



Sumar

✓ Bibliografie

✓ 0 Organizarea temelor de laborator

✓ 8 Modelarea și predicția seriilor de timp

✓ 8.1 Estimarea modelului polinomial al tendinței

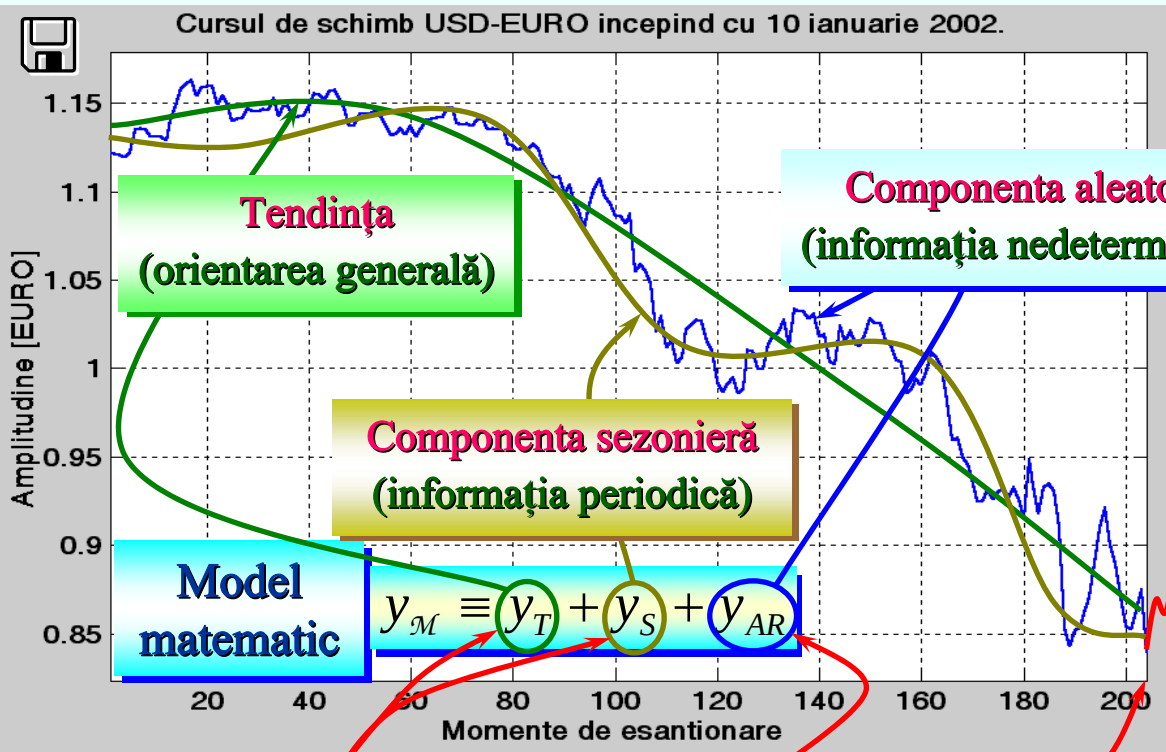
✓ 8.2 Estimarea componentei sezoniere

✓ 8.3 Estimarea componentei nedeterministe (aleatoare)

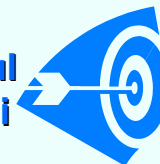
☞ 8.4 Predicția seriei de timp

Modelarea și predicția seriilor de timp

§.4 Predicția seriei de timp



Obiectivul
modelării



Predicția/Proгноza
seriei de timp.

Nu doar **valorile predictate** trebuie estimate, ci și **dispersiile erorilor de predicție aferente**.

$$\{\hat{\sigma}_k^2\}_{k \in \overline{1, K}}$$

Acestea se pot determina numai folosind **componenta nedeterministă** a modelului.

Deterministe

Nedeterministă

$T_{\max}$

$$t_{N+1} < t_{N+2} < \dots < t_{N+K}$$

momente de predicție

$$\hat{y}(t_{N+k} | N) = y_T(t_{N+k}) + y_S(t_{N+k}) + \hat{y}_{AR}(t_{N+k} | N)$$

$$\forall k \in \overline{1, K}$$

valoarea **curent predictată** a seriei de timp, pe baza celor N date măsurate

valori **extrapolate, deterministe**

valoarea **optimală de predicție** curent returnată de către componenta aleatoare, pe baza celor N date măsurate



Modelarea și predicția seriilor de timp

§.4 Predicția seriei de timp

Valori extrapolate?



Se determină **direct**
în cazul **tendinței**.

$$y_T(t_{N+k}) = a_0 + a_1 t_{N+k} + \dots + a_p t_{N+k}^p \quad \forall k \in \overline{1, K}$$

Necesită o **abordare în doi pași**
pentru **componenta sezonieră**.

① Valorile componentei sezoniere trebuie **regrupate în perioada principală**,
după pozițiile ocupate de acestea în perioadele curente.

coeficienți sezonieri **grupați în perioada principală**

⚡ Acest pas nu este necesar în cazul
eșantionării uniforme.

$$[0, PT_s]$$

$$y_S^P \left(\left\{ \frac{t_n}{PT_s} \right\} PT_s \right)^{def} = y_S(t_n) \quad \forall n \in \overline{1, N}$$

partea fracționară

② Interpolarea și re-eșantionarea coeficienților sezonieri
grupați în perioada principală.

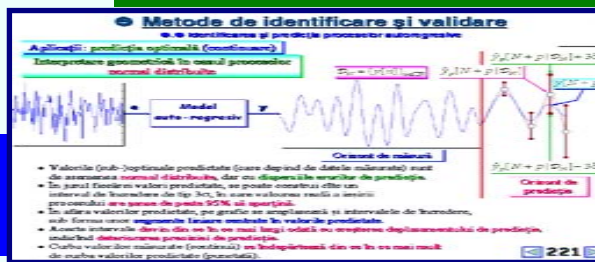
⚡ Acest pas se simplifică în cazul
eșantionării uniforme.

$$y_S(t_{N+k}) = y_S^P \left(\left\{ \frac{t_{N+k}}{PT_s} \right\} PT_s \right) \quad \forall k \in \overline{1, K}$$

Predictor optimal?



Determinat în cadrul
cursului de **IS**.



Modelarea și predicția seriilor de timp

§.4 Predicția seriei de timp

Predictorul optimal al componentei aleatoare

$$\hat{y}_{AR}(t_{N+k} | N) = v(t_{N+k}) \quad \forall k \leq 0$$

$$\hat{y}_{AR}(t_{N+k} | N) = -\hat{a}_{na,1} \hat{y}_{AR}(t_{N+k-1} | N) - \dots - \hat{a}_{na,k-1} \hat{y}_{AR}(t_{N+1} | N) - \hat{a}_{na,k} v(t_N) - \dots - \hat{a}_{na,na} v(t_{N+k-na}) \quad \forall k \in \overline{1, na}$$

$$\hat{y}_{AR}(t_{N+k} | N) = -\hat{a}_{na,1} \hat{y}_{AR}(t_{N+k-1} | N) - \dots - \hat{a}_{na,na} \hat{y}_{AR}(t_{N-na} | N) \quad \forall k \in \overline{na+1, K}$$

Practic, se folosește ecuația modelului AR, forțând zgomotul alb la valori nule.

Dispersiile erorilor de predicție aferente se evaluează **recursiv**.

$$\begin{cases} \hat{\sigma}_k^2 = \hat{\sigma}_{k-1}^2 + \hat{\lambda}_{N,na}^2 \hat{\alpha}_{k-1}^2 \\ \hat{\sigma}_0^2 = 0 \end{cases} \quad \forall k \in \overline{1, K}$$

pătratele coeficienților obținuți prin împărțirea infinită a polinoamelor 1 și $A(q^{-1})$

Dispersia erorii de predicție crește odată cu îndepărtarea de orizontul de măsură.

$$K \leq 5$$

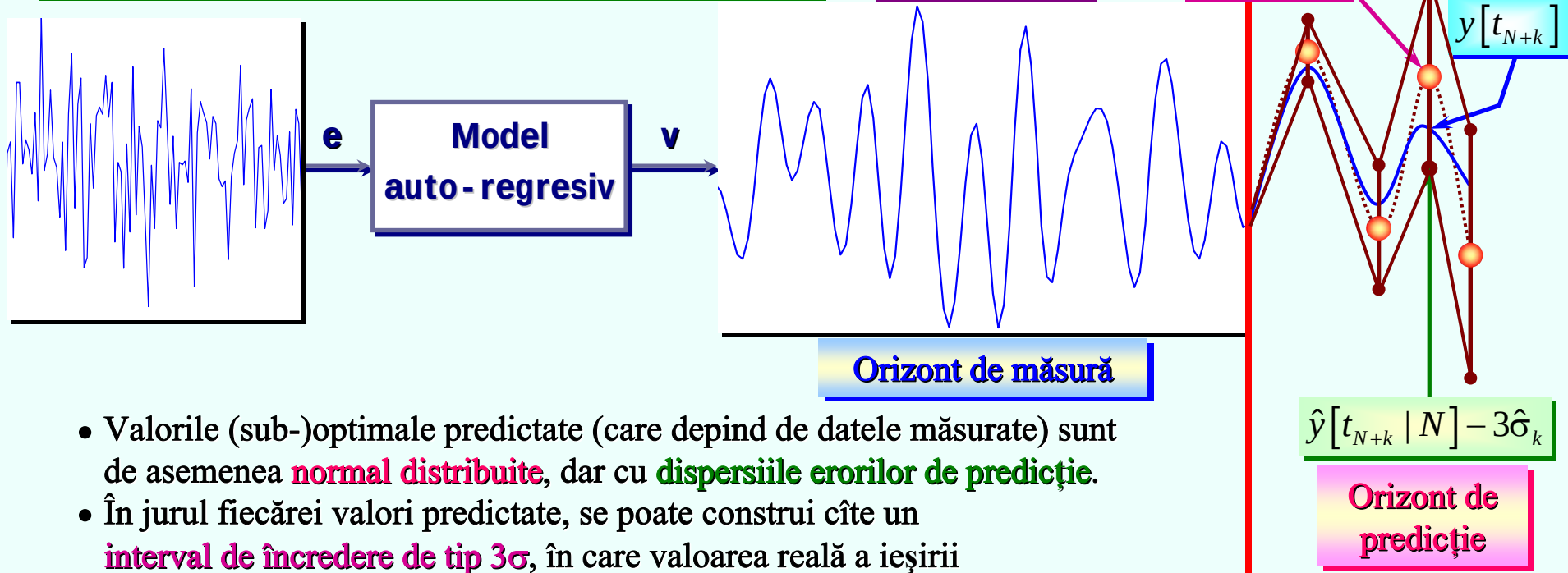
Împărțirea infinită a polinoamelor

$$\begin{array}{l} \cancel{1} \\ \hline \cancel{-1} - \hat{a}_1 q^{-1} - \hat{a}_2 q^{-2} - \dots \\ \hline -\hat{a}_1 q^{-1} - \hat{a}_2 q^{-2} - \dots \\ \hline \hat{a}_1 q^{-1} + (\hat{a}_1)^2 q^{-2} + \dots \\ \hline \vdots \\ \hline (\hat{a}_1)^2 - \hat{a}_2 q^{-2} + \dots \end{array} \quad \begin{array}{l} 1 + \hat{a}_1 q^{-1} + \hat{a}_2 q^{-2} + \dots + \hat{a}_{na} q^{-na} \\ \hline \hat{1} = \hat{a}_1 q^{-1} + \dots + \hat{\alpha}_{k-1} q^{1-k} + \dots \\ \hline \hat{\alpha}_0 \quad \hat{\alpha}_1 \\ \hline \hat{\alpha}_2 \end{array}$$

Modelarea și predicția seriilor de timp

§.4 Predicția seriei de timp

Reprezentarea geometrică în cazul proceselor
normal distribuite



- Valorile (sub-)optimale predictate (care depind de datele măsurate) sunt de asemenea **normal distribuite**, dar cu **dispersiile erorilor de predicție**.
- În jurul fiecărei valori predictate, se poate construi câte un **interval de încredere de tip 3σ** , în care valoarea reală a ieșirii procesului **are șanse de peste 95% să aparțină**.
- **În afara valorilor predictate, pe grafic se amplasează și intervalele de încredere, sub forma unor segmente liniare centrate în valorile predictate.**
- Aceste intervale **devin din ce în ce mai largi odată cu creșterea deplasamentului de predicție**, indicând **deteriorarea preciziei de predicție**.
- Curba valorilor măsurate (continuă) **se îndepărtează statistic din ce în ce mai mult** de curba valorilor predictate (punctată).

Modelarea și predicția seriilor de timp

§.4 Predicția seriei de timp

Cum se pot determina cei doi indici structurali?

Se utilizează criteriul de selecție:

Calitatea predicției

☞ Care trebuie maximizată!

$$PQ_K[p, P, na] \stackrel{\text{def}}{=} \frac{100}{1 + \frac{\sum_{k=1}^K \hat{\sigma}_k |y(t_{N+k}) - \hat{y}[t_{N+k} | N]|}{\sigma_y \sum_{k=1}^K \hat{\sigma}_k}} \quad [\%]$$

Dependența de indicii structurali este **implicită** și nu **explicită**.

Perioada componentei sezoniere este **unic determinată** de gradul polinomului tendință.

Criteriul necesită **cunoașterea datelor pe orizontul de predicție**.

☞ Media ponderată și normalizată a distanțelor dintre datele măsurate și cele predictate.

Atunci... de ce mai este necesară predicția?

$$\sigma_y \stackrel{\text{def}}{=} \sqrt{\frac{1}{N} \sum_{n=1}^N (y(t_n) - \langle y \rangle)^2}$$

deviația standard a datelor originale

$$\langle y \rangle \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N y(t_n)$$

media datelor originale

Se apelează la următoarea strategie:

- ① Se fixează lungimea orizontului de predicție (K) și se determină lungimea seriei de timp (N).
- ② Se utilizează **numai** ($N-K$) date pentru construcția celor 3 modele.
- ③ Se utilizează **ultimele** K date pentru selecția structurii modelelor.



Modelarea și predicția seriilor de timp

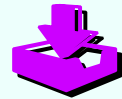
§.4 Predicția seriei de timp

Care este tema de laborator?



Proiectarea și implementarea unei rutine de predicție a seriilor de timp, bazată pe modelul cu 3 componente, predictorul optimal și criteriul calității predicției.

Pași principali ai algoritmului de predicție



Date de intrare

- $\mathcal{D}$ (seria de timp)
- K (lungimea orizontului de predicție; implicit: $K = 5$)
- $P_{\max}$ (gradul maxim al polinomului tendință; implicit: $P_{\max} = 10$)
- Na (ordinul maxim al modelului AR; implicit: $Na = 100$)



Inițializare

Determinarea lungimii seriei de timp. $\Rightarrow N$

Separarea datelor în două.

$$\mathcal{D}_{\text{model}} = \{y(t_n)\}_{n \in \overline{1, N-K}}$$

date de modelare

$$\mathcal{D}_{\text{struct}} = \{y(t_n)\}_{n \in \overline{N-K+1, N}}$$

date de alegere a structurii

Evaluarea perioadei de eșantionare

$$T_s$$

Unitară, pentru seriile eșantionate uniform.

Evaluarea mediei și a deviației standard pentru datele de modelare.

$$\langle y \rangle \quad \sigma_y$$



84.7



⑧ Modelarea și predicția seriilor de timp

⑧.④ Predicția seriei de timp

👉 Pași principali ai algoritmului de predicție (continuare)



Bucle iterative

❶ Pentru $p \in \overline{0, P_{\max}}$

① Se apelează rutina **trend** cu datele de modelare și gradul p .

• se obțin: y_{sta} y_T $\hat{\theta}_{N,p}$

② Se apelează rutina **seasonal** cu datele staționarizate curente.

• se obțin: v y_S P $\mathcal{V}$

❷ Pentru $na \in \overline{0, Na}$

① Se apelează rutina **stochastic** cu zgomotul colorat curent și ordinul na .

• se obțin: $\hat{e}$ $\hat{\lambda}_{N,na}^2$ y_{AR} $\hat{\theta}_{N,na}^{AR}$

② Se extrapolează tendința și componenta sezonieră pe orizontul de predicție.

• se obțin: $y_{T,extra}$ $y_{S,extra}$

③ Se predictează zgomotul colorat.

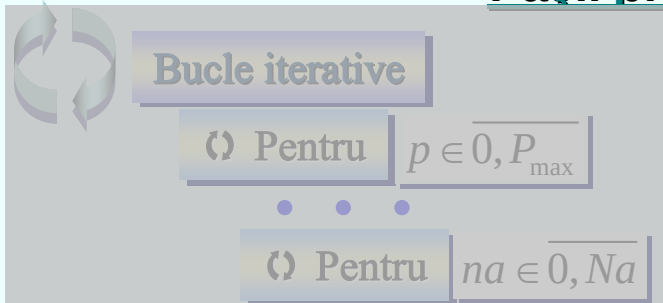
• se obțin: $y_{AR,pred}$ $\{\hat{\sigma}_k^2\}_{k \in \overline{1, K}}$



⑧ Modelarea și predicția seriilor de timp

⑧.④ Predicția seriei de timp

👉 Pași principali ai algoritmului de predicție (continuare)



④ Se predictează seria de timp. $y_{\text{pred}} \equiv y_{T,\text{extra}} + y_{S,\text{extra}} + y_{AR,\text{pred}}$

⑤ Se evaluează calitatea predicției, folosind datele predictate curente și cele de alegere a structurii. $PQ_K[p, P, na]$

Se determină punctul de maxim al calității predicției. $(p_0, P_0, na_0, PQ_K[p_0, P_0, na_0])$

🔥 Calitatea predicției este o funcție de două variabile: p și na .

Se apelează rutina **trend** cu datele de modelare și gradul p_0

Se apelează rutina **seasonal** cu datele staționarizate curente.

Se apelează rutina **stochastic** cu zgomotul colorat curent și ordinul na_0 .

Se predictează seria de timp folosind modelele precedente.

Date de ieșire

y_{pred} $\{\hat{\sigma}_k^2\}_{k \in \overline{1, K}}$ y_T $\hat{\theta}_{N, p_0}$ y_{sta} y_S P_0 v y_{AR} $\hat{\theta}_{N, na_0}^{AR}$ $\hat{e}$

⑧ Modelarea și predicția seriilor de timp

⑧.④ Predicția seriei de timp

Rutină ce trebuie proiectată

vectorul componentei sezoniere
pe orizontul de măsură

perioada componentei
sezoniere

vectorul datelor
staționarizate

vectorul
zgomotului colorat

vectorul parametrilor
modelului polinomial

vectorul
modelului AR

vectorul tendinței pe
orizontul de măsură

vectorul parametrilor
modelului AR

vectorul dispersiilor
erorilor de predicție

vectorul zgomotului
alb rezidual

matricea calității
predicției

vectorul datelor
predictate

```
>> [ypred, sigma2, yT, theta, ysta, yS, P0, v, yAR, theta_AR, e, PQ] =  
    = ts_predict(y, K, Pmax, Na);
```

vectorul de date care
conține seria de timp

ordinul maxim al
modelului AR

lungimea orizontului
de predicție

gradul maxim al
polinomului tendință

⑧ Modelarea și predicția seriilor de timp

⑧.④ Predicția seriei de timp

Programul de test al rutinei `ts_predict`

ISLAB_8D

👉 Pași principali

Se introduce numărul fișierului care conține seria de timp.

nts

Se introduce lungimea orizontului de predicție.

K

Se introduce gradul maxim al polinomului tendință.

Pmax

Se introduce ordinul maxim al modelului AR.

Na

Se încarcă seria de timp în spațiul de lucru.

• cu ajutorul funcției: eval

Se determină lungimea și dispersia seriei de timp.

N

lambda_y2

Se apelează rutina `ts_predict`.

Se trasează graficele seriei de timp restrânse la $(N-K)$ date și al tendinței (în aceeași fereastră, cu precizarea gradului polinomului).

Dacă

$P_0 \neq 0$

Se trasează graficele seriei de timp staționarizate și al componentei sezoniere (în aceeași fereastră, cu precizarea perioadei).

Altfel

Utilizatorul este informat printr-un mesaj corespunzător.



84.11



⑧ Modelarea și predicția seriilor de timp

⑧.④ Predicția seriei de timp

Programul de test al rutinei `ts_predict`

ISLAB_8D

lambda_y2

👉 Pași principali (continuare)

Se estimează raportul semnal-zgomot.

$$\text{SNR} = \frac{\hat{\lambda}_y^2}{\hat{\lambda}_e^2}$$

👉 Dispersia zgomotului alb estimat este egală cu primul element al vectorului dispersiilor erorilor de predicție.

Se trasează graficele zgomotului colorat și al componentei aleatoare (în aceeași fereastră, cu precizarea SNR).

Se trasează graficul zgomotului alb rezidual (cu precizarea dispersiei estimate pe grafic).

Se trasează graficul suprafeței calității de predicție (în funcție de p și na , cu precizarea punctului de maxim pe grafic).

• cu ajutorul funcției:

surf

Se trasează graficul detaliat al datelor măsurate și predictate pe orizontul de predicție, cu reprezentarea tubului de încredere și precizarea valorii maxime a calității de predicție.

👉 Graficele trebuie trasate cu axele corect scalate și marcate, cu titluri sugestive și legendă de discriminare (dacă este cazul).

