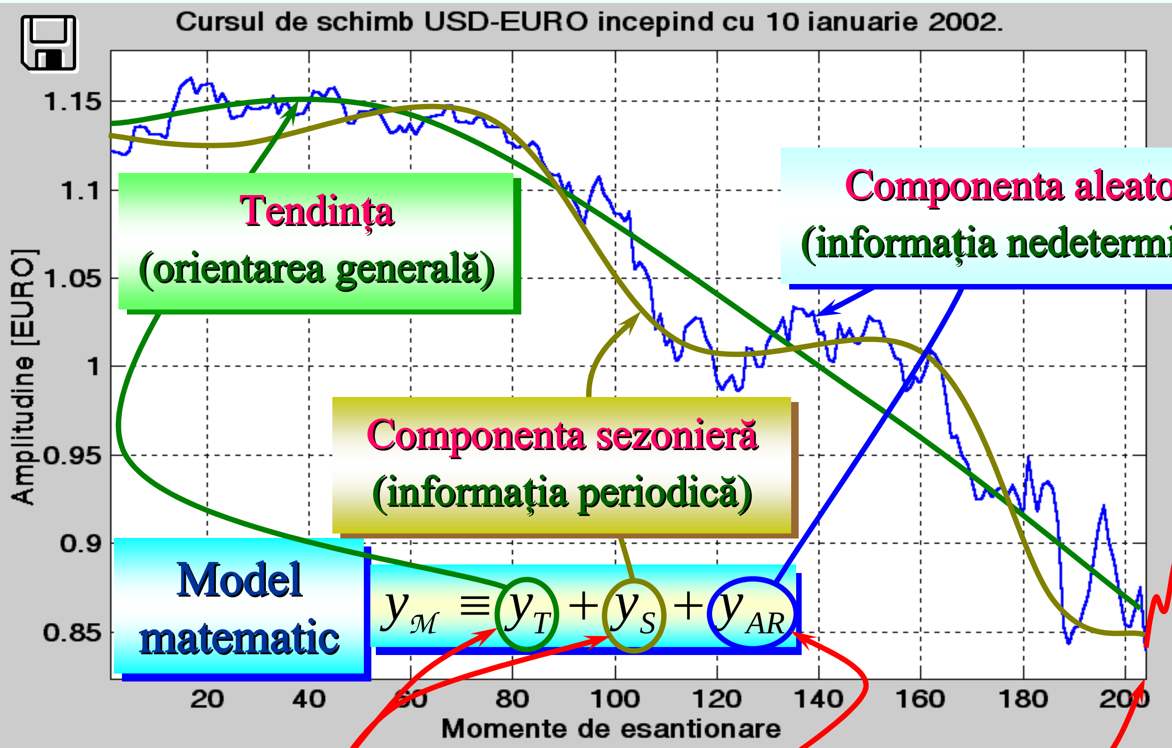
Sumar

- ✓ Bibliografie
- ✓ ⓿ Organizarea temelor de laborator
- ✓ ⑧ Modelarea şi predicţia seriilor de timp
  - ✓ ⑧.① Estimarea modelului polinomial al tendinţei
  - ✓ ⑧.② Estimarea componentei sezoniere
  - ✓ ⑧.③ Estimarea componentei nedeterministe (aleatoare)
  - ☞ ⑧.❹ Predicţia seriei de timp

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia seriei de timp



Tendinţa (orientarea generală)

Componenta sezonieră (informaţia periodică)

Componenta aleatoare (informaţia nedeterministă)

Model matematic

$$y_{\mathcal{M}} \equiv y_T + y_S + y_{AR}$$

Deterministe

Nedeterministă

$T_{\max}$

Obiectivul modelării: Predicţia/Prognoza seriei de timp.

Nu doar valorile predictate trebuie estimate, ci şi dispersiile erorilor de predicţie aferente.

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

Acestea se pot determina numai folosind componenta nedeterministă a modelului.

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

momente de predicţie

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

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

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

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

valori extrapolate, deterministe

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia seriei de timp

Se determină **direct** în cazul **tendinţei**.

$$y_T\left(t_{N+k}\right) = a_0 + a_1\, t_{N+k} + \cdots + a_p\, t_{N+k}^{p}$$

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

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

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

coeficienţi sezonieri **grupaţi în perioada principală** → y_S^P

$$y_S^P\left(\left\{\frac{t_n}{PT_s}\right\} PT_s\right) \overset{def}{=} y_S\left(t_n\right)$$

$\forall n \in \overline{1,N}$

$\left[0, PT_s\right]$

partea **fracţionară**

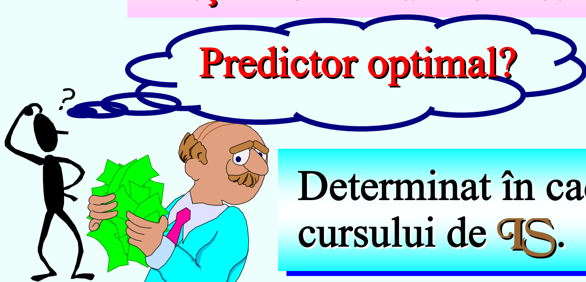
☝ Acest pas nu este necesar în cazul eşantionării uniforme.

② **Interpolarea şi re-eşantionarea coeficienţilor sezonieri grupaţi în perioada principală.**

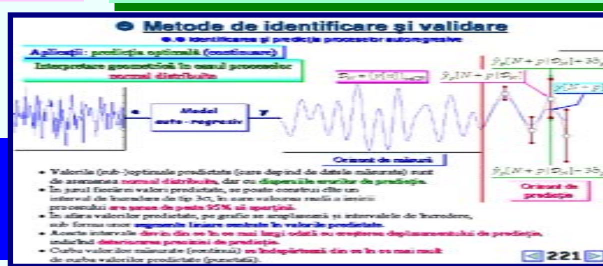
$$y_S\left(t_{N+k}\right) = y_S^P\left(\left\{\frac{t_{N+k}}{PT_s}\right\} PT_s\right)$$

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

☝ Acest pas se simplifică în cazul eşantionării uniforme.



Determinat în cadrul cursului de IS.



❽.❹ Predicţia seriei de timp

Predictorul optimal al componentei aleatoare

$$\begin{cases} \hat{y}_{AR}(t_{N+k} \mid N) = v(t_{N+k}) & \forall k \le 0 \\ \hat{y}_{AR}(t_{N+k} \mid N) = -\hat{a}_{na,1}\hat{y}_{AR}(t_{N+k-1} \mid N) - \cdots - \hat{a}_{na,k-1}\hat{y}_{AR}(t_{N+1} \mid N) - \hat{a}_{na,k}v(t_N) - \cdots - \hat{a}_{na,na}v(t_{N+k-na}) & \forall k \in \overline{1,na} \\ \hat{y}_{AR}(t_{N+k} \mid N) = -\hat{a}_{na,1}\hat{y}_{AR}(t_{N+k-1} \mid N) - \cdots - \hat{a}_{na,na}\hat{y}_{AR}(t_{N-na} \mid N) & \forall k \in \overline{na+1,K} \end{cases}$$

☝ Practic, se foloseşte ecuaţia modelului AR, forţînd zgomotul alb la valoari nule.

Dispersiile erorilor de predicţie aferente se evaluează recursiv.

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

pătratele coeficienţilor obţinuţi prin împărţirea infinită a polinoamelor **1** şi $A(q^{-1})$

Împărţirea infinită a polinoamelor

$$1 \;\Big|\; 1 + \hat{a}_1 q^{-1} + \hat{a}_2 q^{-2} + \cdots + \hat{a}_{na} q^{-na}$$

$$\cancel{-1} - \hat{a}_1 q^{-1} - \hat{a}_2 q^{-2} - \cdots \qquad 1 - \hat{a}_1 q^{-1} + \cdots + \hat{\alpha}_{k-1} q^{1-k} + \cdots$$

$$\cancel{-\hat{a}_1 q^{-1}} - \hat{a}_2 q^{-2} - \cdots$$

$$\cancel{\hat{a}_1 q^{-1}} + (\hat{a}_1)^2 q^{-2} + \cdots$$

$$\vdots \qquad \left[(\hat{a}_1)^2 - \hat{a}_2\right] q^{-2} + \cdots$$

$\hat{\alpha}_0$ $\quad \hat{\alpha}_1$ $\quad \hat{\alpha}_2$

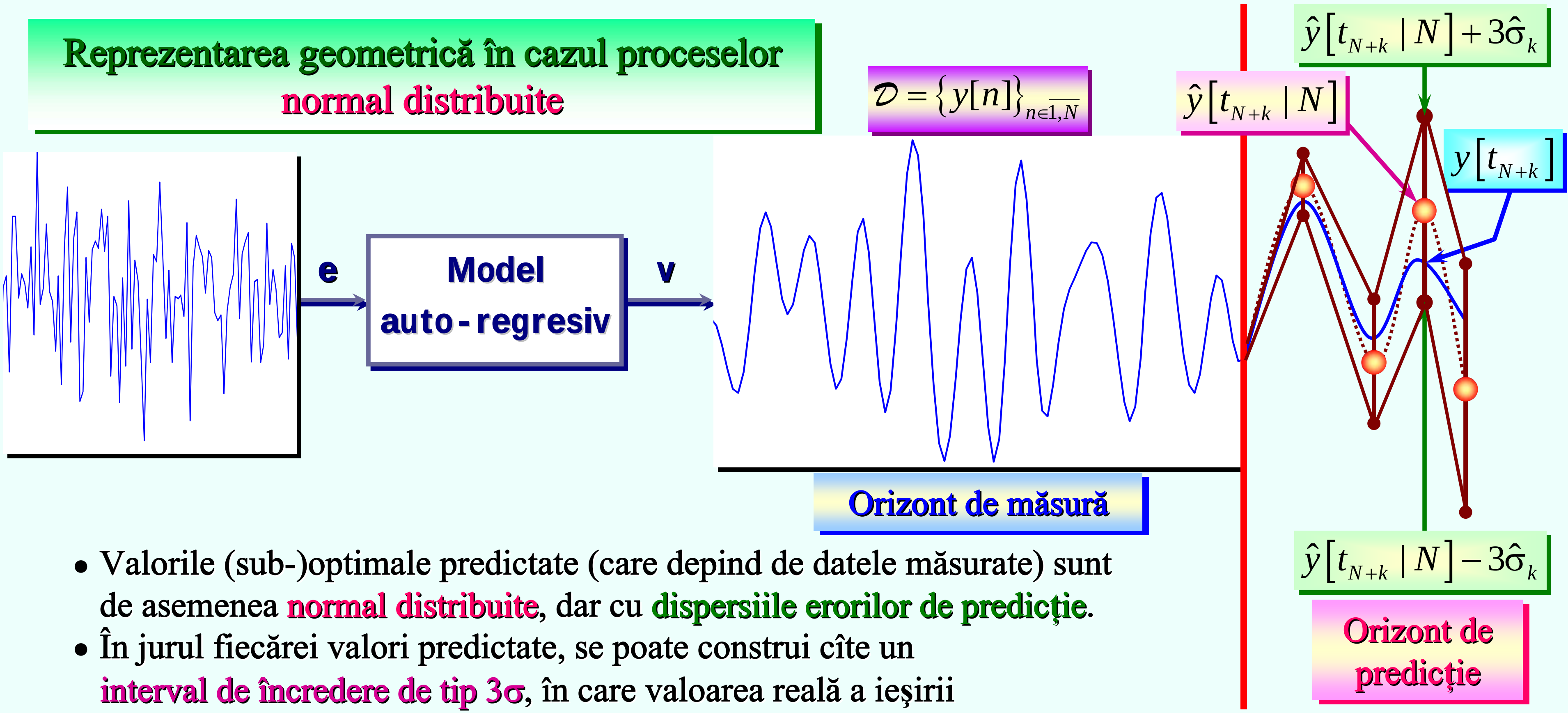
☝ Dispersia erorii de predicţie creşte odată cu îndepărtarea de orizontul de măsură.

$$K \le 5$$

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia 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 predicţie.
- În jurul fiecărei valori predictate, se poate construi cîte un interval de încredere de tip 3σ, în care valoarea reală a ieşirii procesului are şanse de peste 95% să aparţină.
- În afara valorilor predictate, pe grafic se amplasează şi intervalele de încredere, sub forma unor segmente liniare centrate în valorile predictate.
- Aceste intervale devin din ce în ce mai largi odată cu creşterea deplasamentului de predicţie, indicînd deteriorarea preciziei de predicţie.
- Curba valorilor măsurate (continuă) se îndepărtează statistic din ce în ce mai mult de curba valorilor predictate (punctată).

❽.❹ Predicţia seriei de timp

Cum se pot determina cei doi indici structurali?

Se utilizează criteriul de selecţie: **Calitatea predicţiei**

☝ Care trebuie maximizată!

$$\mathrm{PQ}_K[p,P,na] \overset{def}{=} \frac{100}{1+\dfrac{\sum_{k=1}^{K}\hat{\sigma}_k\left|y(t_{N+k})-\hat{y}[t_{N+k}\,|\,N]\right|}{\sigma_y\sum_{k=1}^{K}\hat{\sigma}_k}}\;[\%]$$

- Dependenţa de indicii structurali este **implicită** şi nu **explicită**.
- Perioada componentei sezoniere este **unic determinată** de gradul polinomului tendinţă.
- Criteriul necesită **cunoaşterea datelor pe orizontul de predicţie**.

☝ Media ponderată şi normalizată a distanţelor dintre datele măsurate şi cele predictate.

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

deviaţia standard a datelor originale

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

media datelor originale

Atunci… de ce mai este necesară predicţia?

Se apelează la următoarea strategie:

① **Se fixează lungimea orizontului de predicţie (K) şi se determină lungimea seriei de timp (N).**

② **Se utilizează numai (N-K) date pentru construcţia celor 3 modele.**

③ **Se utilizează ultimele K date pentru selecţia structurii modelelor.**

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia seriei de timp

Care este tema de laborator?

Proiectarea şi implementarea unei rutine de predicţie a seriilor de timp, bazată pe modelul cu 3 componente, predictorul optimal şi criteriul calităţii predicţiei.

☞ Paşii principali ai algoritmului de predicţie

Date de intrare

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

Iniţializare

- Determinarea lungimii seriei de timp. ⇒ N
- Separarea datelor în două. ⇒
  - $\mathcal{D}_{\text{model}} = \left\{y(t_n)\right\}_{n\in\overline{1,N-K}}$ ☜ date de modelare
  - $\mathcal{D}_{\text{struct}} = \left\{y(t_n)\right\}_{n\in\overline{N-K+1,N}}$ ☜ date de alegere a structurii
- Evaluarea perioadei de eşantionare ⇒ T_s ☜ Unitară, pentru seriile eşantionate uniform.
- Evaluarea mediei şi a deviaţiei standard pentru datele de modelare. ⇒ $\langle y\rangle$ σ_y

❽.❹ Predicţia seriei de timp

☞ Paşii principali ai algoritmului de predicţie (continuare)

Bucle iterative

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

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

- se obţin: y_{sta} y_T $\hat{\theta}_{N,p}$

② Se apelează rutina `seasonal` cu datele staţionarizate curente.

- se obţin: v y_S P $\mathcal{V}$

Pentru $na \in \overline{0, Na}$

❶ Se apelează rutina `stochastic` cu zgomotul colorat curent şi ordinul *na*.

- se obţin: $\hat{e}$ $\hat{\lambda}^2_{N,na}$ y_{AR} $\hat{\theta}^{\mathrm{AR}}_{N,na}$

❷ Se extrapolează tendinţa şi componenta sezonieră pe orizontul de predicţie.

- se obţin: $y_{T,\mathrm{extra}}$ $y_{S,\mathrm{extra}}$

❸ Se predictează zgomotul colorat.

- se obţin: $y_{AR,\mathrm{pred}}$ $\left\{\hat{\sigma}^2_k\right\}_{k \in \overline{1,K}}$

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia seriei de timp

☞ Paşii principali ai algoritmului de predicţie (continuare)

Bucle iterative

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

• • •

Pentru $na \in \overline{0, Na}$

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

❺ Se evaluează calitatea predicţiei, folosind datele predictate curente şi cele de alegere a structurii. $\mathrm{PQ}_K[p,P,na]$

Se determină punctul de maxim al calităţii predicţiei. $\left(p_0, P_0, na_0, \mathrm{PQ}_K\left[p_0, P_0, na_0\right]\right)$

☝ Calitatea predicţiei este o funcţie de două variabile: *p* şi *na*.

Se apelează rutina **`trend`** cu datele de modelare şi gradul p_0.

Se apelează rutina **`seasonal`** cu datele staţionarizate curente.

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

Se predictează seria de timp folosind modelele precedente.

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

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia 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

vectorul datelor predictate

matricea calității predicției

```
>> [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

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia seriei de timp

Programul de test al rutinei `ts_predict` — **`ISLAB_8D`**

☞ **Paşi principali**

- Se introduce numărul fişierului care conţine seria de timp. `nts`
- Se introduce lungimea orizontului de predicţie. `K`
- Se introduce gradul maxim al polinomului tendinţă. `Pmax`
- Se introduce ordinul maxim al modelului AR. `Na`
- Se încarcă seria de timp în spaţiul de lucru. • cu ajutorul funcţei: `eval`
- Se determină lungimea şi dispersia seriei de timp. `N` `lambda_y2`
- Se apelează rutina `ts_predict`.
- Se trasează graficele seriei de timp restrînse la (N-K) date şi al tendinţei (în aceeaşi fereastră, cu precizarea gradului polinomului).

Dacă $P_0 \neq 0$: Se trasează graficele seriei de timp staţionarizate şi al componentei sezoniere (în aceeaşi fereastră, cu precizarea perioadei).

Altfel: Utilizatorul este informat printr-un mesaj corespunzător.

❽ Modelarea şi predicţia seriilor de timp

❽.❹ Predicţia seriei de timp

Programul de test al rutinei `ts_predict` **`ISLAB_8D`**

`lambda_y2`

☞ **Paşi principali (continuare)**

Se estimează raportul semnal-zgomot.

$$\mathrm{SNR} = \frac{\hat{\lambda}_y^2}{\hat{\lambda}_{\hat{e}}^2}$$

☝ Dispersia zgomotului alb estimat este egală cu primul element al vectorului dispersiilor erorilor de predicţie.

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

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

Se trasează graficul suprafeţei calităţii de predicţie (în funcţie de *p* şi *na*, cu precizarea punctului de maxim pe grafic).

- cu ajutorul funcţiei: `surf`

Se trasează graficul detaliat al datelor măsurate şi predictate pe orizontul de predicţie, cu reprezentarea tubului de încredere şi precizarea valorii maxime a calităţii de predicţie.

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