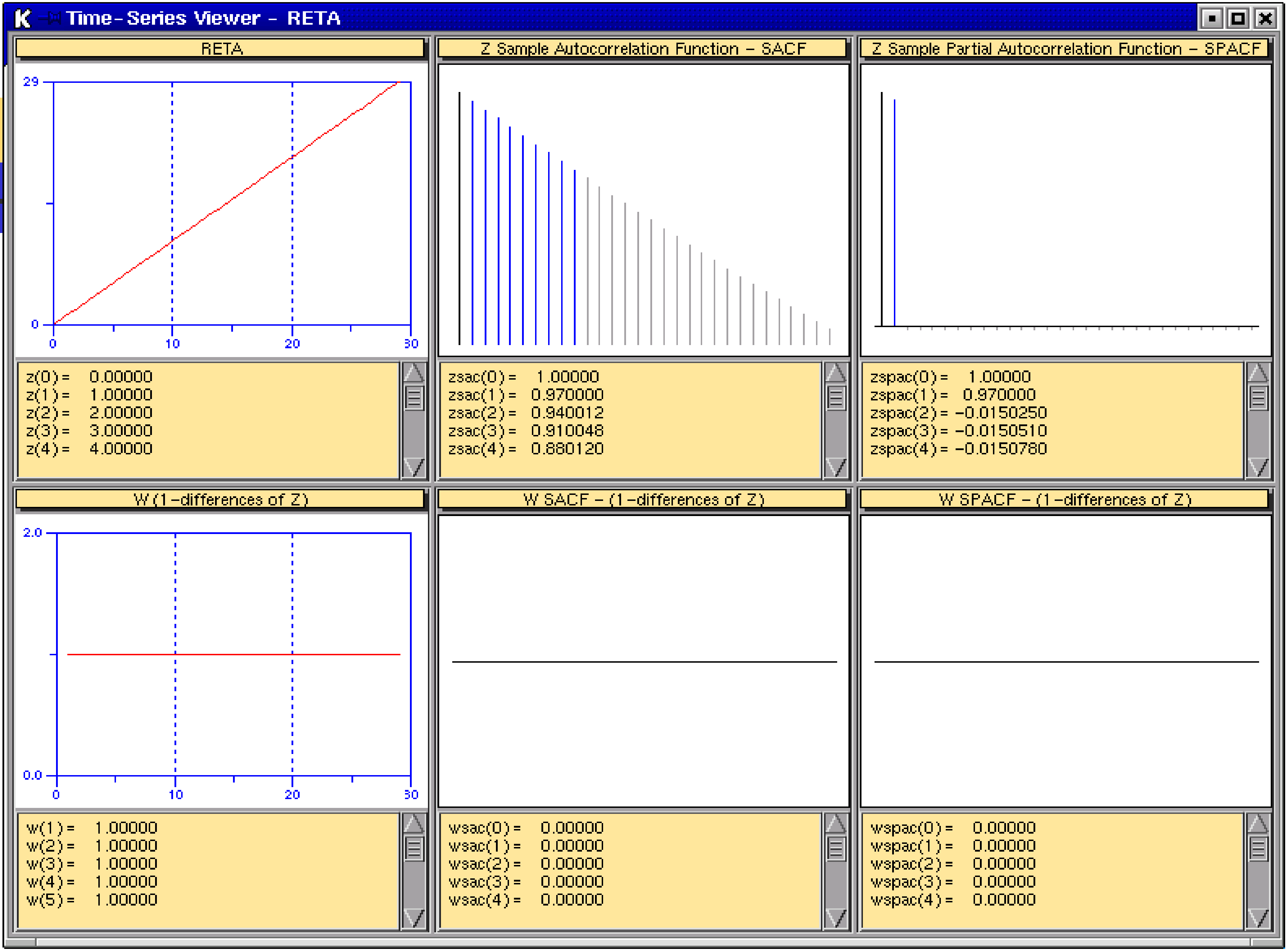
TSTOOL/ARIMA- *Benchmak* com os exemplos de Box-Jenkins

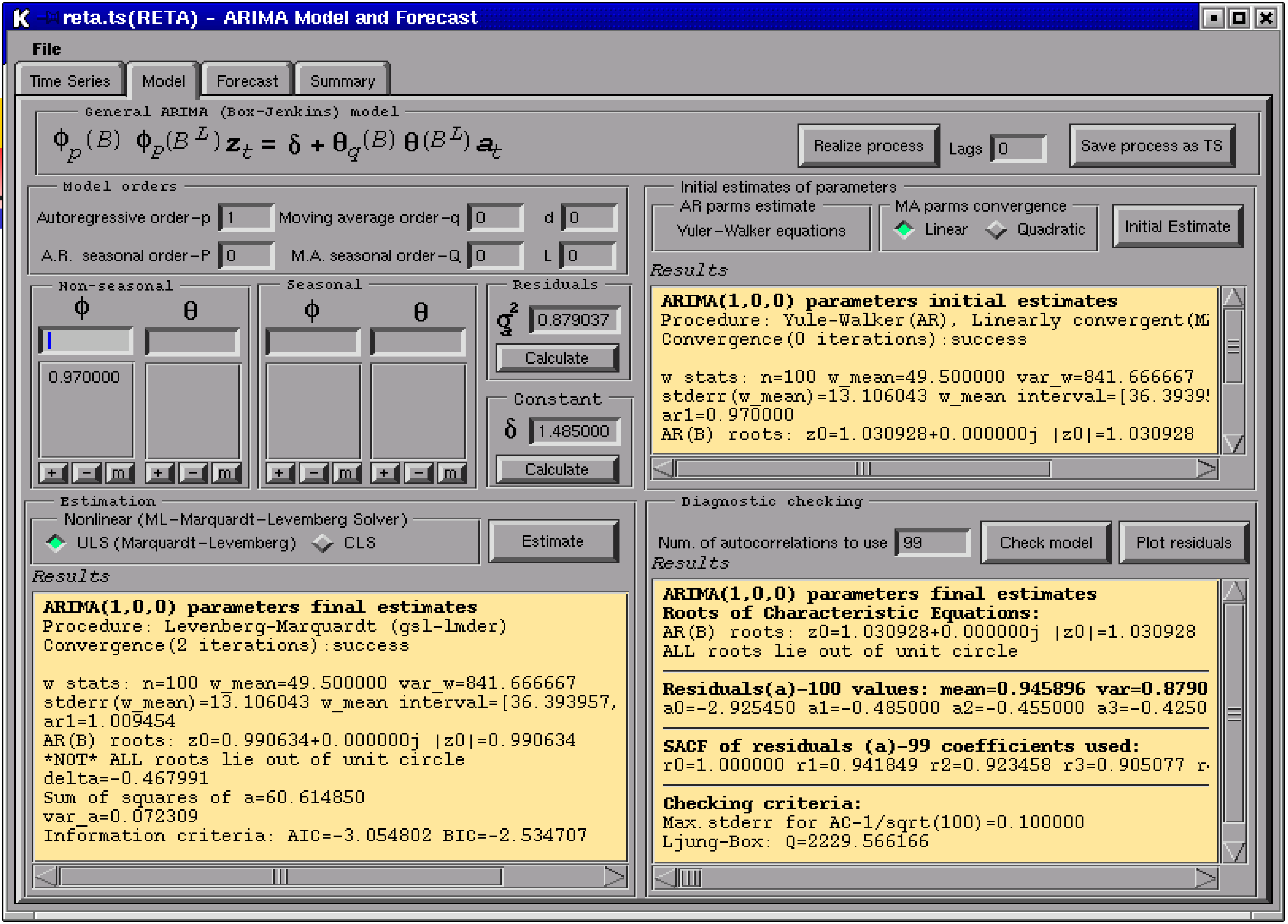
Wolfer sunspot (Box-Jenkins series E)	(2,0,0)	**ar1**=1.32 **ar2**=-0.63 **delta**=14.9 **var_a**=289.0 **S**=23538.354642	**ar1**=0.806244 **ar2**=-0.221924 **delta**=19.507863 **var_a**=388.158329 **S**=39547.779856	**ar1**=1.42 **ar2**=-0.73 **delta**=14.35 **var_a**=228.0 **S**=23085.065862	**ar1**=1.410409, **ar2**=-0.700836, **delta**=13.629755 **var_a**=226.406298 **S**=23096.631910
Wolfer sunspot (Box-Jenkins series E)	(3,0,0)	**ar1**=1.37 **ar2**=-0.74 **ar3**=0.08 **delta**=13.7 **var_a**=287.0 **S**=23048.543429	**ar1**=0.806244 **ar2**=-0.221924 **ar3**=-0.096621 **delta**=24.042304 **var_a**=373.486098 **S**=38054.067240	**ar1**=1.57 **ar2**=-1.02 **ar3**=0.21 **delta**=11.31 **var_a**=218.0 **S**=22284.559126	**ar1**=1.547141, **ar2**=-0.973827, **ar3**=0.189886, **delta**=11.113035 **var_a**=218.056043 **S**=22264.205800
Chemical yields (Box-Jenkins series F)	(2,0,0)	**ar1**=-0.32 **ar2**=-0.18 **delta**=58.3 **var_a**=115.0 **S**=7923.092342	**ar1**=-0.389878 **ar2**=-0.152389 **delta**=63.271062 **var_a**=113.105096 **S**=7928.507849	**ar1**=-0.34 **ar2**=-0.19 **delta**=58.87 **var_a**=113.0 **S**=7916.294670	**ar1**=-0.348361, **ar2**=0.193983, **delta**=59.021704 **var_a**=112.913167 **S**=7916.976328

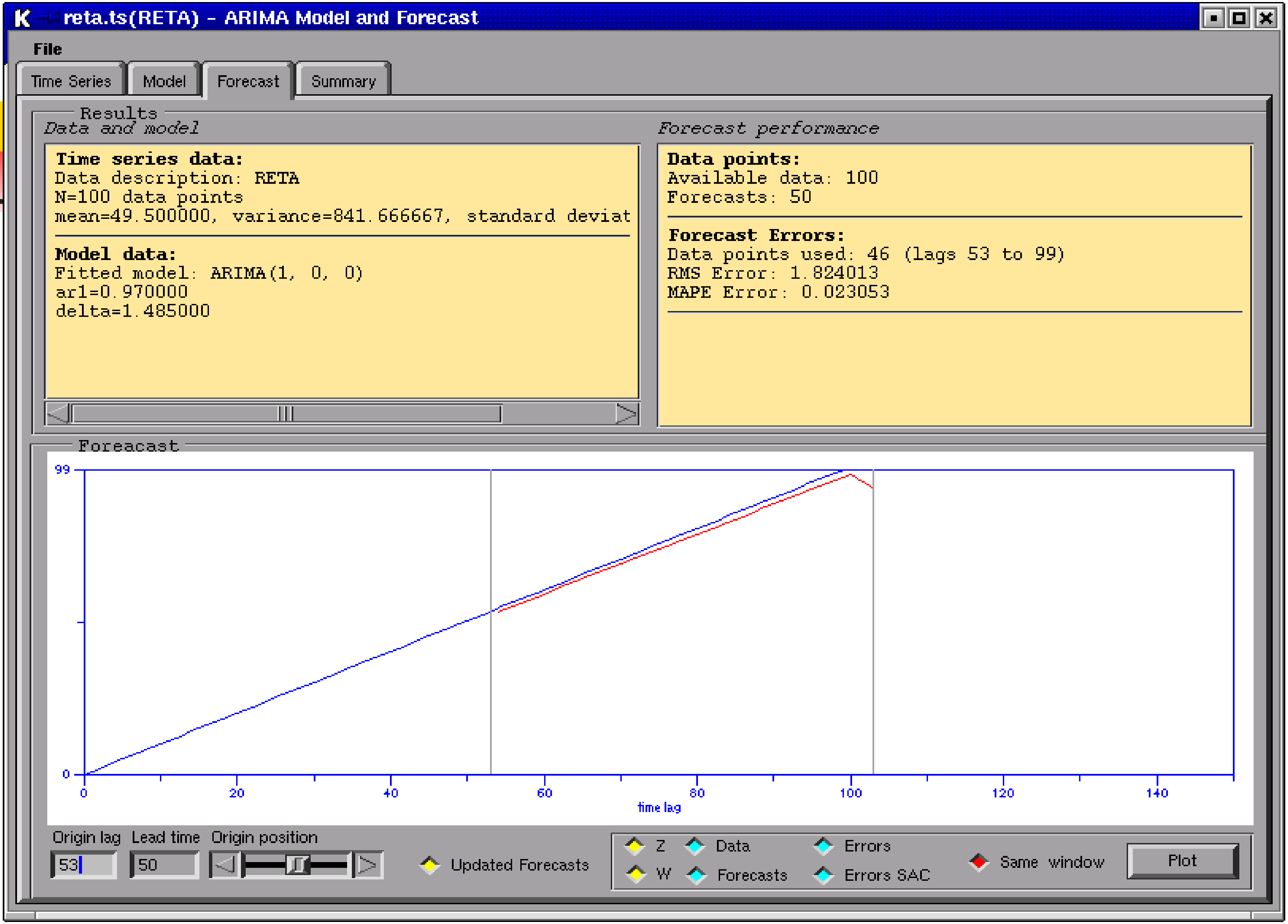
TSTOOL/ARIMA- Exemplos de modelagem

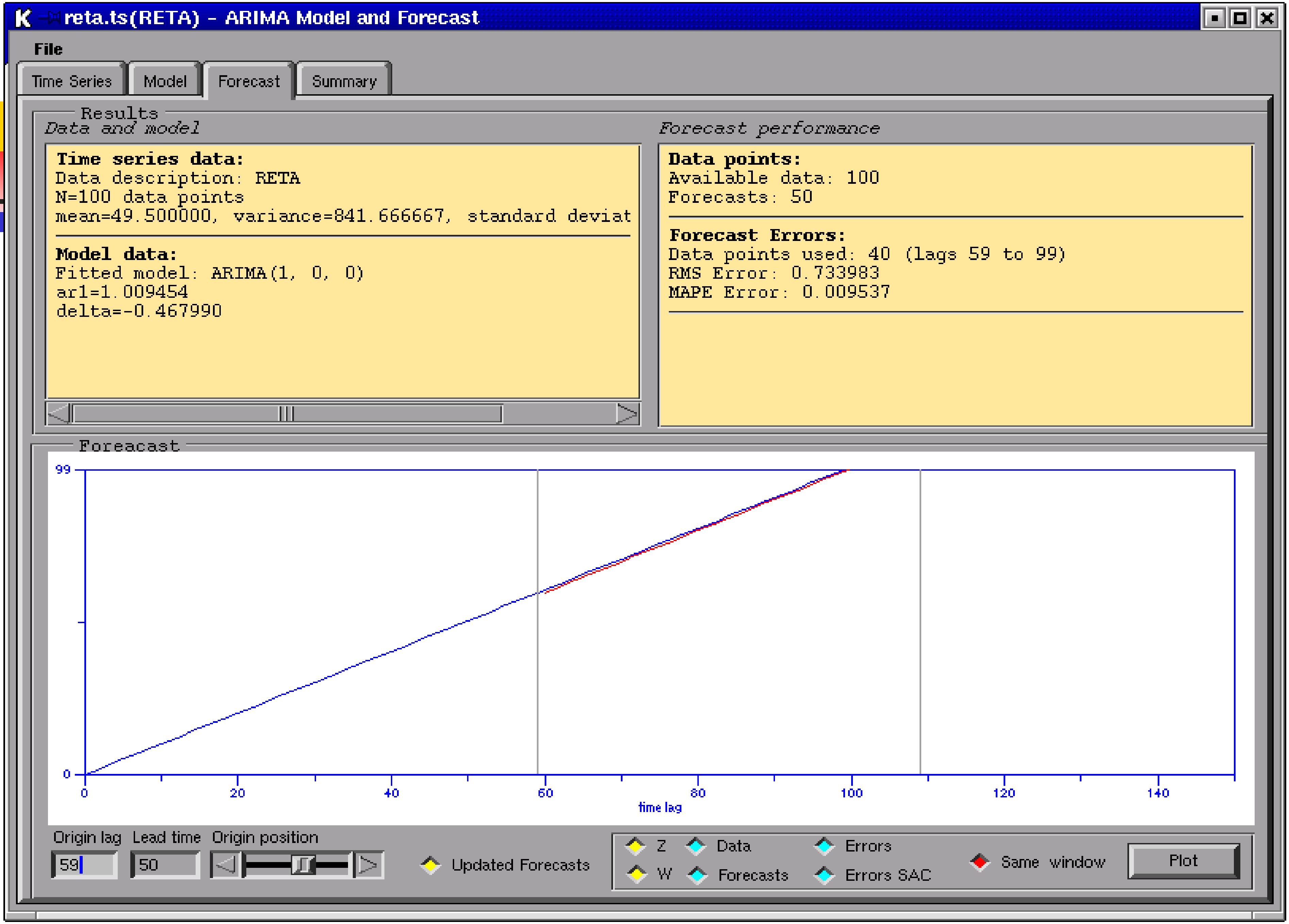
- Reta
- Série triangular.
- Cotações de COFAP (FAP4)

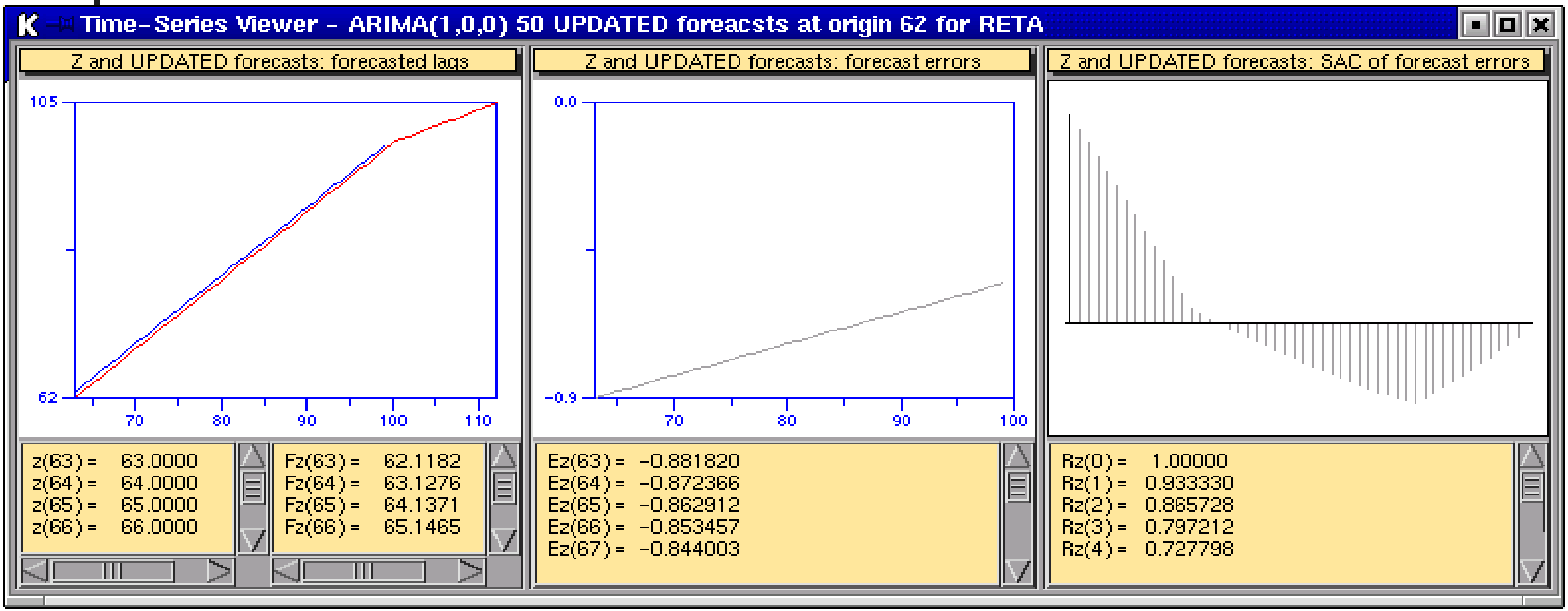

reta.ts(RETA) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Data selection
Original series (z) statistics
Z series -
Basic statistics:
N=100 data points
mean=49.500000, var=841.666667, stddev=29.011492
y0=0.000000 y1=1.000000 y2=2.000000 y3=3.000000 y
SACF-99 coefficients:
r1=0.970000 r2=0.940012 r3=0.910048 r4=0.880120 r
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.970000| > se=0.196987
q=2 - |r2=0.940012| > se=0.327689
q=3 - |r3=0.910048| > se=0.414312
q=4 - |r4=0.880120| > se=0.481425
q=5 - |r5=0.850240| > se=0.536539
q=6 - |r6=0.820420| > se=0.583190
99
0
0
20
40
60
80
100
time lag
From lag
to lag
window
window position
0
99
100
Select
Reset
Reduction to stationarity
Reducted series (w) statistics
W series -
Basic statistics:
N=100 data points
mean=49.500000, var=841.666667, stddev=29.011492
y0=0.000000 y1=1.000000 y2=2.000000 y3=3.000000 y
SACF-99 coefficients:
r1=0.970000 r2=0.940012 r3=0.910048 r4=0.880120 r
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.970000| > se=0.196987
q=2 - |r2=0.940012| > se=0.327689
q=3 - |r3=0.910048| > se=0.414312
q=4 - |r4=0.880120| > se=0.481425
q=5 - |r5=0.850240| > se=0.536539
q=6 - |r6=0.820420| > se=0.583190
Method
Differencing order-d
0
Detrending
Mean for W_tilde
0.000000
Apply
Plot
Z
W
Data
SAC
SPAC
Same window
Lags to plot
30
Plot

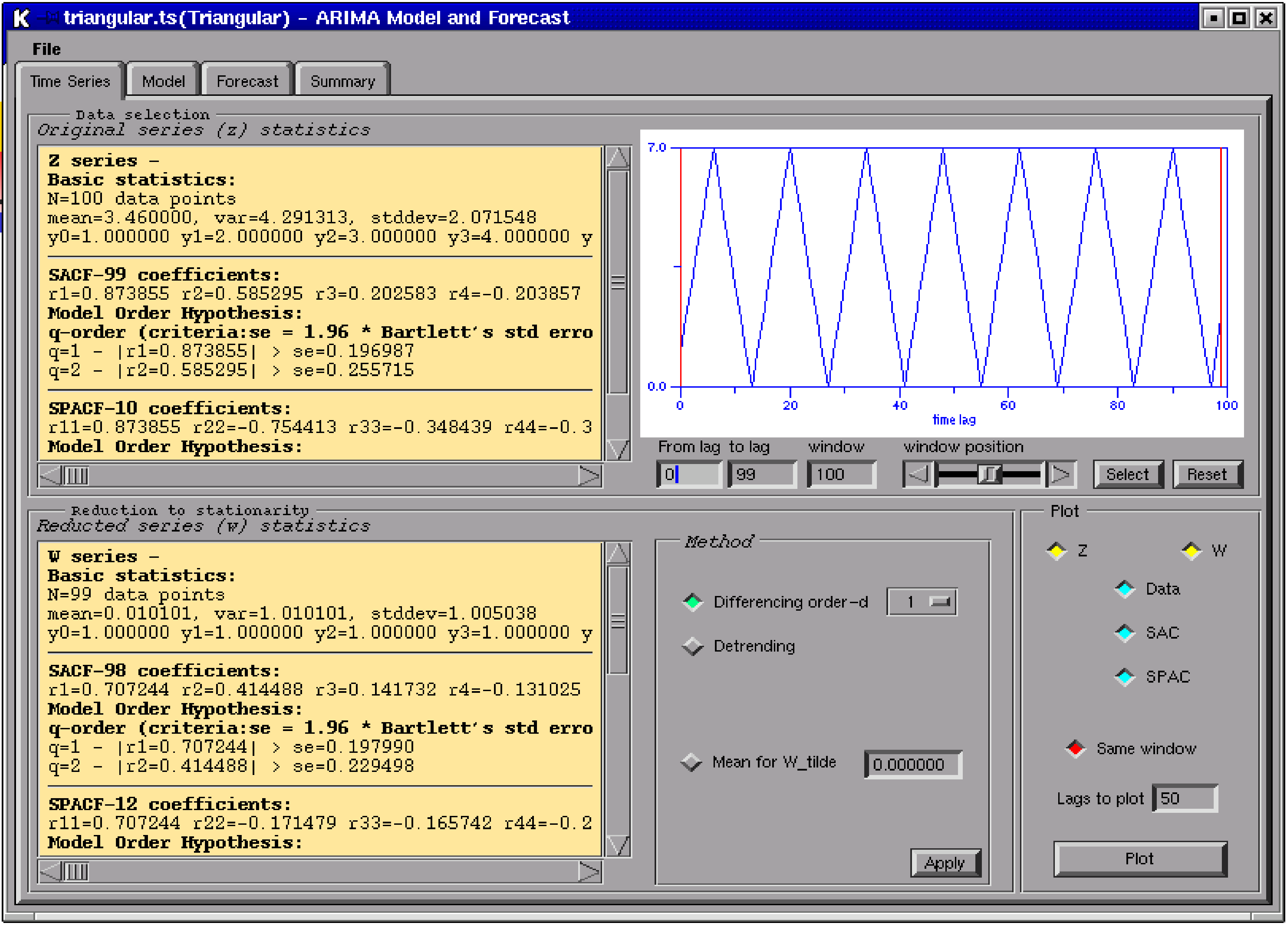

Time-Series Viewer - RETA
RETA
29
0
0
10
20
30
z(0) = 0.00000
z(1) = 1.00000
z(2) = 2.00000
z(3) = 3.00000
z(4) = 4.00000
Z Sample Autocorrelation Function – SACF
zsac(0) = 1.00000
zsac(1) = 0.970000
zsac(2) = 0.940012
zsac(3) = 0.910048
zsac(4) = 0.880120
Z Sample Partial Autocorrelation Function – SPACF
zspac(0) = 1.00000
zspac(1) = 0.970000
zspac(2) = –0.0150250
zspac(3) = –0.0150510
zspac(4) = –0.0150780
W (1–differences of Z)
2.0
0.0
0
10
20
30
w(1) = 1.00000
w(2) = 1.00000
w(3) = 1.00000
w(4) = 1.00000
w(5) = 1.00000
W SACF – (1–differences of Z)
wsac(0) = 0.00000
wsac(1) = 0.00000
wsac(2) = 0.00000
wsac(3) = 0.00000
wsac(4) = 0.00000
W SPACF – (1–differences of Z)
wspac(0) = 0.00000
wspac(1) = 0.00000
wspac(2) = 0.00000
wspac(3) = 0.00000
wspac(4) = 0.00000

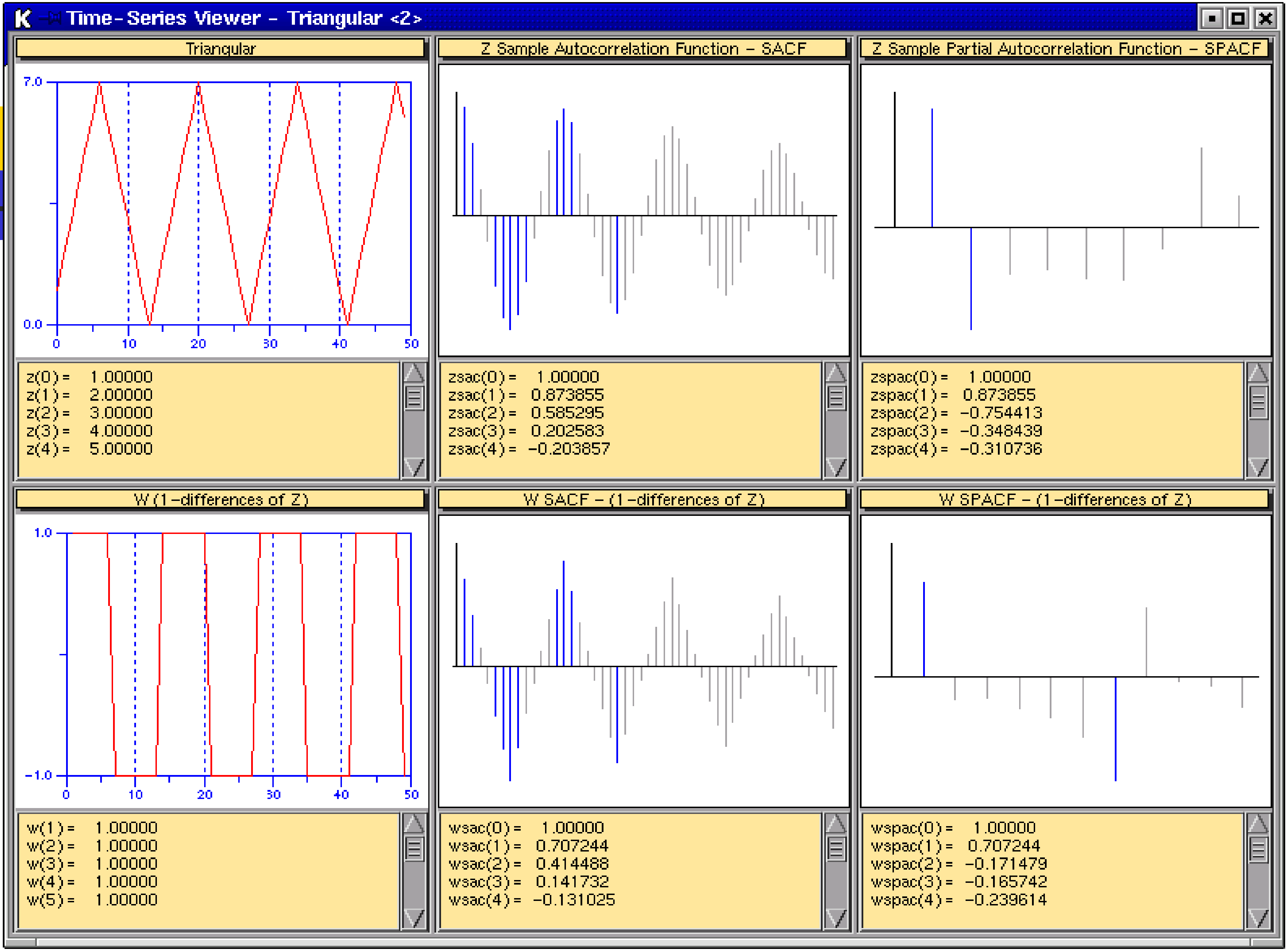

reta.ts(RETA) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
General ARIMA (Box-Jenkins) model
$\phi_p(B)\ \Phi_P(B^L)\,z_t = \delta + \theta_q(B)\ \Theta(B^L)\,a_t$
Realize process
Lags
0
Save process as TS
Model orders
Autoregressive order-p
1
Moving average order-q
0
d
0
A.R. seasonal order-P
0
M.A. seasonal order-Q
0
L
0
Non-seasonal
φ
θ
0.970000
Seasonal
φ
θ
Residuals
σ^2
0.879037
Calculate
Constant
δ
1.485000
Calculate
Initial estimates of parameters
AR parms estimate
Yuler-Walker equations
MA parms convergence
Linear
Quadratic
Initial Estimate
Results
ARIMA(1,0,0) parameters initial estimates
Procedure: Yule-Walker(AR), Linearly convergent(M
Convergence(0 iterations):success
w stats: n=100 w_mean=49.500000 var_w=841.666667
stderr(w_mean)=13.106043 w_mean interval=[36.3939
ar1=0.970000
AR(B) roots: z0=1.030928+0.000000j |z0|=1.030928
Estimation
Nonlinear (ML-Marquardt-Levemberg Solver)
ULS (Marquardt-Levemberg)
CLS
Estimate
Results
ARIMA(1,0,0) parameters final estimates
Procedure: Levenberg-Marquardt (gsl-lmder)
Convergence(2 iterations):success
w stats: n=100 w_mean=49.500000 var_w=841.666667
stderr(w_mean)=13.106043 w_mean interval=[36.393957,
ar1=1.009454
AR(B) roots: z0=0.990634+0.000000j |z0|=0.990634
NOT ALL roots lie out of unit circle
delta=-0.467991
Sum of squares of a=60.614850
var_a=0.072309
Information criteria: AIC=-3.054802 BIC=-2.534707
Diagnostic checking
Num. of autocorrelations to use
99
Check model
Plot residuals
Results
ARIMA(1,0,0) parameters final estimates
Roots of Characteristic Equations:
AR(B) roots: z0=1.030928+0.000000j |z0|=1.030928
ALL roots lie out of unit circle
Residuals(a)-100 values: mean=0.945896 var=0.8790
a0=-2.925450 a1=-0.485000 a2=-0.455000 a3=-0.4250
SACF of residuals (a)-99 coefficients used:
r0=1.000000 r1=0.941849 r2=0.923458 r3=0.905077 r
Checking criteria:
Max.stderr for AC-1/sqrt(100)=0.100000
Ljung-Box: Q=2229.566166

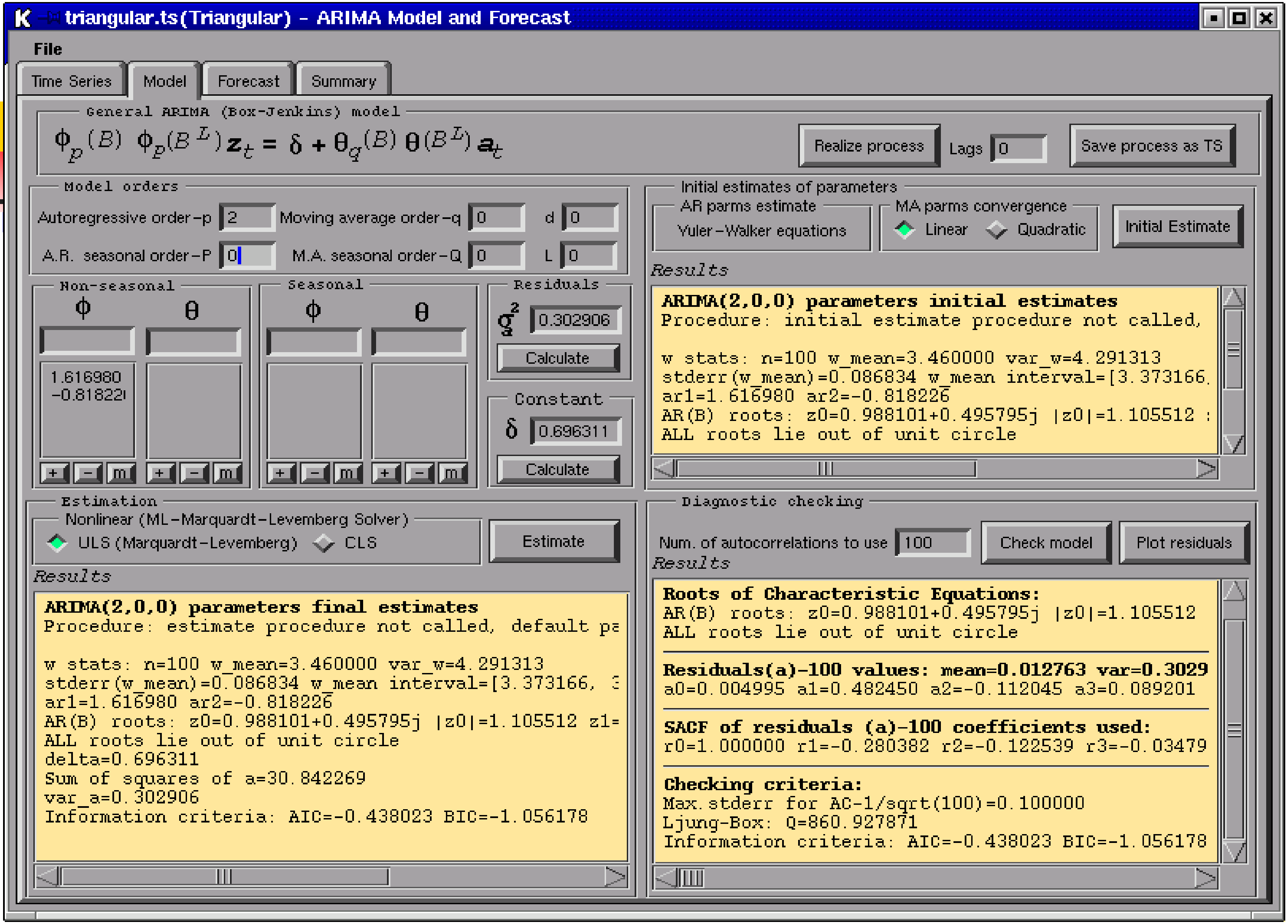

reta.ts(RETA) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Results
Data and model
Time series data:
Data description: RETA
N=100 data points
mean=49.500000, variance=841.666667, standard deviat
Model data:
Fitted model: ARIMA(1, 0, 0)
ar1=0.970000
delta=1.485000
Forecast performance
Data points:
Available data: 100
Forecasts: 50
Forecast Errors:
Data points used: 46 (lags 53 to 99)
RMS Error: 1.824013
MAPE Error: 0.023053
Foreacast
99
0
0
20
40
60
80
100
120
140
time lag
Origin lag
Lead time
Origin position
53
50
Updated Forecasts
Z
W
Data
Forecasts
Errors
Errors SAC
Same window
Plot

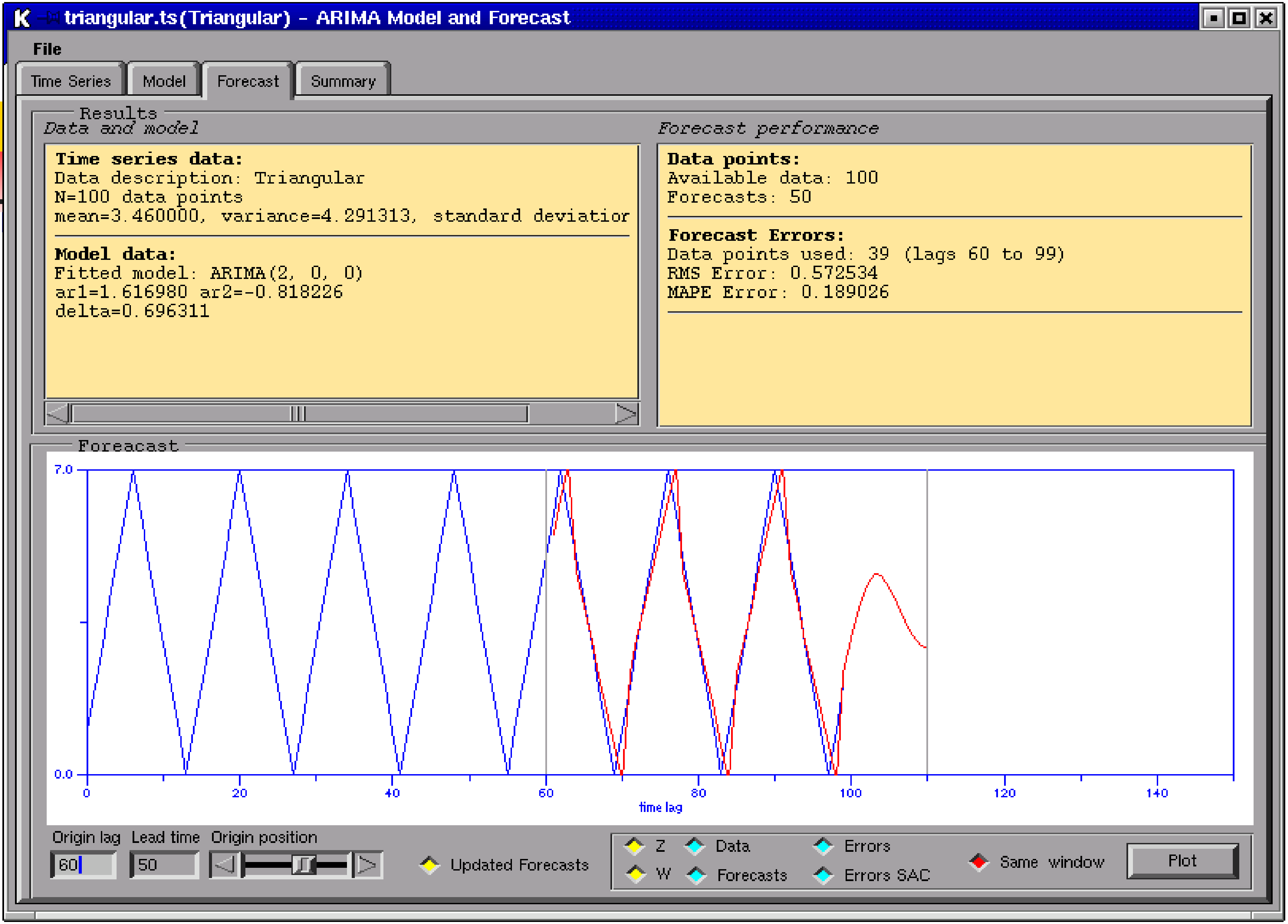

reta.ts(RETA) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Results
Data and model
Time series data:
Data description: RETA
N=100 data points
mean=49.500000, variance=841.666667, standard deviat
Model data:
Fitted model: ARIMA(1, 0, 0)
ar1=1.009454
delta=-0.467990
Forecast performance
Data points:
Available data: 100
Forecasts: 50
Forecast Errors:
Data points used: 40 (lags 59 to 99)
RMS Error: 0.733983
MAPE Error: 0.009537
Foreacast
99
0
0
20
40
60
80
100
120
140
time lag
Origin lag
Lead time
Origin position
59
50
Updated Forecasts
Z
W
Data
Forecasts
Errors
Errors SAC
Same window
Plot

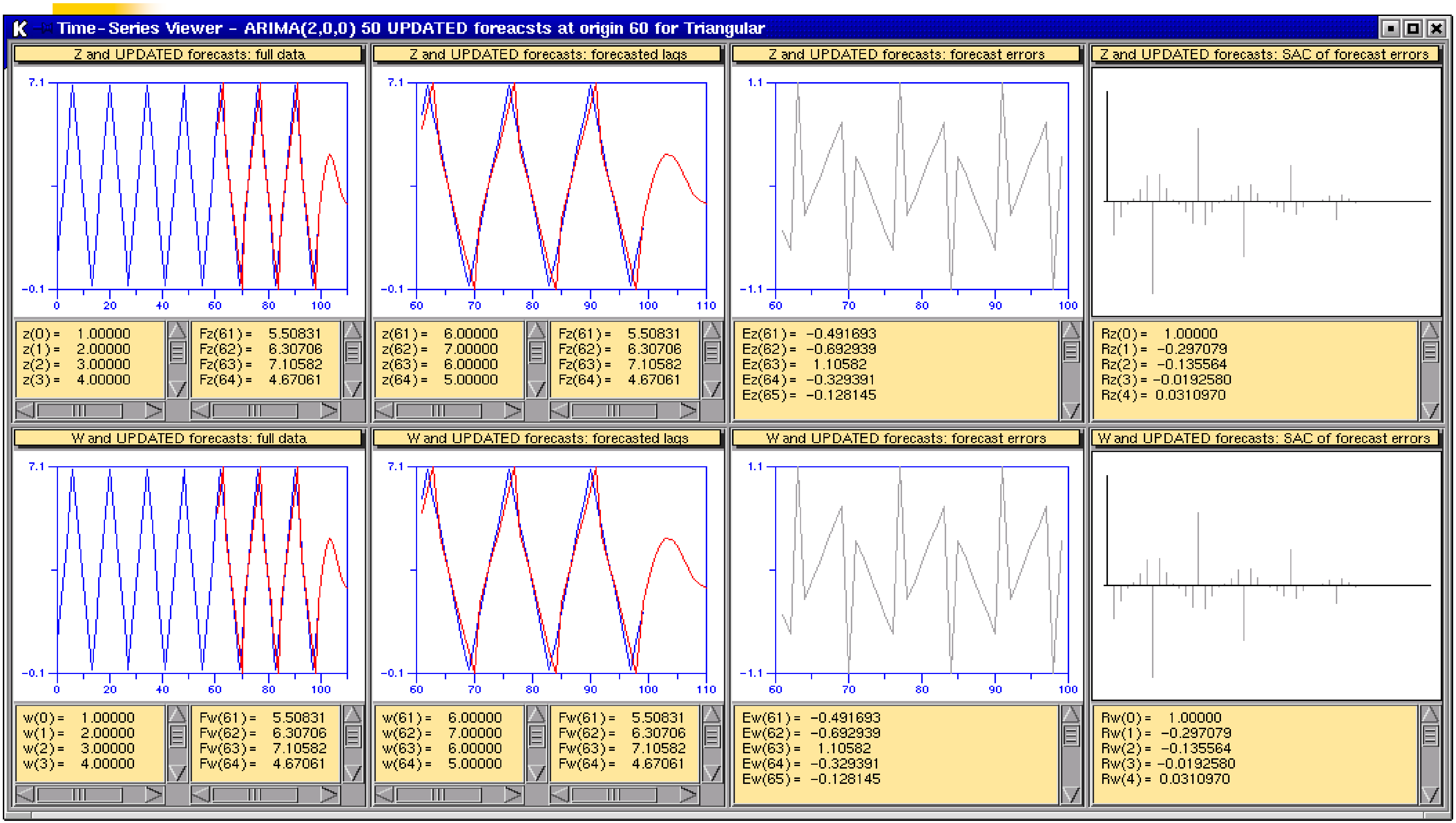

Time-Series Viewer - ARIMA(1,0,0) 50 UPDATED foreacsts at origin 62 for RETA
Z and UPDATED forecasts: forecasted lags
Z and UPDATED forecasts: forecast errors
Z and UPDATED forecasts: SAC of forecast errors
105
62
70
80
90
100
110
0.0
-0.9
70
80
90
100
z(63) = 63.0000
z(64) = 64.0000
z(65) = 65.0000
z(66) = 66.0000
Fz(63) = 62.1182
Fz(64) = 63.1276
Fz(65) = 64.1371
Fz(66) = 65.1465
Ez(63) = -0.881820
Ez(64) = -0.872366
Ez(65) = -0.862912
Ez(66) = -0.853457
Ez(67) = -0.844003
Rz(0) = 1.00000
Rz(1) = 0.933330
Rz(2) = 0.865728
Rz(3) = 0.797212
Rz(4) = 0.727798

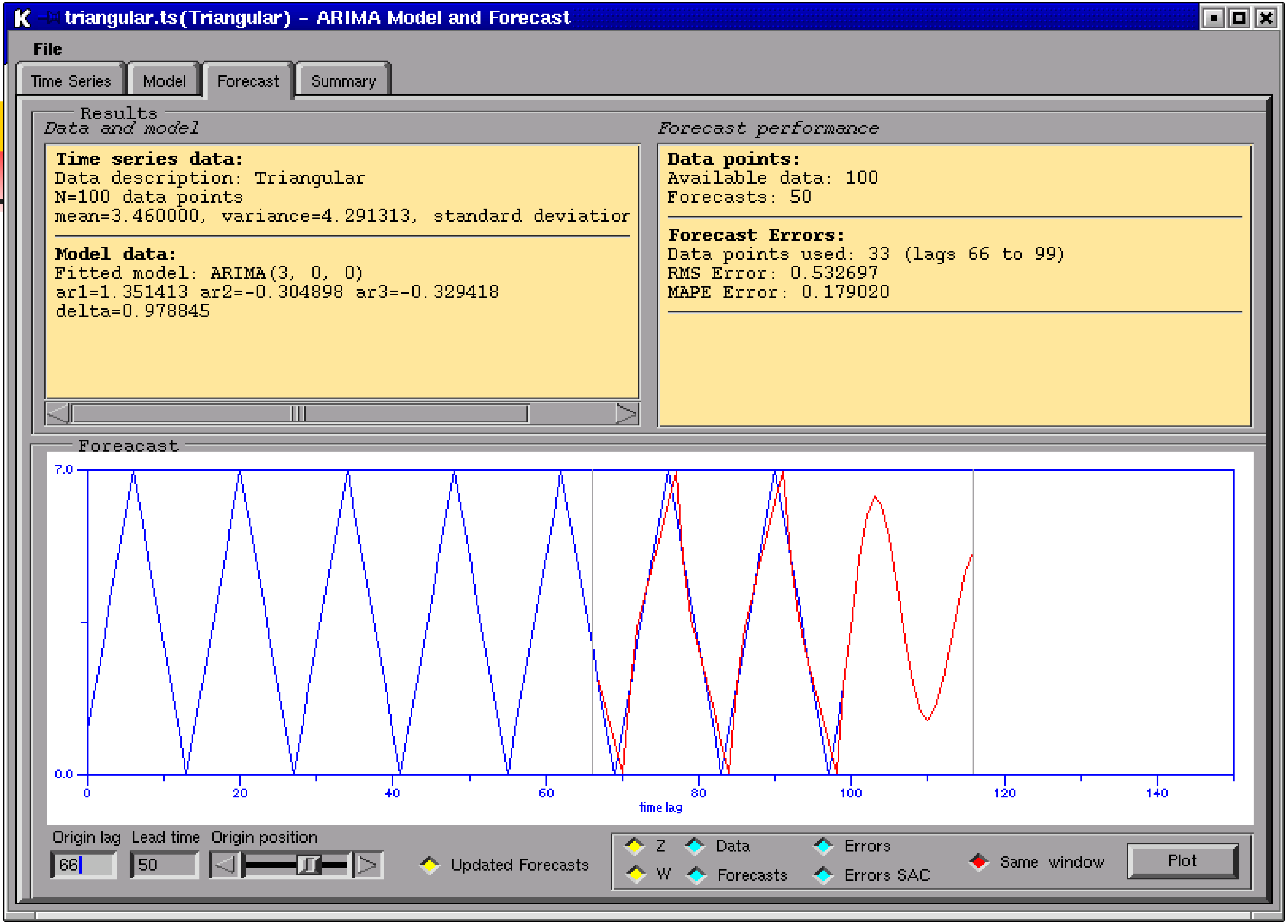

triangular.ts(Triangular) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Data selection
Original series (z) statistics
Z series -
Basic statistics:
N=100 data points
mean=3.460000, var=4.291313, stddev=2.071548
y0=1.000000 y1=2.000000 y2=3.000000 y3=4.000000 y
SACF-99 coefficients:
r1=0.873855 r2=0.585295 r3=0.202583 r4=-0.203857
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.873855| > se=0.196987
q=2 - |r2=0.585295| > se=0.255715
SPACF-10 coefficients:
r11=0.873855 r22=-0.754413 r33=-0.348439 r44=-0.3
Model Order Hypothesis:
7.0
0.0
0
20
40
60
80
100
time lag
From lag
to lag
window
window position
0
99
100
Select
Reset
Reduction to stationarity
Reducted series (w) statistics
W series -
Basic statistics:
N=99 data points
mean=0.010101, var=1.010101, stddev=1.005038
y0=1.000000 y1=1.000000 y2=1.000000 y3=1.000000 y
SACF-98 coefficients:
r1=0.707244 r2=0.414488 r3=0.141732 r4=-0.131025
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.707244| > se=0.197990
q=2 - |r2=0.414488| > se=0.229498
SPACF-12 coefficients:
r11=0.707244 r22=-0.171479 r33=-0.165742 r44=-0.2
Model Order Hypothesis:
Method
Differencing order-d
1
Detrending
Mean for W_tilde
0.000000
Apply
Plot
Z
W
Data
SAC
SPAC
Same window
Lags to plot
50
Plot

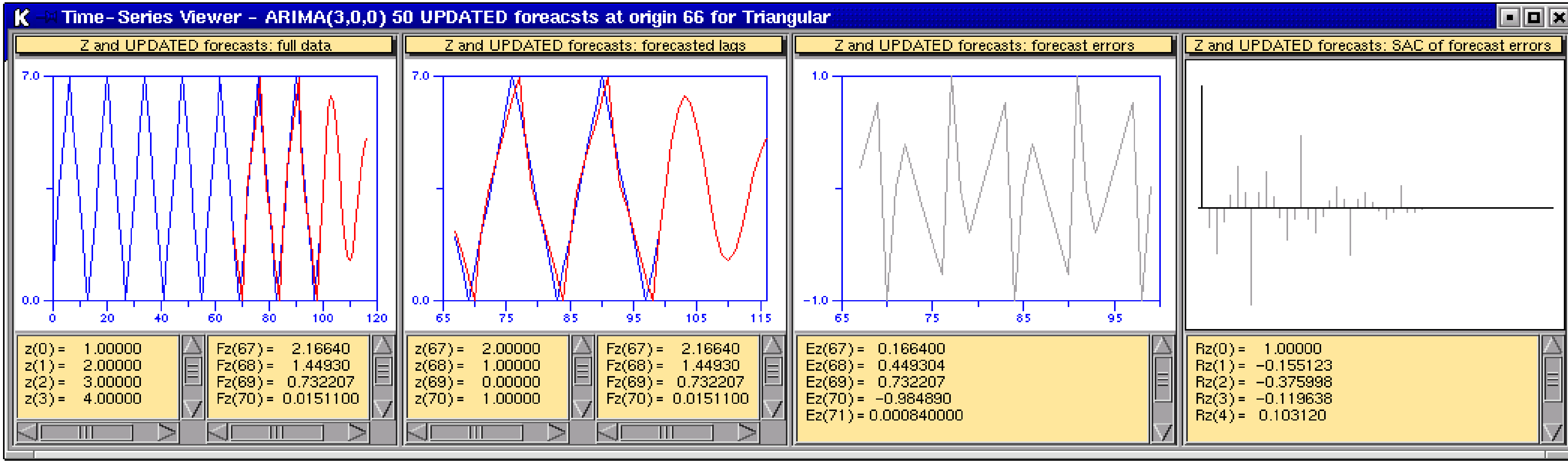

Time-Series Viewer - Triangular <2>
Triangular
7.0
0.0
0 10 20 30 40 50
Z Sample Autocorrelation Function – SACF
Z Sample Partial Autocorrelation Function – SPACF
z(0) = 1.00000
z(1) = 2.00000
z(2) = 3.00000
z(3) = 4.00000
z(4) = 5.00000
zsac(0) = 1.00000
zsac(1) = 0.873855
zsac(2) = 0.585295
zsac(3) = 0.202583
zsac(4) = −0.203857
zspac(0) = 1.00000
zspac(1) = 0.873855
zspac(2) = −0.754413
zspac(3) = −0.348439
zspac(4) = −0.310736
W (1−differences of Z)
1.0
−1.0
0 10 20 30 40 50
W SACF – (1−differences of Z)
W SPACF – (1−differences of Z)
w(1) = 1.00000
w(2) = 1.00000
w(3) = 1.00000
w(4) = 1.00000
w(5) = 1.00000
wsac(0) = 1.00000
wsac(1) = 0.707244
wsac(2) = 0.414488
wsac(3) = 0.141732
wsac(4) = −0.131025
wspac(0) = 1.00000
wspac(1) = 0.707244
wspac(2) = −0.171479
wspac(3) = −0.165742
wspac(4) = −0.239614

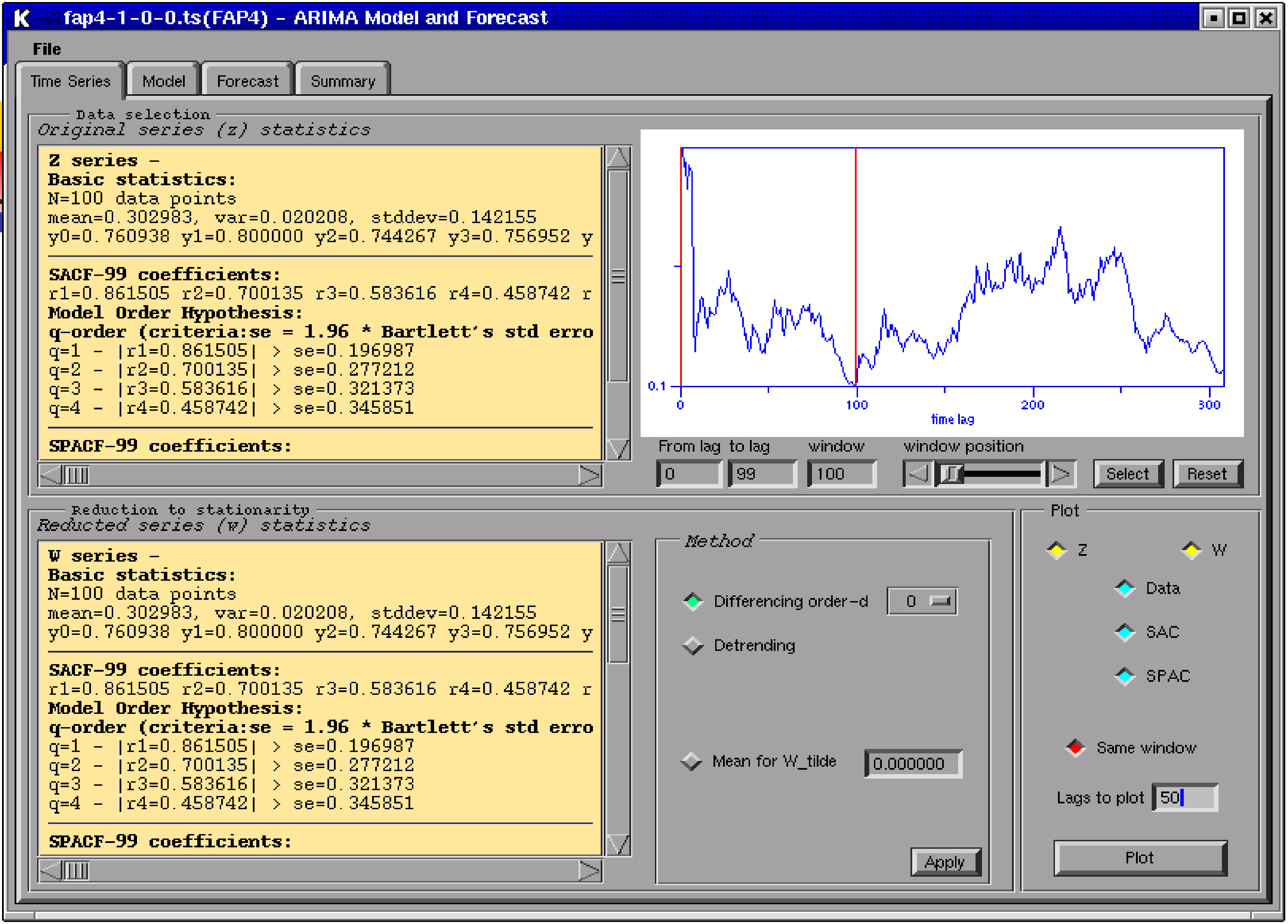

triangular.ts (Triangular) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
General ARIMA (Box-Jenkins) model
$\phi_p(B)\ \Phi_P(B^L)\, z_t = \delta + \theta_q(B)\ \Theta(B^L)\, a_t$
Realize process
Lags 0
Save process as TS
Model orders
Autoregressive order-p 2
Moving average order-q 0
d 0
A.R. seasonal order-P 0
M.A. seasonal order-Q 0
L 0
Non-seasonal
Seasonal
Residuals
σ^2 0.302906
Calculate
Constant
δ 0.696311
Calculate
1.616980
-0.81822
Initial estimates of parameters
AR parms estimate
Yuler-Walker equations
MA parms convergence
Linear
Quadratic
Initial Estimate
Results
ARIMA(2,0,0) parameters initial estimates
Procedure: initial estimate procedure not called,
w stats: n=100 w_mean=3.460000 var_w=4.291313
stderr(w_mean)=0.086834 w_mean interval=[3.373166,
ar1=1.616980 ar2=-0.818226
AR(B) roots: z0=0.988101+0.495795j |z0|=1.105512 :
ALL roots lie out of unit circle
Estimation
Nonlinear (ML-Marquardt-Levemberg Solver)
ULS (Marquardt-Levemberg)
CLS
Estimate
Results
ARIMA(2,0,0) parameters final estimates
Procedure: estimate procedure not called, default pa
w stats: n=100 w_mean=3.460000 var_w=4.291313
stderr(w_mean)=0.086834 w_mean interval=[3.373166,
ar1=1.616980 ar2=-0.818226
AR(B) roots: z0=0.988101+0.495795j |z0|=1.105512 z1=
ALL roots lie out of unit circle
delta=0.696311
Sum of squares of a=30.842269
var_a=0.302906
Information criteria: AIC=-0.438023 BIC=-1.056178
Diagnostic checking
Num. of autocorrelations to use 100
Check model
Plot residuals
Results
Roots of Characteristic Equations:
AR(B) roots: z0=0.988101+0.495795j |z0|=1.105512
ALL roots lie out of unit circle
Residuals(a)-100 values: mean=0.012763 var=0.3029
a0=0.004995 a1=0.482450 a2=-0.112045 a3=0.089201
SACF of residuals (a)-100 coefficients used:
r0=1.000000 r1=-0.280382 r2=-0.122539 r3=-0.03479
Checking criteria:
Max.stderr for AC-1/sqrt(100)=0.100000
Ljung-Box: Q=860.927871
Information criteria: AIC=-0.438023 BIC=-1.056178

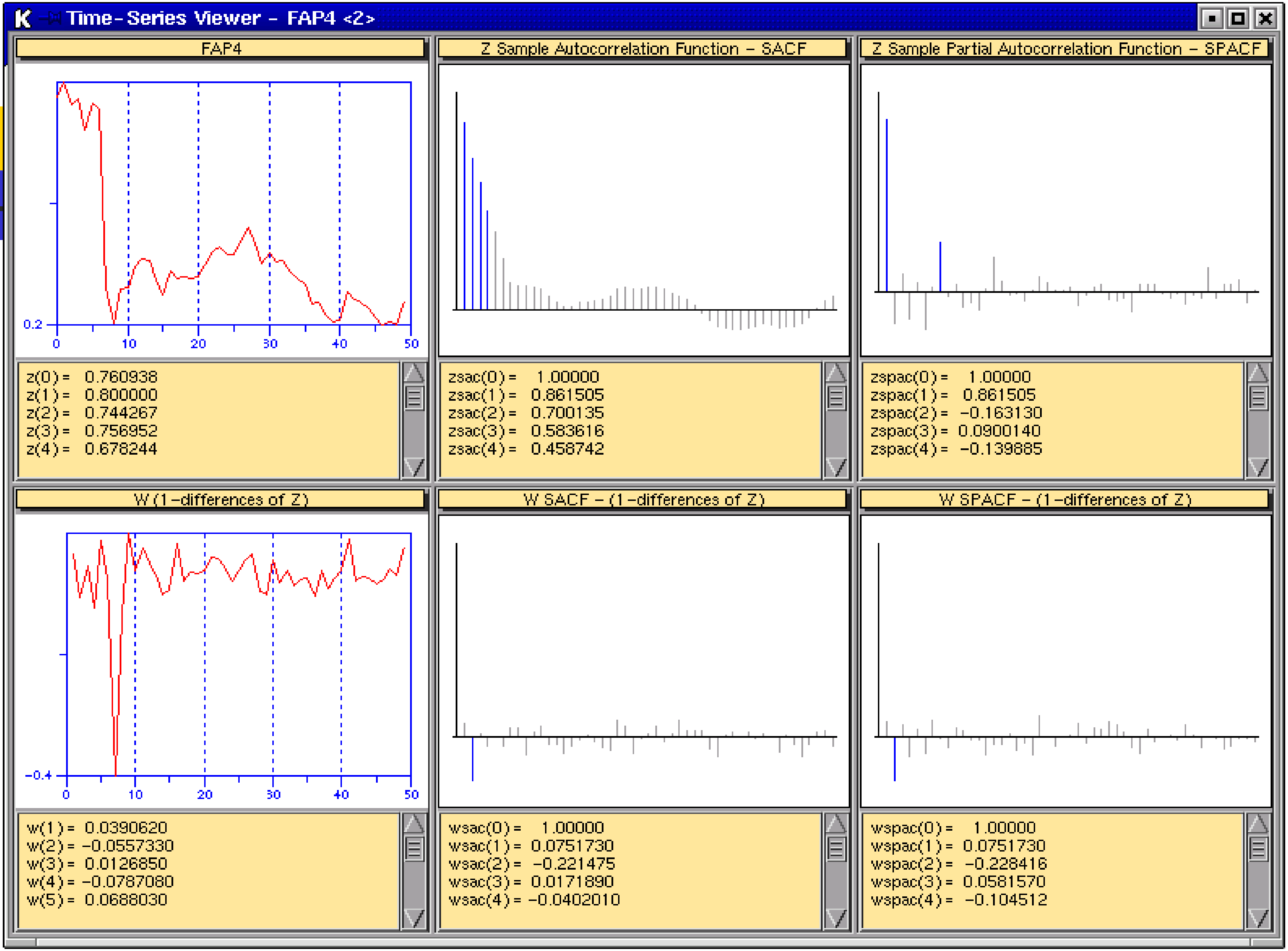

triangular.ts(Triangular) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Results
Data and model
Time series data:
Data description: Triangular
N=100 data points
mean=3.460000, variance=4.291313, standard deviatior
Model data:
Fitted model: ARIMA(2, 0, 0)
ar1=1.616980 ar2=-0.818226
delta=0.696311
Forecast performance
Data points:
Available data: 100
Forecasts: 50
Forecast Errors:
Data points used: 39 (lags 60 to 99)
RMS Error: 0.572534
MAPE Error: 0.189026
Foreacast
7.0
0.0
0
20
40
60
80
100
120
140
time lag
Origin lag
Lead time
Origin position
60
50
Updated Forecasts
Z
W
Data
Forecasts
Errors
Errors SAC
Same window
Plot

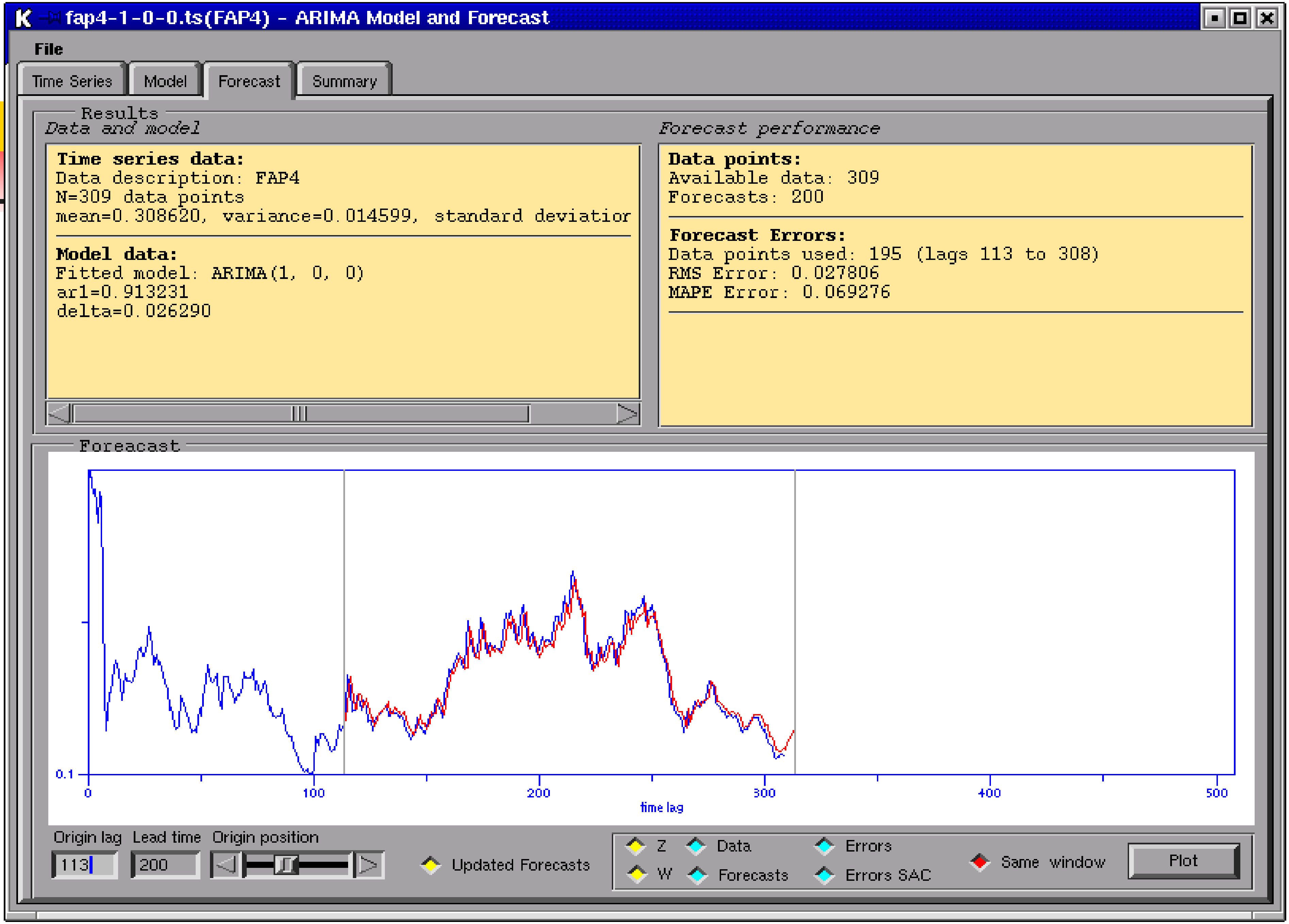

Time-Series Viewer - ARIMA(2,0,0) 50 UPDATED foreacsts at origin 60 for Triangular
Z and UPDATED forecasts: full data
Z and UPDATED forecasts: forecasted lags
Z and UPDATED forecasts: forecast errors
Z and UPDATED forecasts: SAC of forecast errors
z(0) = 1.00000
z(1) = 2.00000
z(2) = 3.00000
z(3) = 4.00000
Fz(61) = 5.50831
Fz(62) = 6.30706
Fz(63) = 7.10582
Fz(64) = 4.67061
z(61) = 6.00000
z(62) = 7.00000
z(63) = 6.00000
z(64) = 5.00000
Fz(61) = 5.50831
Fz(62) = 6.30706
Fz(63) = 7.10582
Fz(64) = 4.67061
Ez(61) = −0.491693
Ez(62) = −0.692939
Ez(63) = 1.10582
Ez(64) = −0.329391
Ez(65) = −0.128145
Rz(0) = 1.00000
Rz(1) = −0.297079
Rz(2) = −0.135564
Rz(3) = −0.0192580
Rz(4) = 0.0310970
W and UPDATED forecasts: full data
W and UPDATED forecasts: forecasted lags
W and UPDATED forecasts: forecast errors
W and UPDATED forecasts: SAC of forecast errors
w(0) = 1.00000
w(1) = 2.00000
w(2) = 3.00000
w(3) = 4.00000
Fw(61) = 5.50831
Fw(62) = 6.30706
Fw(63) = 7.10582
Fw(64) = 4.67061
w(61) = 6.00000
w(62) = 7.00000
w(63) = 6.00000
w(64) = 5.00000
Fw(61) = 5.50831
Fw(62) = 6.30706
Fw(63) = 7.10582
Fw(64) = 4.67061
Ew(61) = −0.491693
Ew(62) = −0.692939
Ew(63) = 1.10582
Ew(64) = −0.329391
Ew(65) = −0.128145
Rw(0) = 1.00000
Rw(1) = −0.297079
Rw(2) = −0.135564
Rw(3) = −0.0192580
Rw(4) = 0.0310970

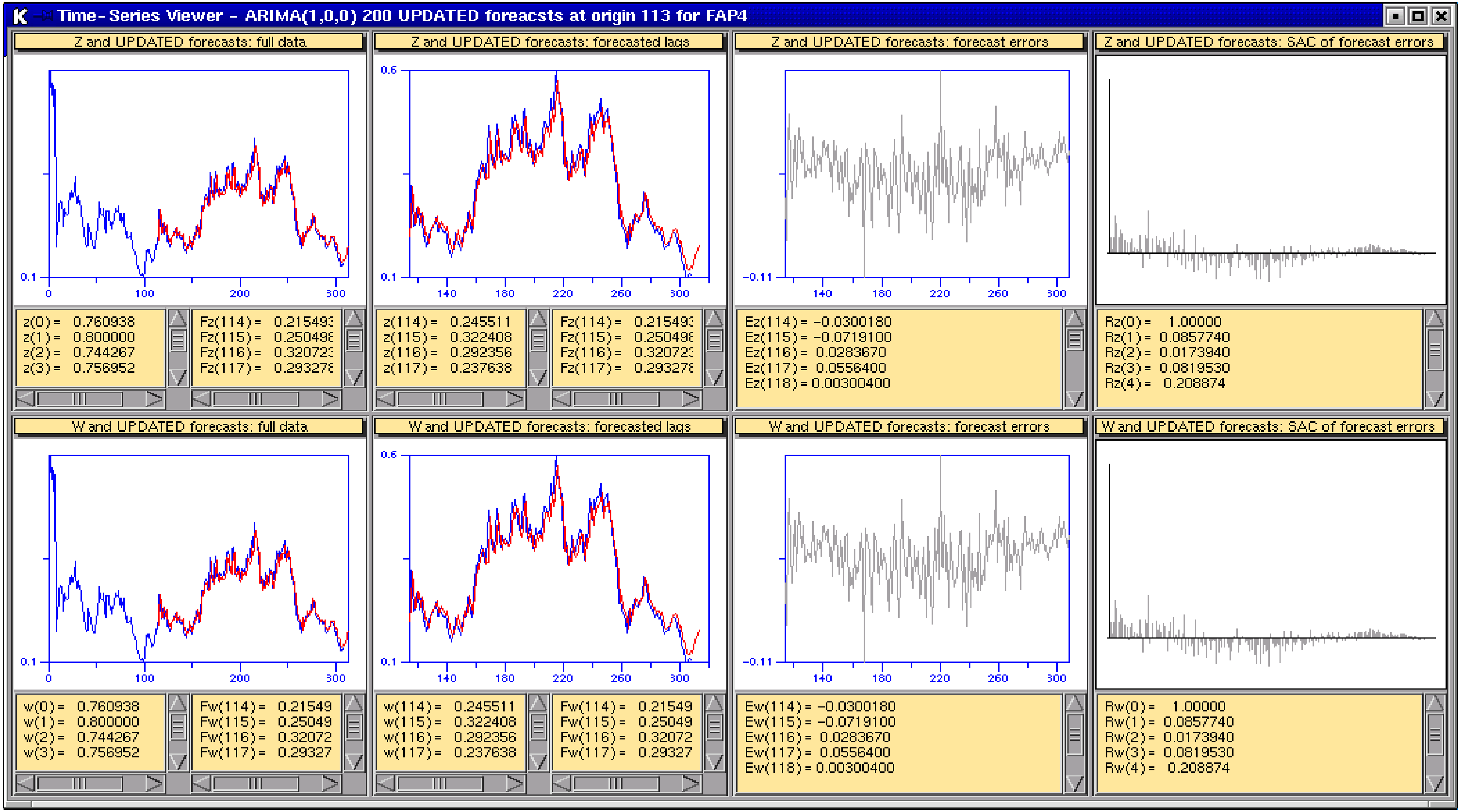

triangular.ts(Triangular) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Results
Data and model
Time series data:
Data description: Triangular
N=100 data points
mean=3.460000, variance=4.291313, standard deviatior
Model data:
Fitted model: ARIMA(3, 0, 0)
ar1=1.351413 ar2=-0.304898 ar3=-0.329418
delta=0.978845
Forecast performance
Data points:
Available data: 100
Forecasts: 50
Forecast Errors:
Data points used: 33 (lags 66 to 99)
RMS Error: 0.532697
MAPE Error: 0.179020
Foreacast
7.0
0.0
0
20
40
60
80
100
120
140
time lag
Origin lag
Lead time
Origin position
66
50
Updated Forecasts
Z
W
Data
Forecasts
Errors
Errors SAC
Same window
Plot


Time-Series Viewer - ARIMA(3,0,0) 50 UPDATED foreacsts at origin 66 for Triangular
Z and UPDATED forecasts: full data
Z and UPDATED forecasts: forecasted lags
Z and UPDATED forecasts: forecast errors
Z and UPDATED forecasts: SAC of forecast errors
z(0) = 1.00000
z(1) = 2.00000
z(2) = 3.00000
z(3) = 4.00000
Fz(67) = 2.16640
Fz(68) = 1.44930
Fz(69) = 0.732207
Fz(70) = 0.0151100
z(67) = 2.00000
z(68) = 1.00000
z(69) = 0.00000
z(70) = 1.00000
Fz(67) = 2.16640
Fz(68) = 1.44930
Fz(69) = 0.732207
Fz(70) = 0.0151100
Ez(67) = 0.166400
Ez(68) = 0.449304
Ez(69) = 0.732207
Ez(70) = -0.984890
Ez(71) = 0.000840000
Rz(0) = 1.00000
Rz(1) = -0.155123
Rz(2) = -0.375998
Rz(3) = -0.119638
Rz(4) = 0.103120


fap4-1-0-0.ts(FAP4) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Data selection
Original series (z) statistics
Z series -
Basic statistics:
N=100 data points
mean=0.302983, var=0.020208, stddev=0.142155
y0=0.760938 y1=0.800000 y2=0.744267 y3=0.756952 y
SACF-99 coefficients:
r1=0.861505 r2=0.700135 r3=0.583616 r4=0.458742 r
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.861505| > se=0.196987
q=2 - |r2=0.700135| > se=0.277212
q=3 - |r3=0.583616| > se=0.321373
q=4 - |r4=0.458742| > se=0.345851
SPACF-99 coefficients:
0.1
0
100
200
300
time lag
From lag
to lag
window
window position
0
99
100
Select
Reset
Reduction to stationarity
Reducted series (w) statistics
W series -
Basic statistics:
N=100 data points
mean=0.302983, var=0.020208, stddev=0.142155
y0=0.760938 y1=0.800000 y2=0.744267 y3=0.756952 y
SACF-99 coefficients:
r1=0.861505 r2=0.700135 r3=0.583616 r4=0.458742 r
Model Order Hypothesis:
q-order (criteria:se = 1.96 * Bartlett's std erro
q=1 - |r1=0.861505| > se=0.196987
q=2 - |r2=0.700135| > se=0.277212
q=3 - |r3=0.583616| > se=0.321373
q=4 - |r4=0.458742| > se=0.345851
SPACF-99 coefficients:
Method
Differencing order-d
0
Detrending
Mean for W_tilde
0.000000
Apply
Plot
Z
W
Data
SAC
SPAC
Same window
Lags to plot
50
Plot

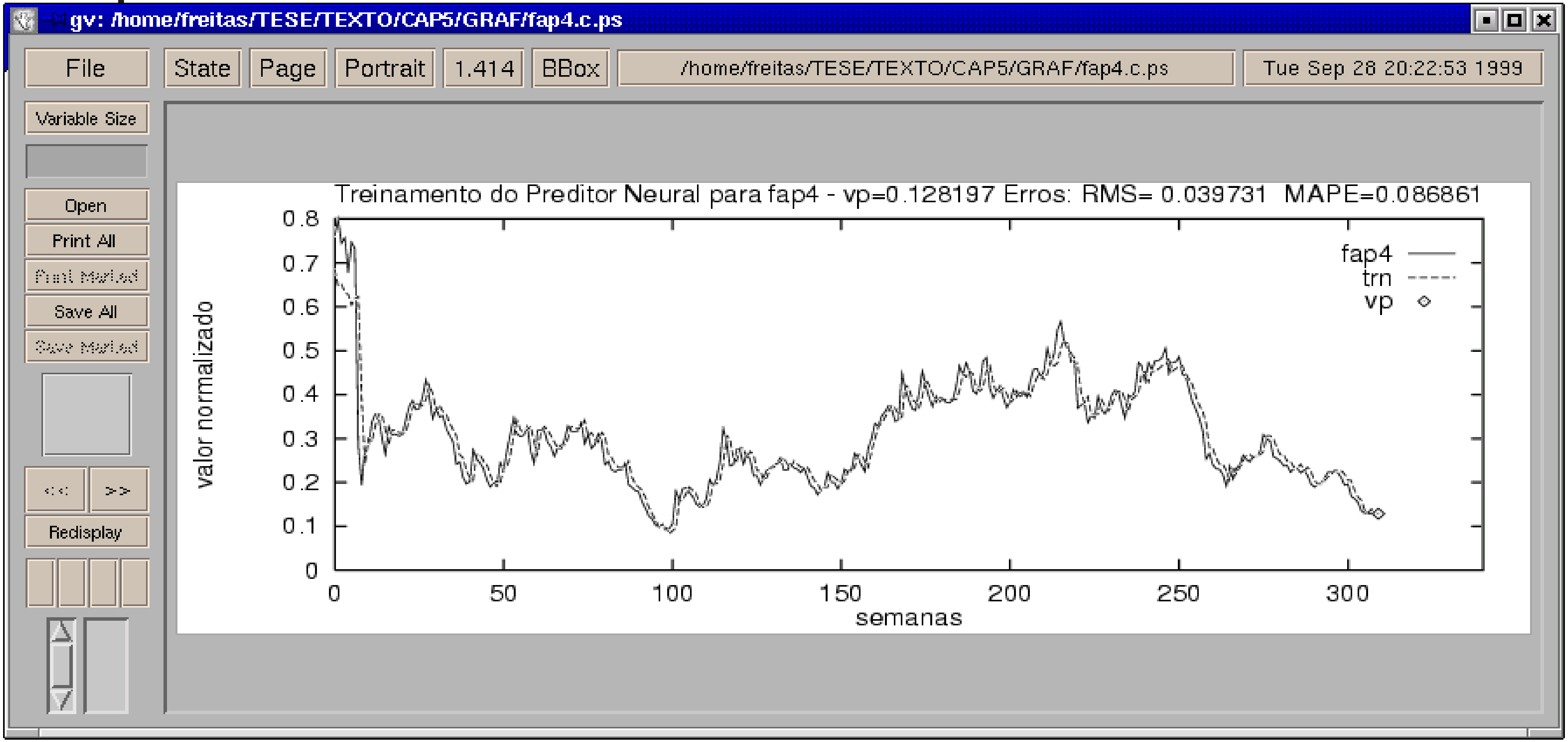

Time-Series Viewer - FAP4 <2>
FAP4
Z Sample Autocorrelation Function – SACF
Z Sample Partial Autocorrelation Function – SPACF
0.2
0
10
20
30
40
50
z(0) = 0.760938
z(1) = 0.800000
z(2) = 0.744267
z(3) = 0.756952
z(4) = 0.678244
zsac(0) = 1.00000
zsac(1) = 0.861505
zsac(2) = 0.700135
zsac(3) = 0.583616
zsac(4) = 0.458742
zspac(0) = 1.00000
zspac(1) = 0.861505
zspac(2) = −0.163130
zspac(3) = 0.0900140
zspac(4) = −0.139885
W (1−differences of Z)
W SACF – (1−differences of Z)
W SPACF – (1−differences of Z)
−0.4
0
10
20
30
40
50
w(1) = 0.0390620
w(2) = −0.0557330
w(3) = 0.0126850
w(4) = −0.0787080
w(5) = 0.0688030
wsac(0) = 1.00000
wsac(1) = 0.0751730
wsac(2) = −0.221475
wsac(3) = 0.0171890
wsac(4) = −0.0402010
wspac(0) = 1.00000
wspac(1) = 0.0751730
wspac(2) = −0.228416
wspac(3) = 0.0581570
wspac(4) = −0.104512


fap4-1-0-0.ts(FAP4) - ARIMA Model and Forecast
File
Time Series
Model
Forecast
Summary
Results
Data and model
Time series data:
Data description: FAP4
N=309 data points
mean=0.308620, variance=0.014599, standard deviatior
Model data:
Fitted model: ARIMA(1, 0, 0)
ar1=0.913231
delta=0.026290
Forecast performance
Data points:
Available data: 309
Forecasts: 200
Forecast Errors:
Data points used: 195 (lags 113 to 308)
RMS Error: 0.027806
MAPE Error: 0.069276
Foreacast
0.1
0
100
200
300
400
500
time lag
Origin lag
Lead time
Origin position
113
200
Updated Forecasts
Z
W
Data
Forecasts
Errors
Errors SAC
Same window
Plot


Time-Series Viewer - ARIMA(1,0,0) 200 UPDATED foreacsts at origin 113 for FAP4
Z and UPDATED forecasts: full data
Z and UPDATED forecasts: forecasted lags
Z and UPDATED forecasts: forecast errors
Z and UPDATED forecasts: SAC of forecast errors
z(0) = 0.760938
z(1) = 0.800000
z(2) = 0.744267
z(3) = 0.756952
Fz(114) = 0.215493
Fz(115) = 0.250498
Fz(116) = 0.320723
Fz(117) = 0.293278
z(114) = 0.245511
z(115) = 0.322408
z(116) = 0.292356
z(117) = 0.237638
Ez(114) = -0.0300180
Ez(115) = -0.0719100
Ez(116) = 0.0283670
Ez(117) = 0.0556400
Ez(118) = 0.00300400
Rz(0) = 1.00000
Rz(1) = 0.0857740
Rz(2) = 0.0173940
Rz(3) = 0.0819530
Rz(4) = 0.208874
W and UPDATED forecasts: full data
W and UPDATED forecasts: forecasted lags
W and UPDATED forecasts: forecast errors
W and UPDATED forecasts: SAC of forecast errors
w(0) = 0.760938
w(1) = 0.800000
w(2) = 0.744267
w(3) = 0.756952
Fw(114) = 0.21549
Fw(115) = 0.25049
Fw(116) = 0.32072
Fw(117) = 0.29327
w(114) = 0.245511
w(115) = 0.322408
w(116) = 0.292356
w(117) = 0.237638
Ew(114) = -0.0300180
Ew(115) = -0.0719100
Ew(116) = 0.0283670
Ew(117) = 0.0556400
Ew(118) = 0.00300400
Rw(0) = 1.00000
Rw(1) = 0.0857740
Rw(2) = 0.0173940
Rw(3) = 0.0819530
Rw(4) = 0.208874

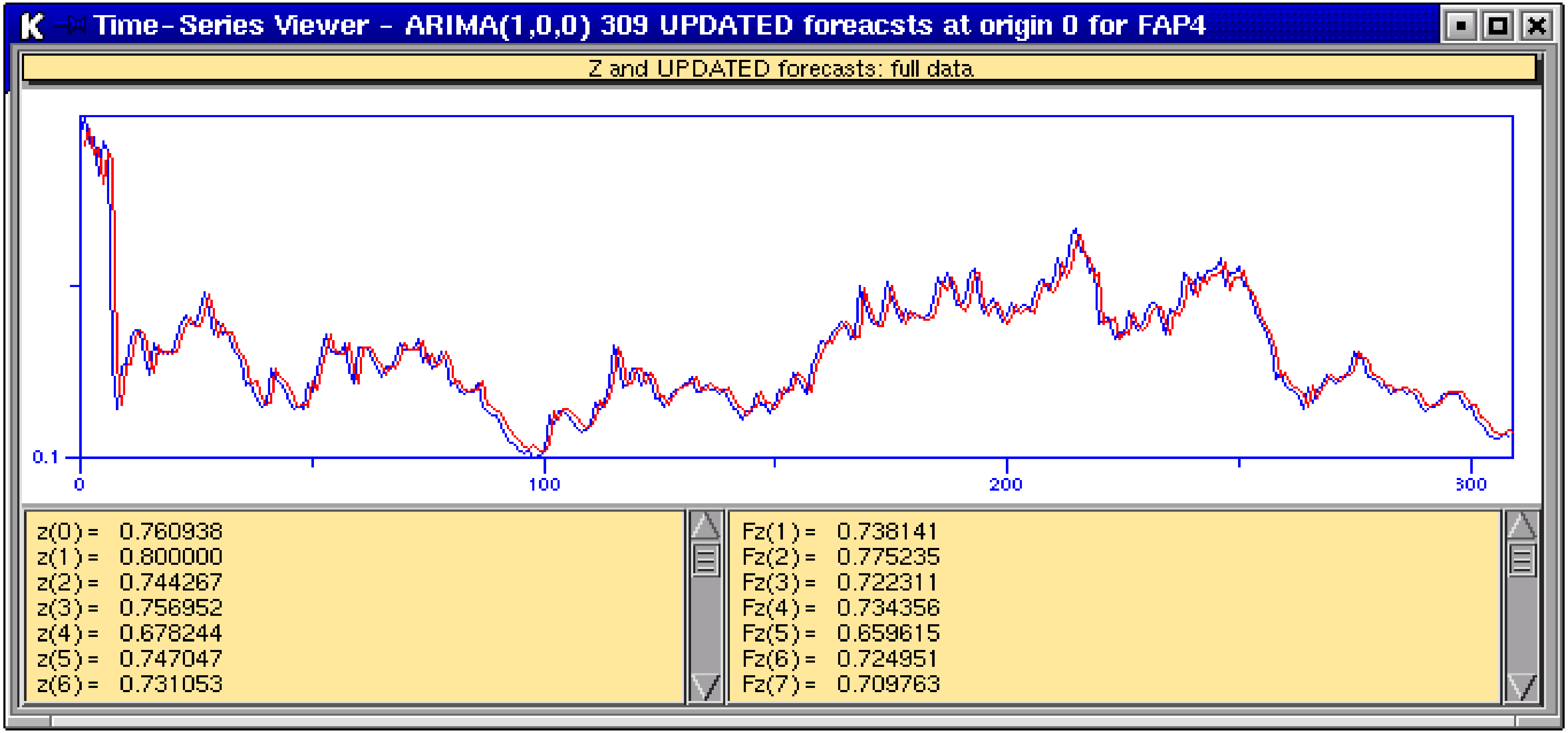
Preditor Neural para COFAP (FAP4)

- RMSE=0.039731 MAPE=0.086861



ARIMA(1,0,0) para COFAP (FAP4)

- RMSE=0.037596 MAPE=0.080951



Conclusão

- Metodologia ARIMA
  - Vantagens:
    - Corpo de conhecimento bastante sólido.
    - Permite calcular os erros e intervalos de confiança para a estimativa dos parâmetros.
    - Relativa imunidade à amplitude dos dados.
    - Necessita de pequena quantidade de dados (*backforecast*).
  - Desvantagens
    - Modelagem fundamentada na intervenção humana – experiência do modelador é determinante.
    - Processos lineares.

Discussão

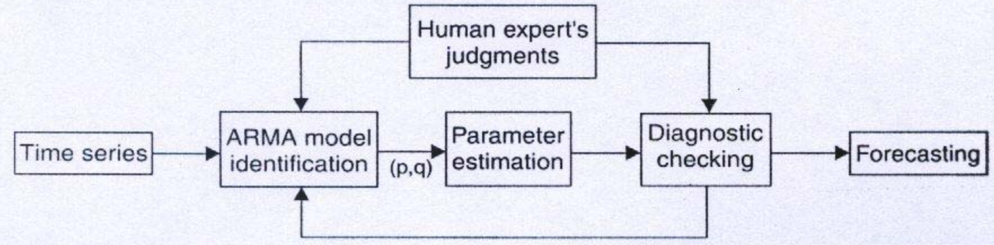
- Identificação
  - Automatização da Identificação dos modelos com redes neurais.
  - Utilização da SACF e SPACF para a seleção de arquiteturas de preditores neurais.

Discussão

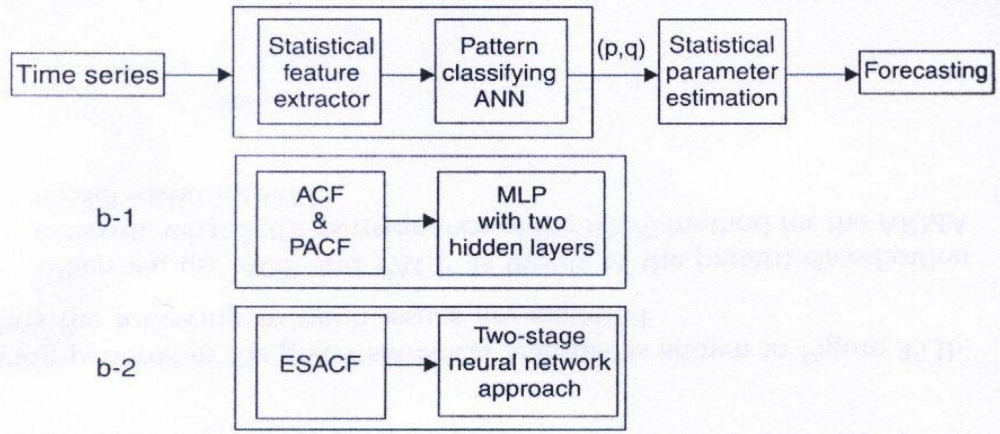
Figure 35.1

ARMA Modeling Procedure and Suggested ANN Approaches

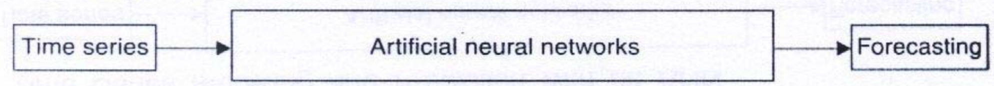
a. Traditional ARMA Model Building Procedure



b. Automating the Identification Step by Adopting the ANN



c. Time Series Modeling and Prediction with an ANN



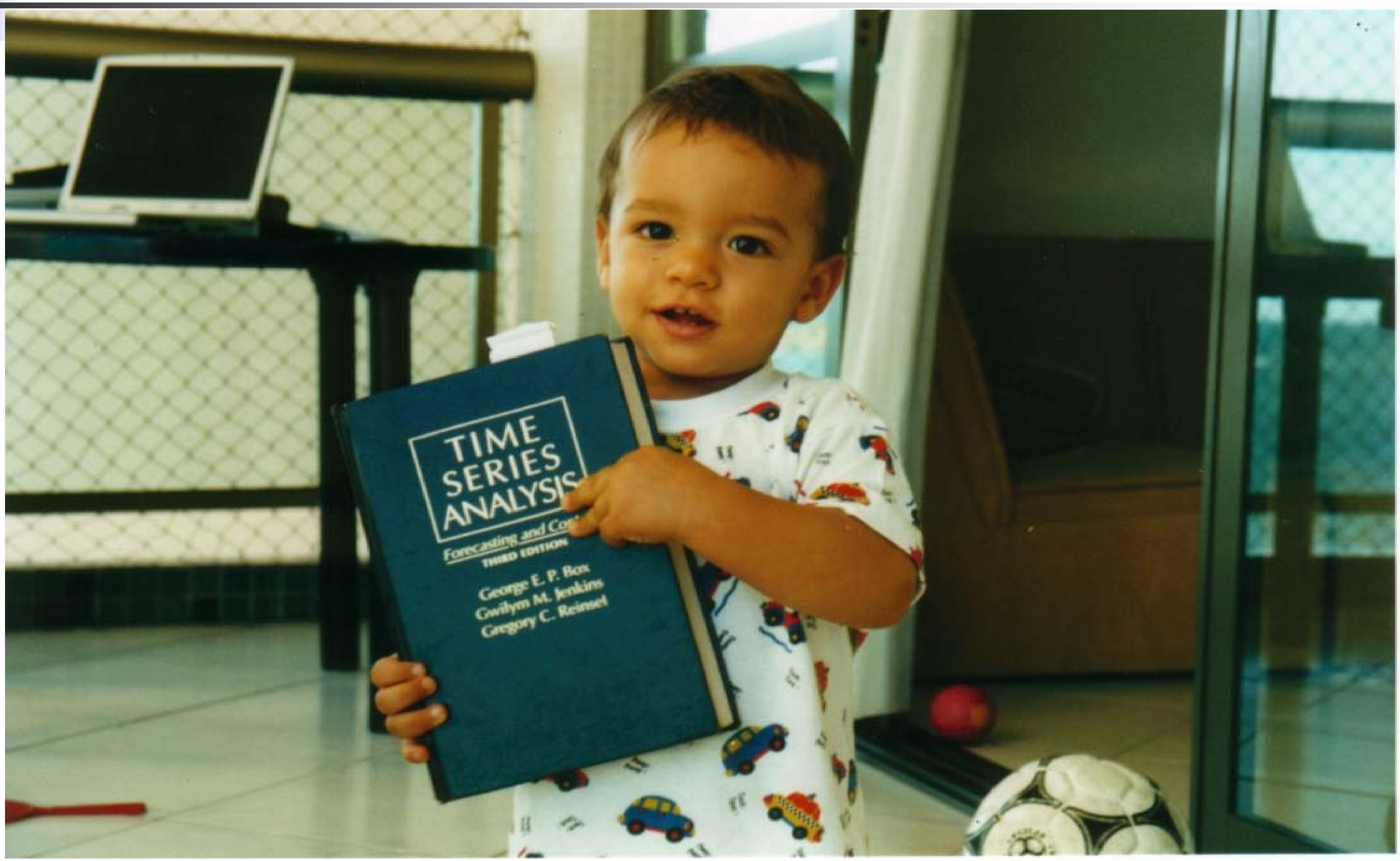
Discussão

- Estimação
  - Estimação dos parâmetros com método de otimização global → Algoritimo genético (GA).

Referências Bibliográficas

- "Time Series Analysis – Forecasting and Control", G.P.Box, G.M.Jenkins, G.C.Reisnel, 3rd edition 1994
- "Applied Time Series Analysis", Charles R. Nelson, 1973
- "Performance of Neural Networks in Managerial Forecasting", W.C.Jhee and J.K. Lee, *Intelligent Systems in Accounting, Finance and Management, April 1993*

Fábio Daros de Freitas
www: www.inf.ufes.br/~freitas
E-mail: freitas@computer.org
2003


TIME
SERIES
ANALYSIS
THIRD EDITION
George E. P. Box
Gwilym M. Jenkins
Gregory C. Reinsel