Χρονοσειρές - Μάθημα 9
Ανάλυση χρονοσειρών και δυναμικά συστήματα

- Ανακατασκευή του χώρου καταστάσεων
  παρατήρηση της πολυπλοκότητας / στοχαστικότητας / δομής του συστήματος

- Εκτίμηση χαρακτηριστικών του συστήματος / ελκυστή
  μέτρηση της πολυπλοκότητας / διάστασης του συστήματος

- Μοντελοποίηση / Πρόβλεψη
  Χρήση μη-γραμμικών μοντέλων για καλύτερες προβλέψεις

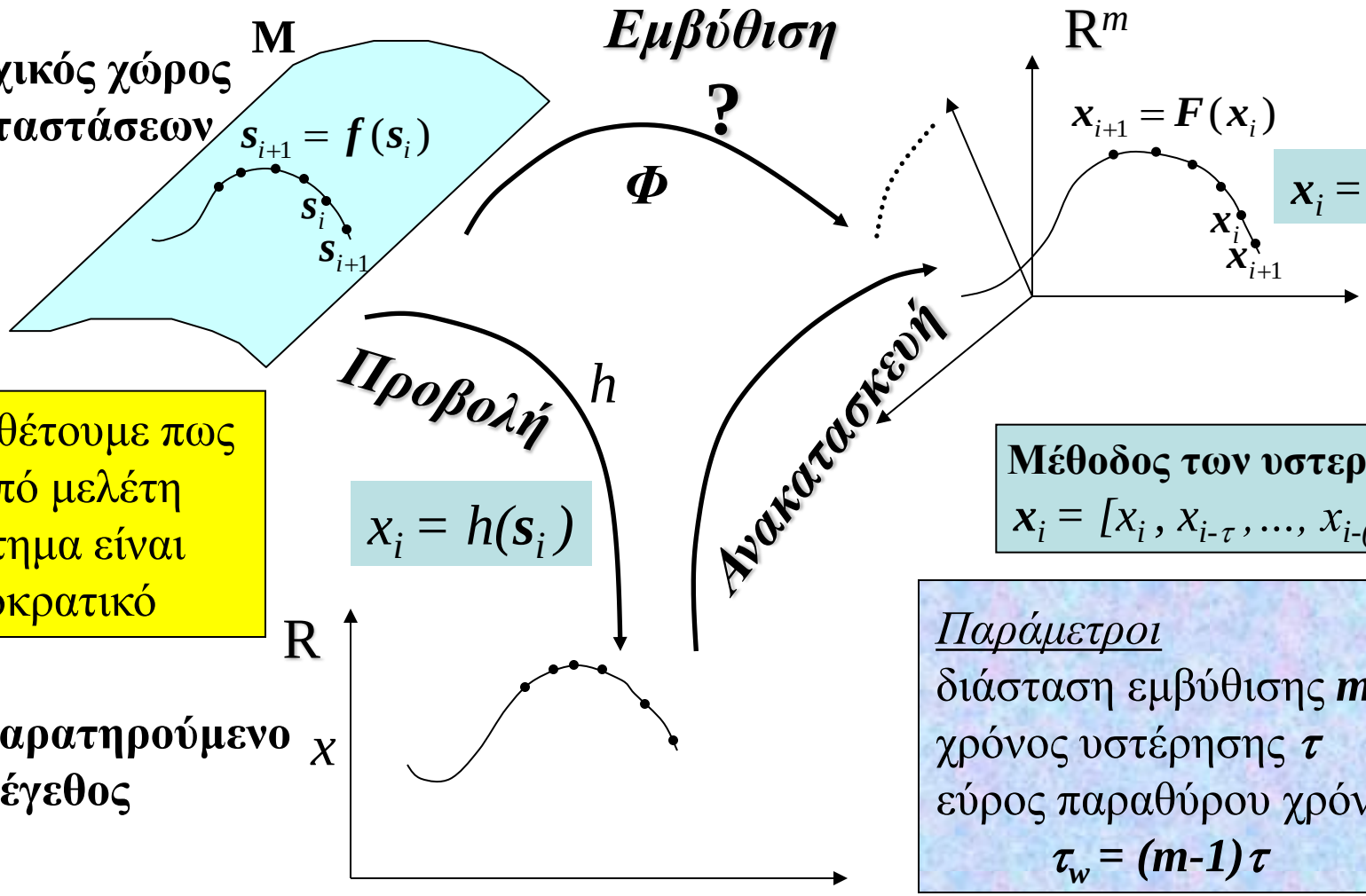
- Άλλα θέματα:
  - Έλεγχος υπόθεσης για γραμμικότητα / μη-γραμμικότητα
  - Έλεγχος εξέλιξης συστήματος (control)
  - Συγχρονισμός δυναμικών συστημάτων
  - …

Ανακατασκευή του χώρου καταστάσεων

συνθήκη: $m \geq 2D+1$

ανακατασκευασμένος χώρος καταστάσεων

αρχικός χώρος καταστάσεων



$x_i = \Phi(s_i)$

$x_i = h(s_i)$

παρατηρούμενο μέγεθος

Υποθέτουμε πως το υπό μελέτη σύστημα είναι αιτιοκρατικό

Μέθοδος των υστερήσεων

$x_i = [x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\tau}]$

Παράμετροι

διάσταση εμβύθισης m

χρόνος υστέρησης τ

εύρος παραθύρου χρόνου τ_w

$\tau_w = (m-1)\tau$

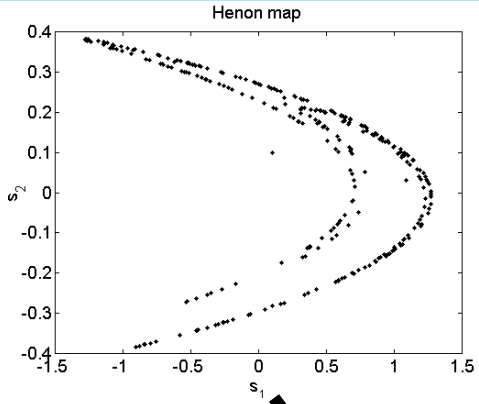
Παράδειγμα: Απεικόνιση Henon

$s(i) = 1 - 1.4\, s(i-1)^2 + 0.3 s(i-2)$

ή

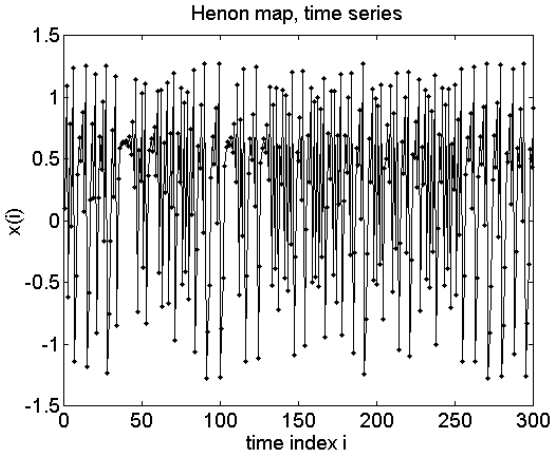
$s_1(i) = 1 - 1.4\, s_1(i-1)^2 + s_2(i-1)$

$s_2(i) = 0.3\, s_1(i-1)$



Προβολή

$x_i = s_1(i)$

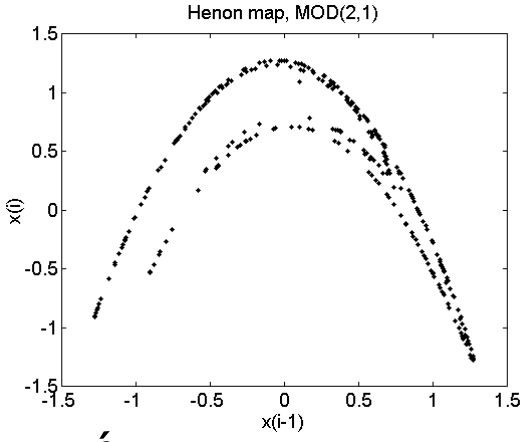


Ανακατασκευή

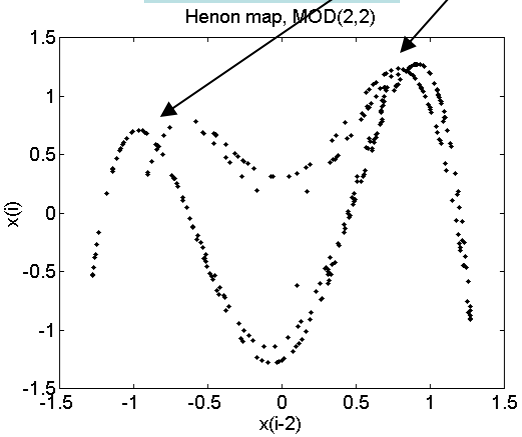
Μέθοδος των υστερήσεων

Αυτο-τομές

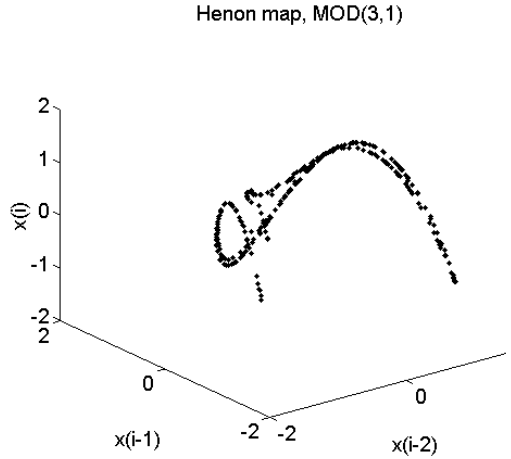
$m=2 \quad \tau=1$



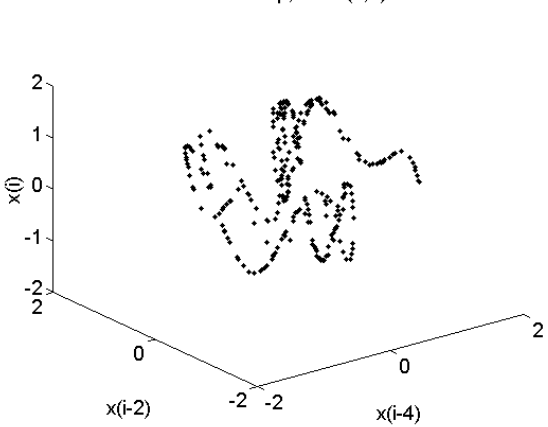
$m=2 \quad \tau=2$



$m=3 \quad \tau=1$



$m=3 \quad \tau=2$

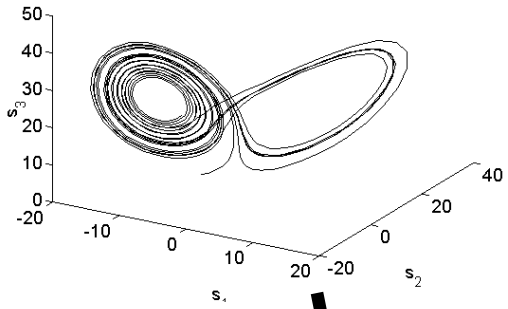


Παράδειγμα: σύστημα Lorenz

$$\dot{s}_1 = -a(s_1 - s_2)$$
$$\dot{s}_2 = -s_1 s_3 + b s_1 - s_2$$
$$\dot{s}_3 = s_1 s_2 - c s_3$$

a=10, b=28, c=8/3

Lorenz system

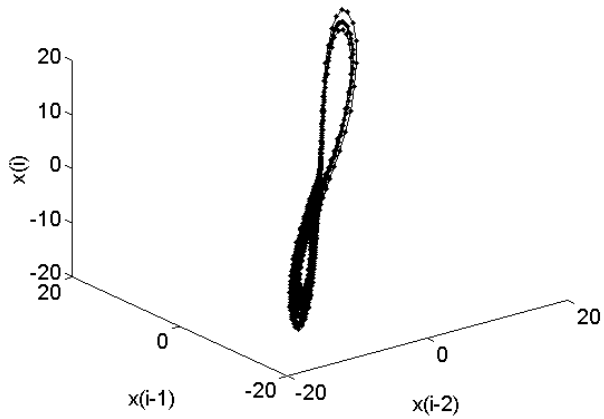


Μέθοδος υστερήσεων, ***m=3***

Βέλτιστο τ ?

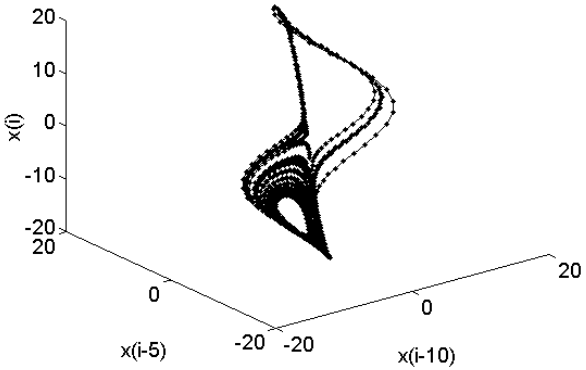
τ=1

x Lorenz, MOD(3,1)



τ =5

x Lorenz, MOD(3,5)



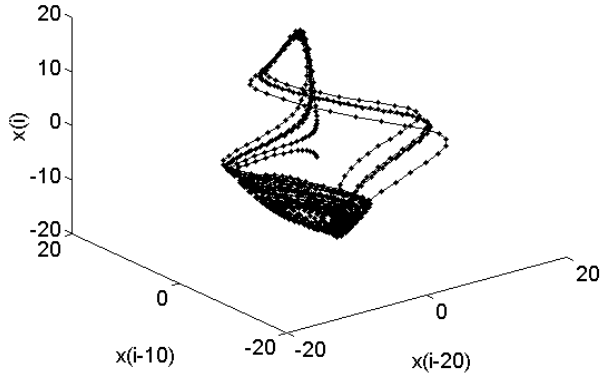
Προβολή

$x_i = s_1(i)$

Ανακατασκευή

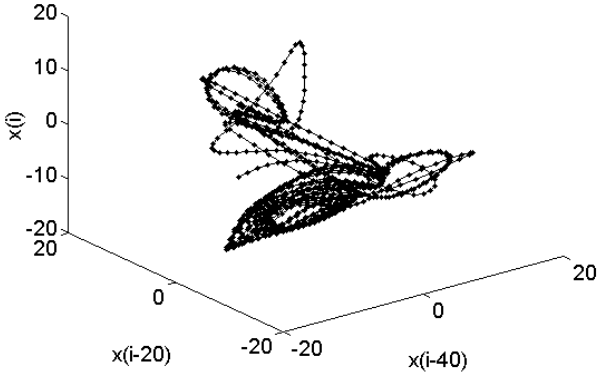
τ =10

x Lorenz, MOD(3,10)

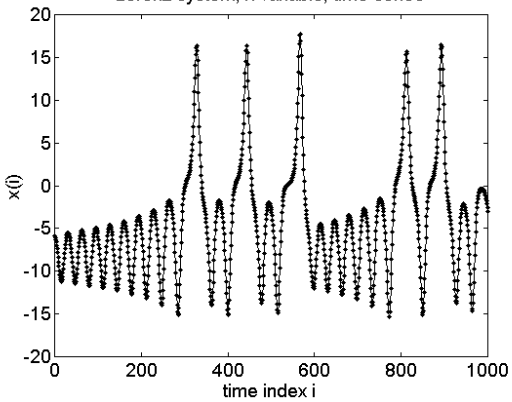


τ =20

x Lorenz, MOD(3,20)



Lorenz system, x-variable, time series



Εκτίμηση του *τ*

Εκτίμηση αμοιβαίας πληροφορίας συνεχών τυχαίων μεταβλητών: εκτιμητές, ιδιότητες (μεροληψία, αποτελεσματικότητα)

- Από την αυτοσυσχέτιση *r(τ)*

  (μετράει γραμμικές συσχετίσεις)

  τ ➔ *r(τ)* =1/e ή *τ* ➔ *r(τ)* =0

- Από την αμοιβαία πληροφορία *I(τ)*

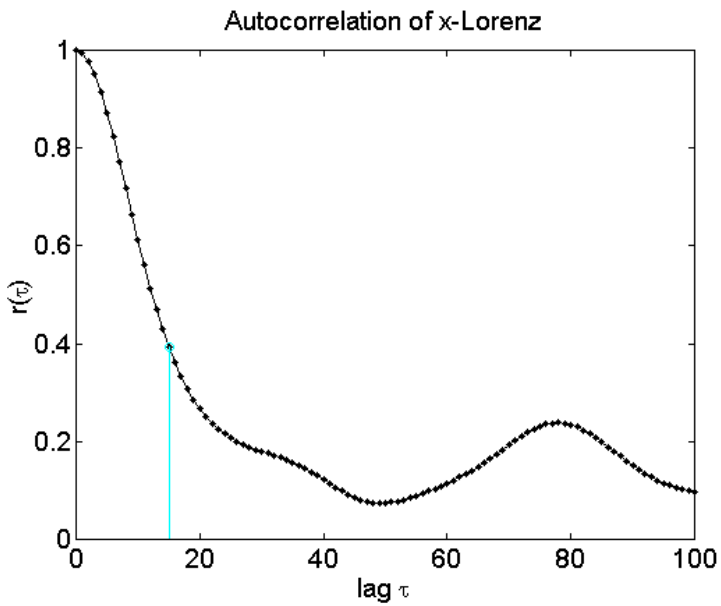
  (μετράει γραμμικές και μη-γραμμικές συσχετίσεις)

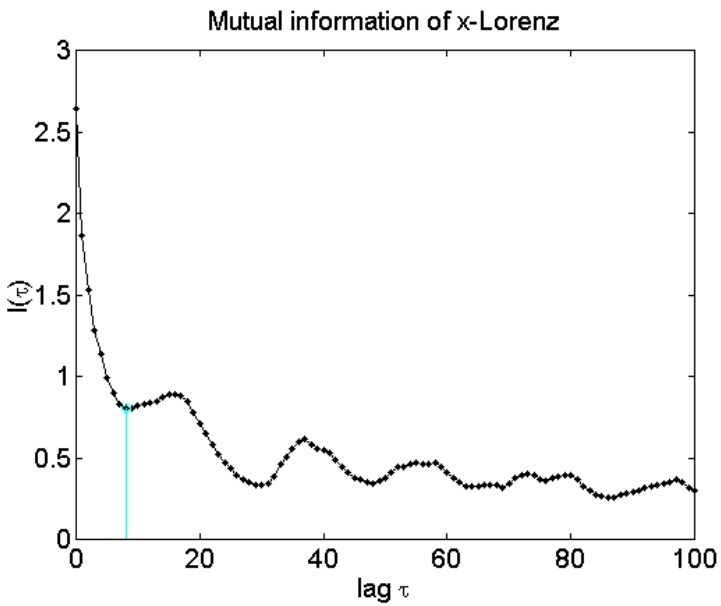
  τ ➔ πρώτο τοπικό ελάχιστο της *I(τ)*

$$I(X,Y)=\sum_{x,y} p_{XY}(x,y)\log\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}$$

$$X \to x_i \qquad Y \to x_{i-\tau}$$

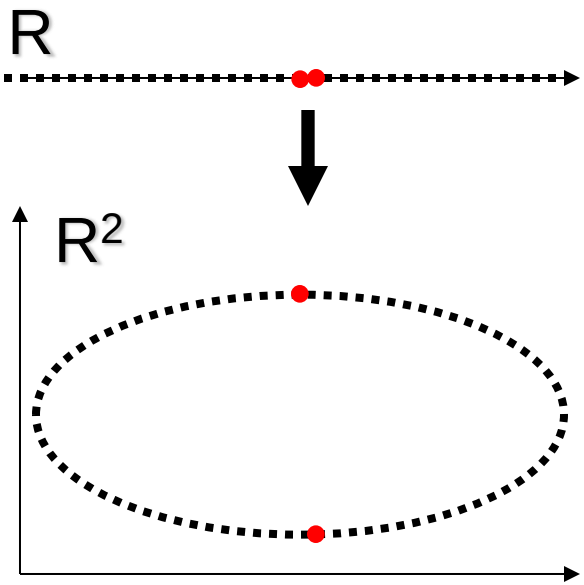
$$I(X,Y) \to I(\tau)$$





Εκτίμηση του *m*

Βέλτιστο *m* ?



- Αν *m* είναι πολύ μικρό, ο ελκυστής
→ παρουσιάζει αυτο-τομές

- Αν *m* είναι πολύ μεγάλο
→ "curse of dimensionality"

- Θεώρημα Takens: $m \geq 2D+1$ αλλά *D* άγνωστο

Μέθοδος των ψευδών γειτόνων (FNN)

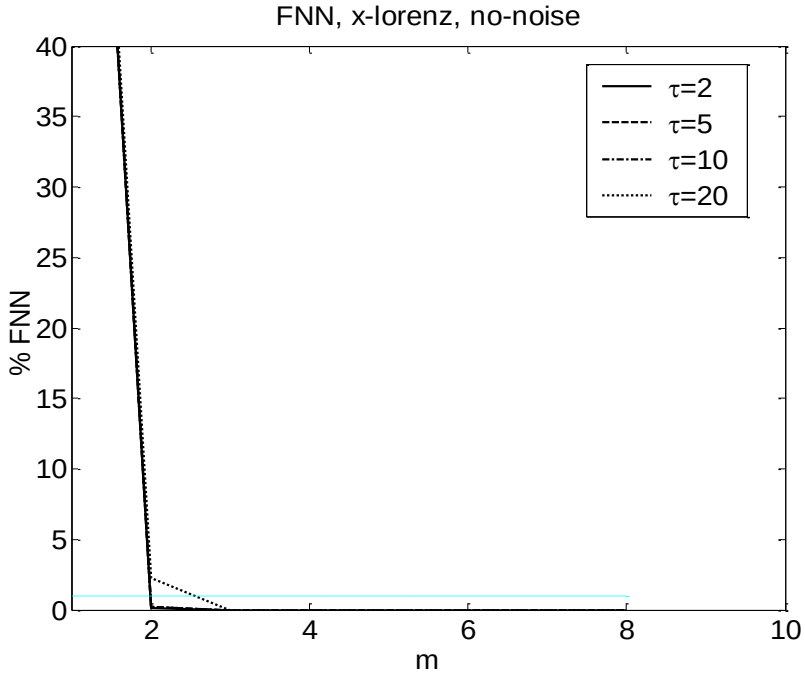
- Κοντινά σημεία του ελκυστή είναι
- είτε *πραγματικά* γειτονικά σημεία (λόγω της δυναμικής του συστήματος)
- ή *ψευδή* γειτονικά σημεία (λόγω των αυτο-τομών και μικρού *m*)

- Σε μεγαλύτερο *m* όπου δεν υπάρχουν αυτο-τομές έχουν φανερωθεί όλα τα ψευδή γειτονικά σημεία αφού δε θα είναι πια γειτονικά.

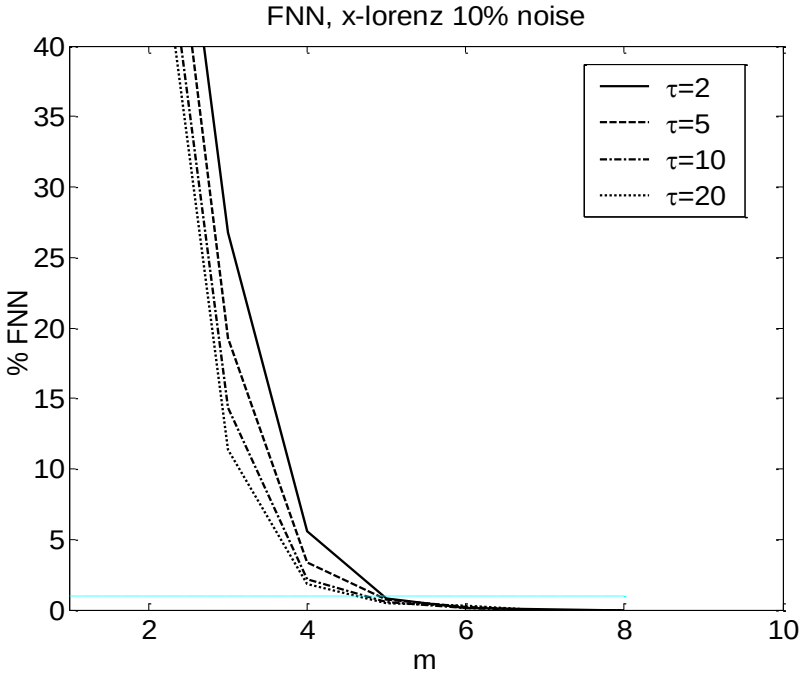
- Το βέλτιστο *m'* είναι αυτό για το οποίο δεν εμφανίζονται πλέον ψευδή γειτονικά σημεία καθώς αυξάνουμε τη διάσταση σε *m'+1*.

Παράδειγμα εκτίμησης του *m* από τη μέθοδο FNN

x-Lorenz χωρίς θόρυβο



x-Lorenz + 10% θόρυβο



Η εκτίμηση του *m* με τη μέθοδο FNN εξαρτάται από:

- την υστέρηση *t*
- θόρυβο

Εκτίμηση μη-γραμμικών χαρακτηριστικών

Μη-γραμμικά χαρακτηριστικά ή **αναλλοίωτα μέτρα**

- **Διάσταση**
  1. Ευκλείδια
  2. Τοπολογική
  3. Μορφοκλασματική

  (συσχέτισης, πληροφορίας, μέτρησης κουτιών, …)

- **Εκθέτες Lyapunov**

  (μεγαλύτερος, όλο το φάσμα)

- Εντροπία

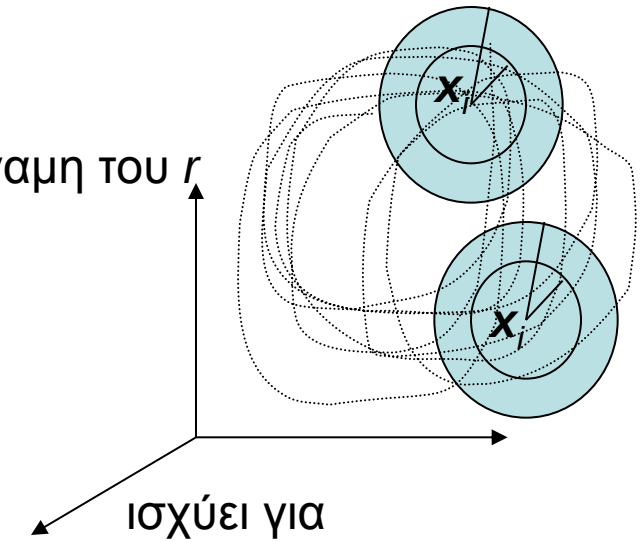
Διάσταση συσχέτισης *ν*

Η διάσταση συσχέτισης ν χαρακτηρίζει τη μορφοκλασματική δομή του ελκυστή (αυτο-ομοιότητα σε διάφορες κλίμακες) χρησιμοποιώντας την πυκνότητα των σημείων του ελκυστή στο χώρο των καταστάσεων

Η βασική ιδέα είναι πως η πιθανότητα δύο σημείων να βρίσκονται σε απόσταση μικρότερη του *r*

$$\mathrm{P}\left(\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\| < r\right)$$

μεταβάλλεται ως προς *r* ανάλογα με κάποια δύναμη του *r*



μ_i : αριθμός των σημείων που βρίσκονται μέσα σε σφαίρα με ακτίνα *r* και κέντρο $\boldsymbol{x}_i$

$$\left\langle \mu_i \right\rangle_x \approx \mathrm{P}\left(\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\| < r\right)$$

νόμο κλιμάκωσης (scaling law) $\left\langle \mu_i \right\rangle_x \sim r^{\nu}$ ισχύει για $r \to 0$, $N \to \infty$

ν ακέραιος ⇒ ελκυστής είναι συνήθης γεωμετρικό αντικείμενο

ν μη-ακέραιος ⇒ ελκυστής είναι μορφοκλασματικό αντικείμενο

Εκτίμηση της διάστασης συσχέτισης *ν*

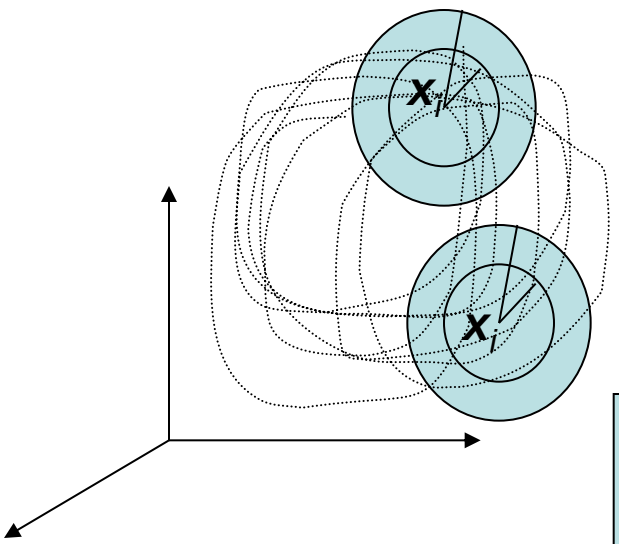
χρονοσειρά $\{x_i\},\ i=1,\ldots,N+(m-1)\tau$

ανακατασκευή $\{\boldsymbol{x}_i\},\ i=1,\ldots,N$

Εκτίμηση του $\langle \mu_i \rangle_x$

Άθροισμα συσχέτισης $$C(r)=\frac{2}{N(N-1)}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\Theta\left(r-\left\|\boldsymbol{x}_i-\boldsymbol{x}_j\right\|\right)$$

Heaviside συνάρτηση $$\Theta(x)=\begin{cases}0 & \text{όταν } x\le 0\\ 1 & \text{όταν } x>0\end{cases}$$



Νόμος κλιμάκωσης $C(r)\propto r^{\nu}$ για *r* μικρό

Εκτίμηση $\nu=\dfrac{\mathrm{d}\log C(r)}{\mathrm{d}\log r}$ για εύρος τιμών του *r*

Σύγκλιση του *ν(m)* για *m* ικανοποιητικά μεγάλο

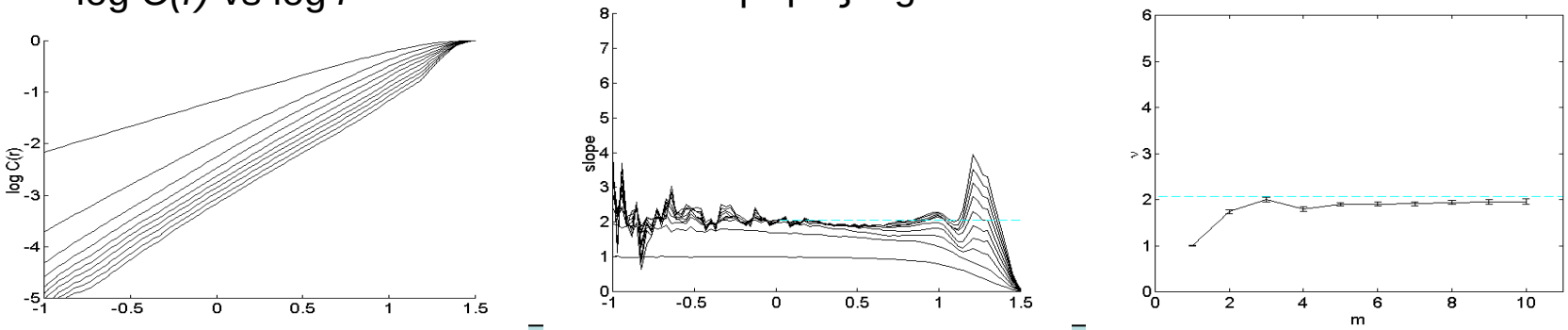
Αν *ν* μικρό και μη-ακέραιο **και** σύστημα αιτιοκρατικό

⇩

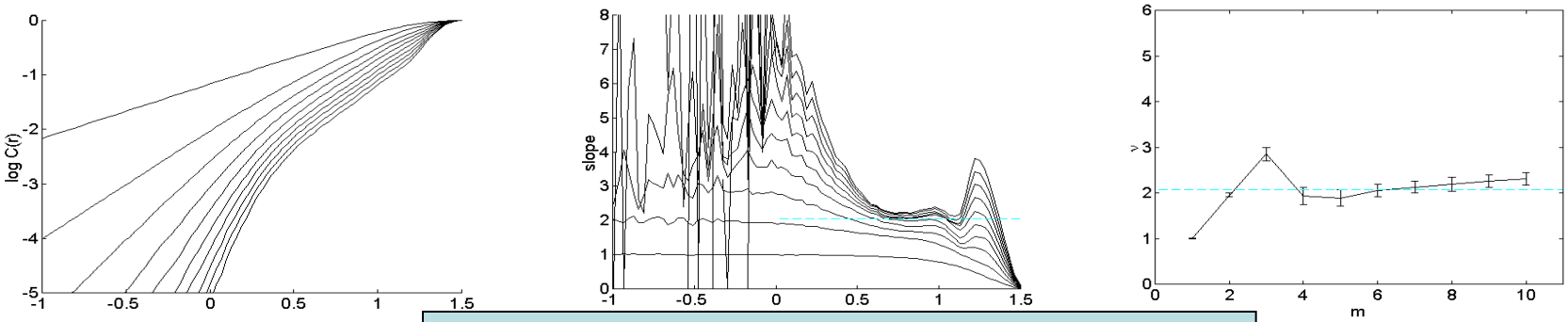
χαμηλή διάσταση και μορφοκλασματική (χαοτική) δομή

x-Lorenz χωρίς θόρυβο, τ=2

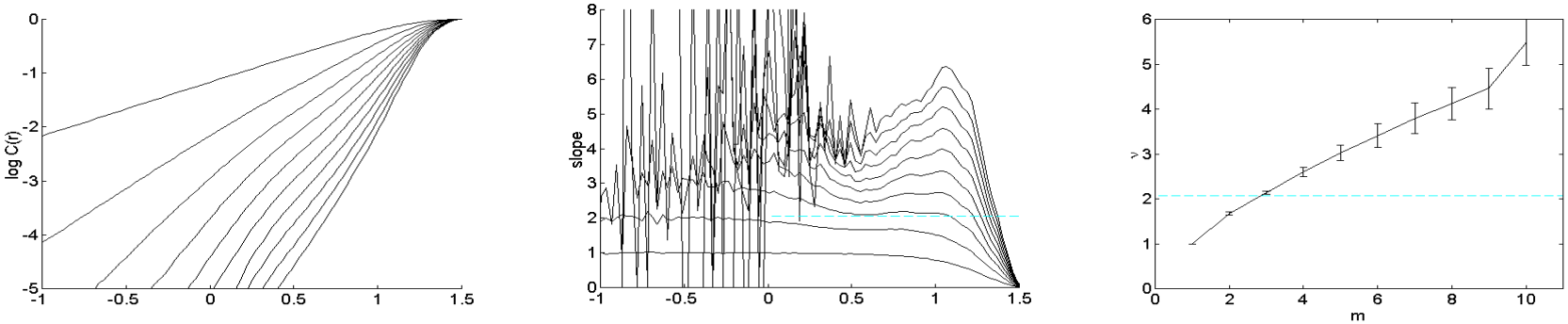
log $C(r)$ vs log r | κλίση προς log r | ν προς m



x-Lorenz + 10% θόρυβο παρατήρησης, τ=2



x-Lorenz + 10% θόρυβο παρατήρησης, τ=10



Η εκτίμηση του *v* επηρεάζεται από τους παρακάτω παράγοντες:

- χρόνος συσχέτισης ➔ $|i - j| < w$

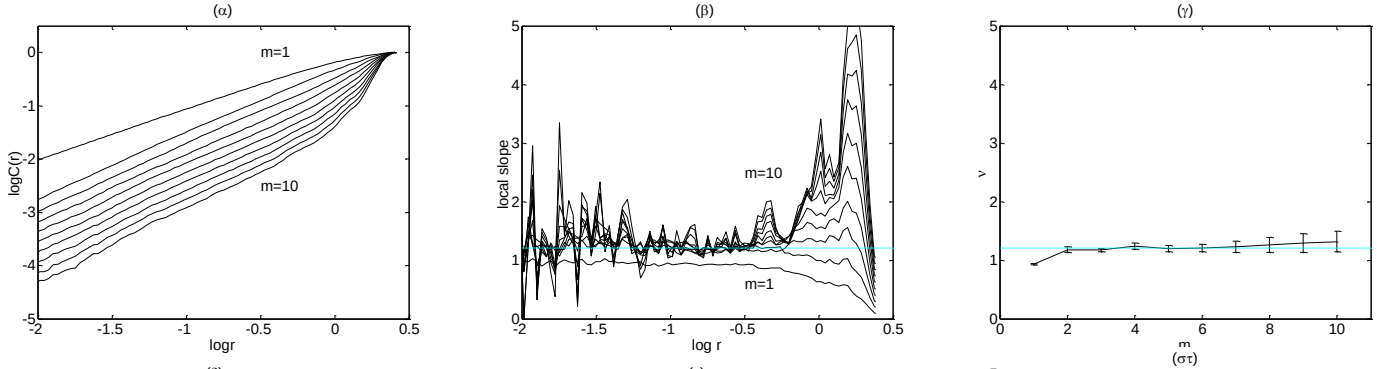
- επιλογή του *τ* και *m*

- θόρυβος

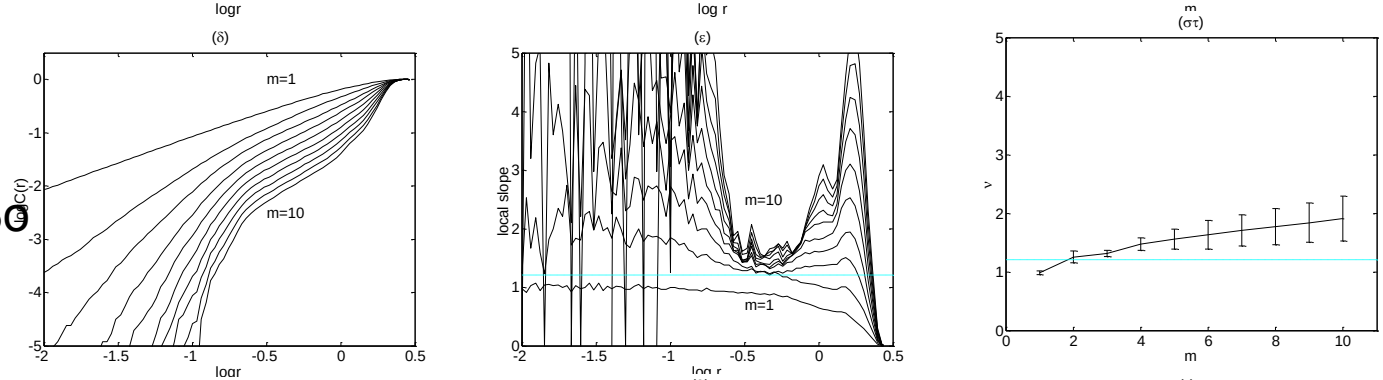
- μήκος χρονοσειράς

n=924

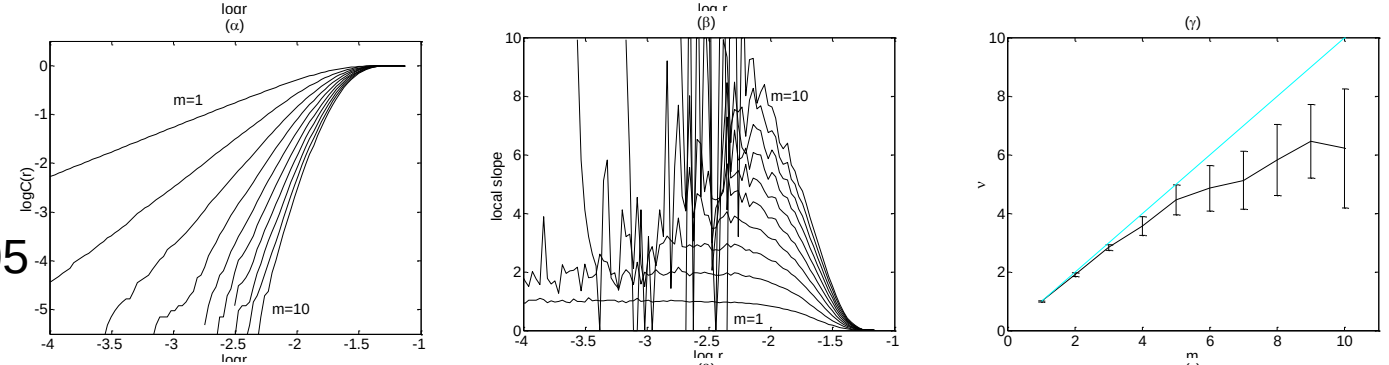
Henon



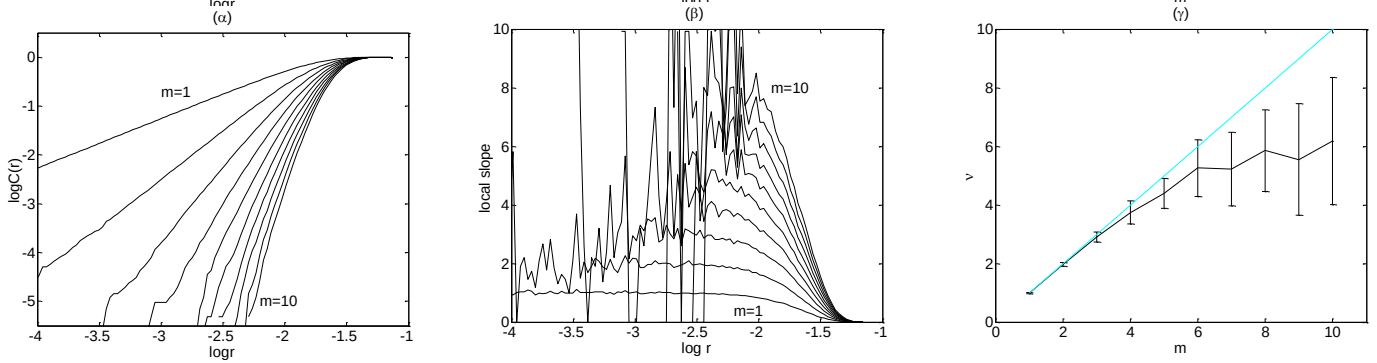
Henon
+ 10% λευκό θόρυβο



Αποδόσεις
Γενικού δείκτη
1.1.2005 – 20.9.2005

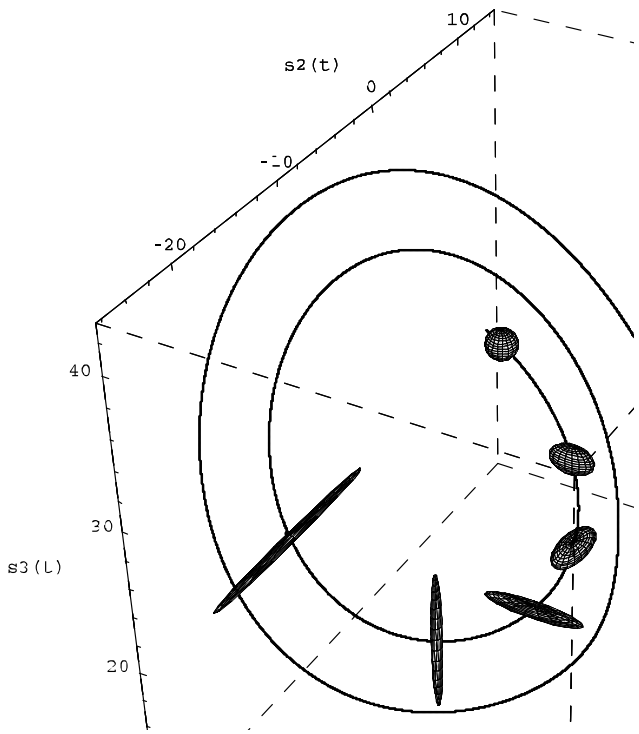


λευκός θόρυβος



Εκθέτες Lyapunov

Οι εκθέτες Lyapunov μετράνε το μέσο βαθμό απόκλισης και σύγκλισης των τροχιών στον ελκυστή, στις κατευθύνσεις του τοπικά αναλυμένου χώρου καταστάσεων



Lyapunov spectrum: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$

$l_i > 0$ ➔ απόκλιση

$l_i < 0$ ➔ σύγκλιση

$l_i = 0$ ➔ κατεύθυνση ροής

Σύστημα απώλειας ενέργειας: $\sum_{i=1}^{m} \lambda_i < 0$

Αν $l_1 > 0$ **και** το σύστημα είναι αιτιοκρατικό

⇩

χάος

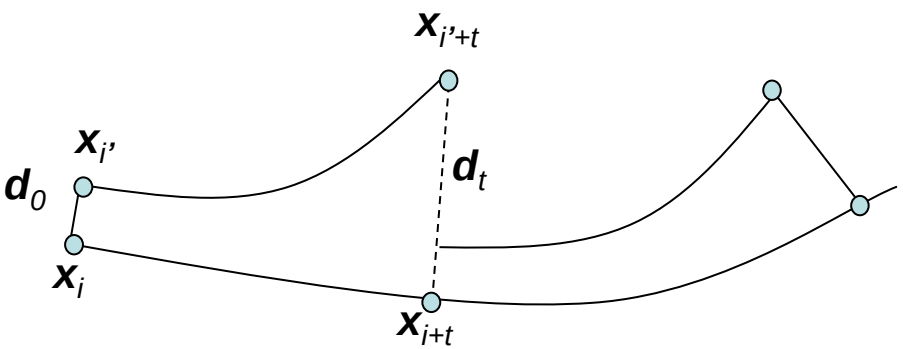
Μέγιστος εκθέτης Lyapunov λ_1

Αρχική απόσταση $\boldsymbol{d}_0 = \boldsymbol{x}_i - \boldsymbol{x}_{i'}$ δύο κοντινών τροχιών θα πρέπει να μεγαλώνει εκθετικά με το χρόνο.

Μετά από χρόνο t: $\boldsymbol{d}_t = \boldsymbol{x}_{i+t} - \boldsymbol{x}_{i'+t}$

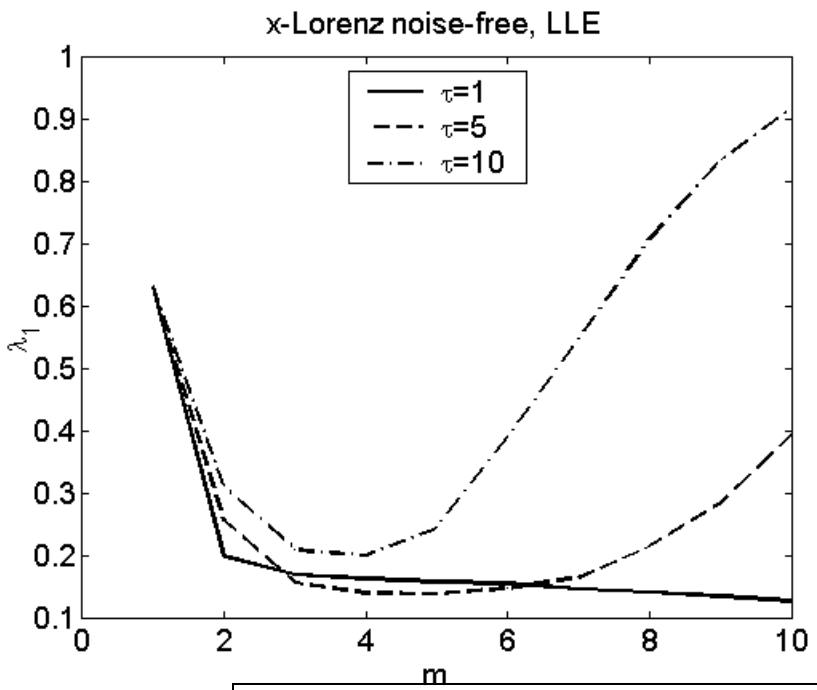
Αν $\delta_t \approx \delta_0 e^{\lambda_1 t}$ ⇒ λ_1 είναι ο μέγιστος εκθέτης Lyapunov

Υπολογισμός: $$\lambda_1 = \frac{1}{Nt} \sum_{j=1}^{N} \ln \frac{\delta_{t,j}}{\delta_{0,j}}$$

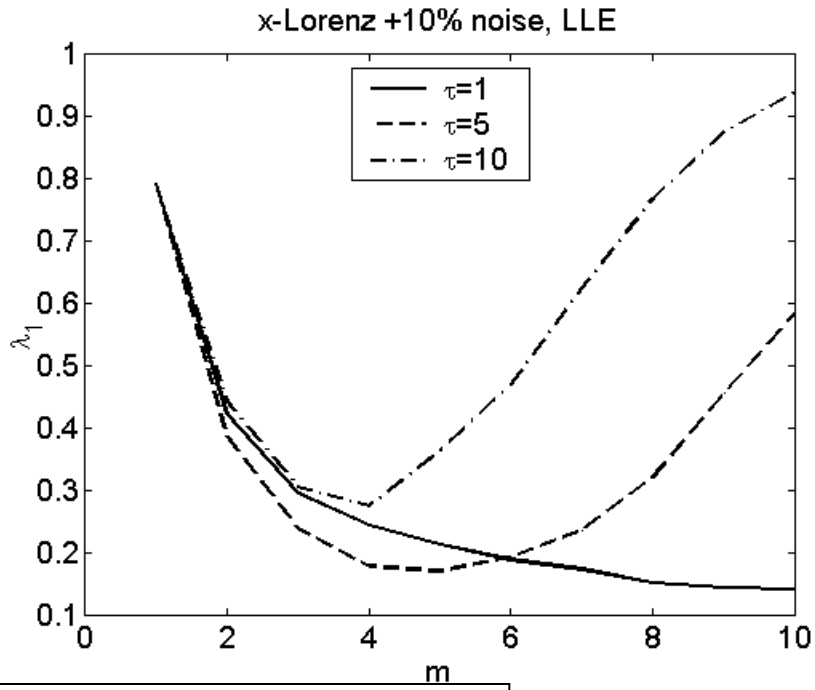


Παράδειγμα: x-Lorenz

χωρίς θόρυβο



με 10%-θόρυβο



Η εκτίμηση του λ_1 εξαρτάται από: ***τ***, ***m***, **θόρυβο**

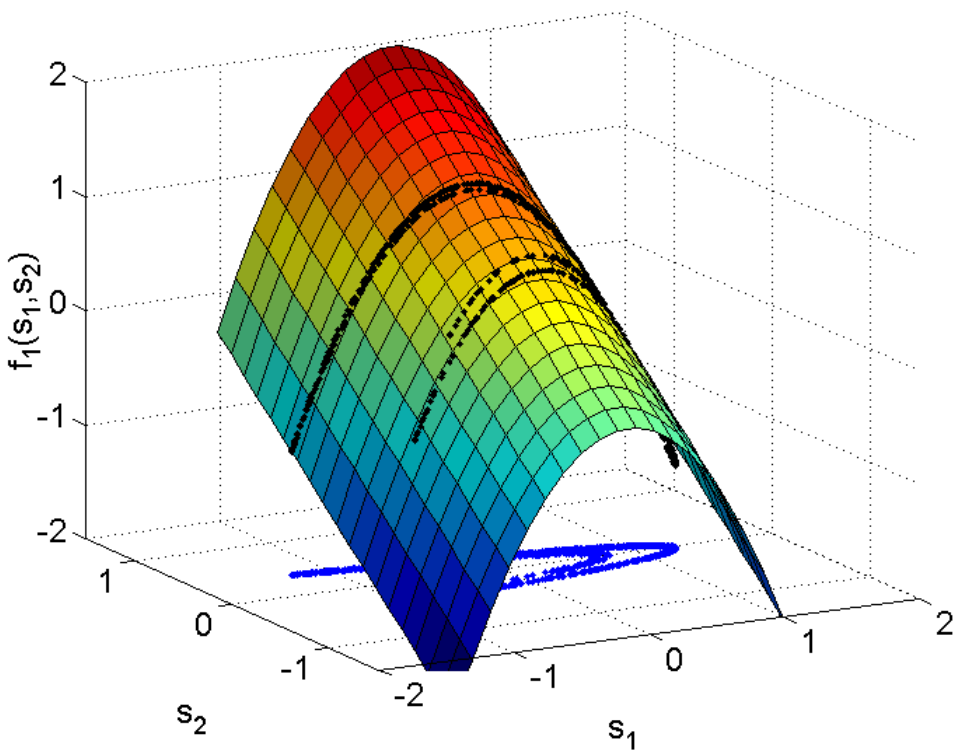
Άλλα χαρακτηριστικά / μέτρα:

26 **Προσεγγιστική εντροπία (approximate entropy): παρουσίαση, παραδείγματα, εφαρμογές**

27 **Κάποιο άλλο μη-γραμμικό χαρακτηριστικό / μέτρο**

Μοντέλα πρόβλεψης

Το σύστημα που παράγει τη χρονοσειρά: $\boldsymbol{s}_{i+1} = \boldsymbol{f}(\boldsymbol{s}_i)$



απεικόνιση Henon

$$s_{1,i+1} = 1 - 1.4 s_{1,i}^2 + s_{2,i}$$

$$s_{2,i+1} = 0.3 s_{1,i}$$

$$\boldsymbol{s}_i \xrightarrow{\boldsymbol{f}} \boldsymbol{s}_{i+1}$$

$$(s_{1,i}, s_{2,i}) \xrightarrow{f_1} s_{1,i+1}$$

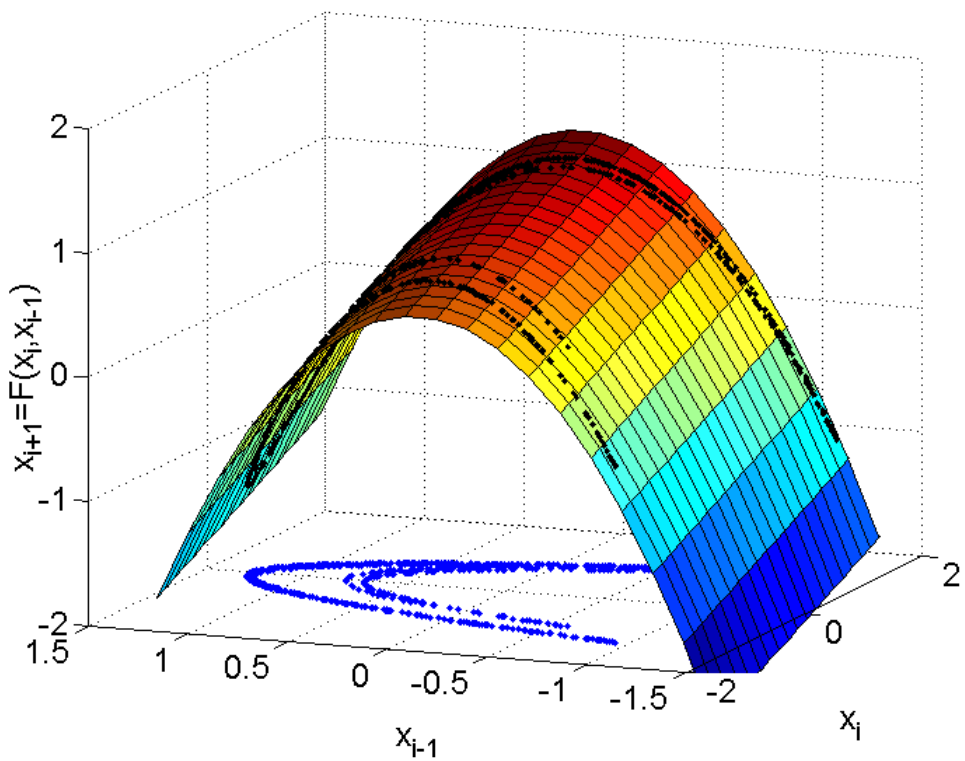
$$(s_{1,i}, s_{2,i}) \xrightarrow{f_2} s_{2,i+1}$$

Μοντέλα πρόβλεψης

Το σύστημα που παράγει τη χρονοσειρά: $\boldsymbol{s}_{i+1} = \boldsymbol{f}(\boldsymbol{s}_i)$ άγνωστο

Το ανακατασκευασμένο σύστημα από τη χρονοσειρά: $\boldsymbol{x}_{i+1} = \boldsymbol{F}(\boldsymbol{x}_i)$ εκτίμηση ?

Το πρόβλημα της μοντελοποίησης / πρόβλεψης χρονοσειρών:
δίνονται $x_1, x_2, \ldots x_i$, εκτίμηση / πρόβλεψη του x_{i+1}



Ανακατασκευή του χώρου καταστάσεων
με τη μέθοδο των υστερήσεων:

$$\boldsymbol{x}_i = [x_i, x_{i-t} \ldots, x_{i-(m-1)t}]$$

Η συνάρτηση που μας ενδιαφέρει να εκτιμήσουμε για χρονοσειρές είναι:

$$\boldsymbol{x}_{i+1} = \boldsymbol{F}(\boldsymbol{x}_i) \quad \boldsymbol{F} : \mathbb{R}^m \mapsto \mathbb{R}^m$$

$$\downarrow$$

$$x_{i+1} = F(\boldsymbol{x}_i) \quad F : \mathbb{R}^m \mapsto \mathbb{R}$$

$$m = 2,\ \tau = 1 \quad x_{i+1} = F(x_i, x_{i-1})$$

Μη-γραμμικά μοντέλα πρόβλεψης

- *Σφαιρικά μοντέλα*, π.χ. πολυώνυμα

  η *F* έχει αναλυτική μορφή και ίδια για όλο το πεδίο ορισμού

- *Τοπικά μοντέλα*, π.χ. τοπικό γραμμικό μοντέλο

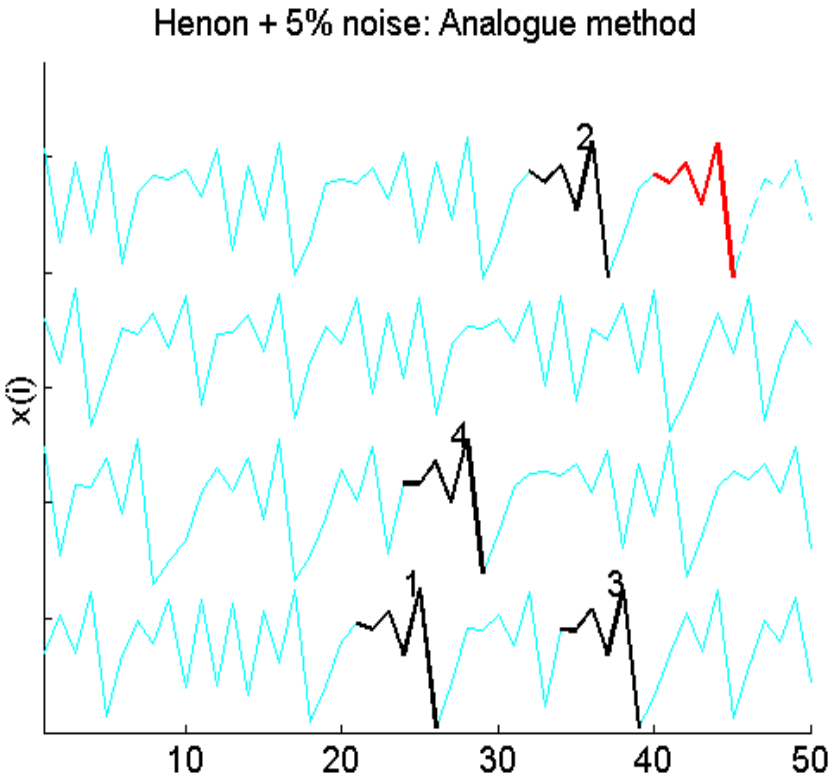
  η *F* ορίζεται διαφορετικά για κάθε σημείο του χώρου

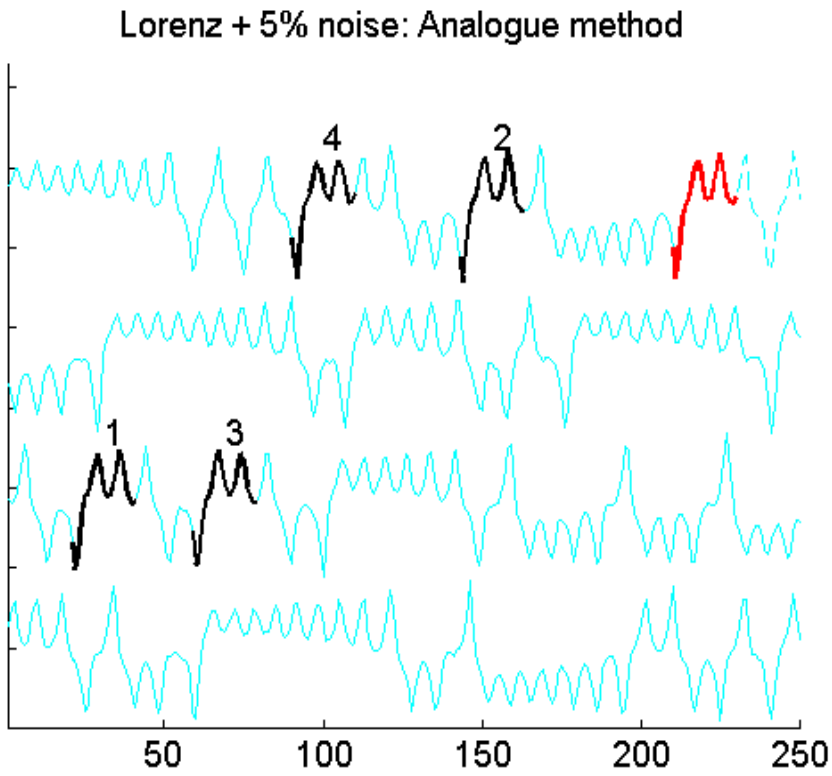
- *Ημι-τοπικά μοντέλα*, π.χ. νευρωνικά δίκτυα

  η *F* έχει μια μορφή που ορίζεται ως σταθμισμένο άθροισμα τοπικών στοιχειωδών συναρτήσεων

Πρόβλεψη με την εύρεση όμοιων τμημάτων της χρονοσειράς

Πρόβλεψη για χρόνο *i+T* από τις εικόνες για T χρονικά βήματα «όμοιων» τμημάτων της χρονοσειράς από το παρελθόν





Τοπικά Μοντέλα Πρόβλεψης

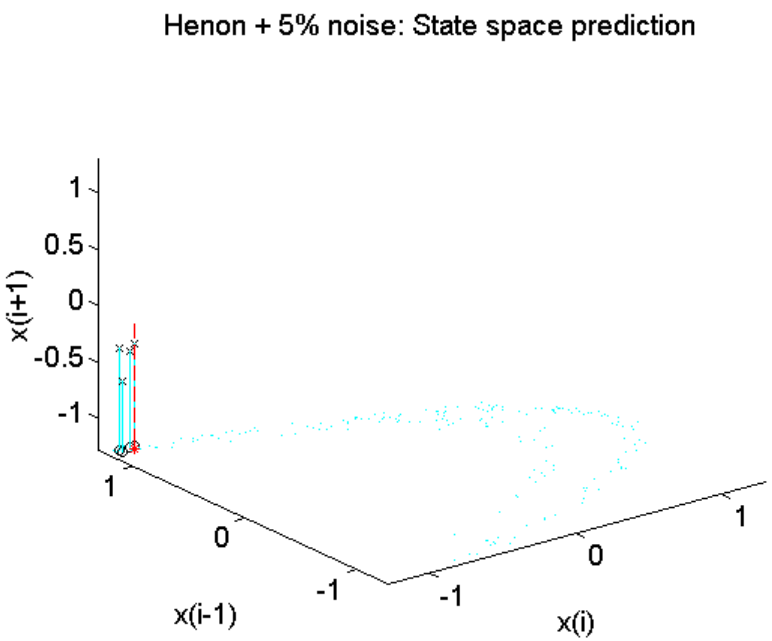
Εφαρμογή της ιδέας «όμοιων» τμημάτων:
τμήματα χρονοσειράς ➔ ανακατασκευασμένα σημεία

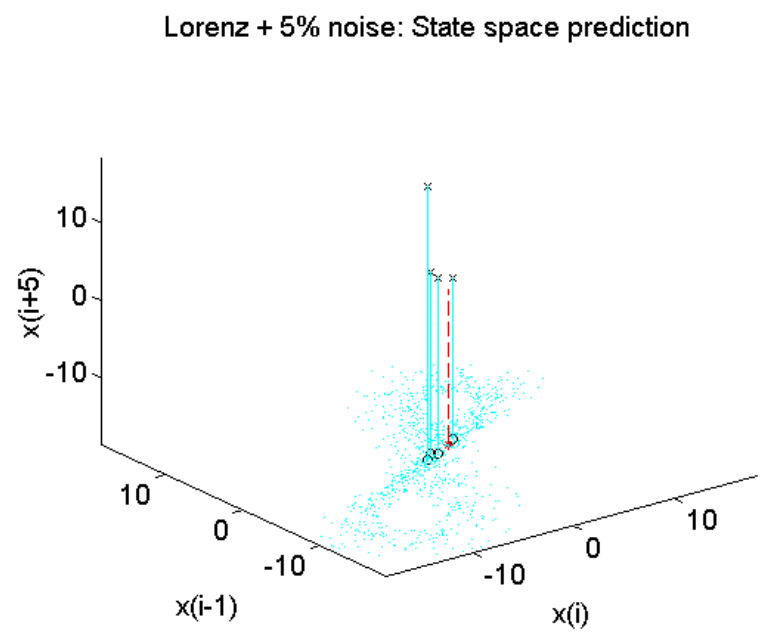
Κοντινότερα γειτονικά σημεία του $\boldsymbol{x}_i$:$\{\boldsymbol{x}_{i(1)}, \boldsymbol{x}_{i(2)}, \ldots, \boldsymbol{x}_{i(K)}\}$

Πρόβλεψη του $\boldsymbol{x}_{i+T}$ από τις εικόνες των γειτόνων: $\{x_{i(1)+T}, x_{i(2)+T}, \ldots, x_{i(K)+T}\}$

Πρόβλεψη μηδενικής τάξης: $\hat{x}_{i+T} \equiv x_i(T) = x_{i(1)+T}$

Πρόβλεψη μέσου όρου: $x_i(T) = \frac{1}{K}\sum_{j=1}^{K} x_{i(j)+T}$





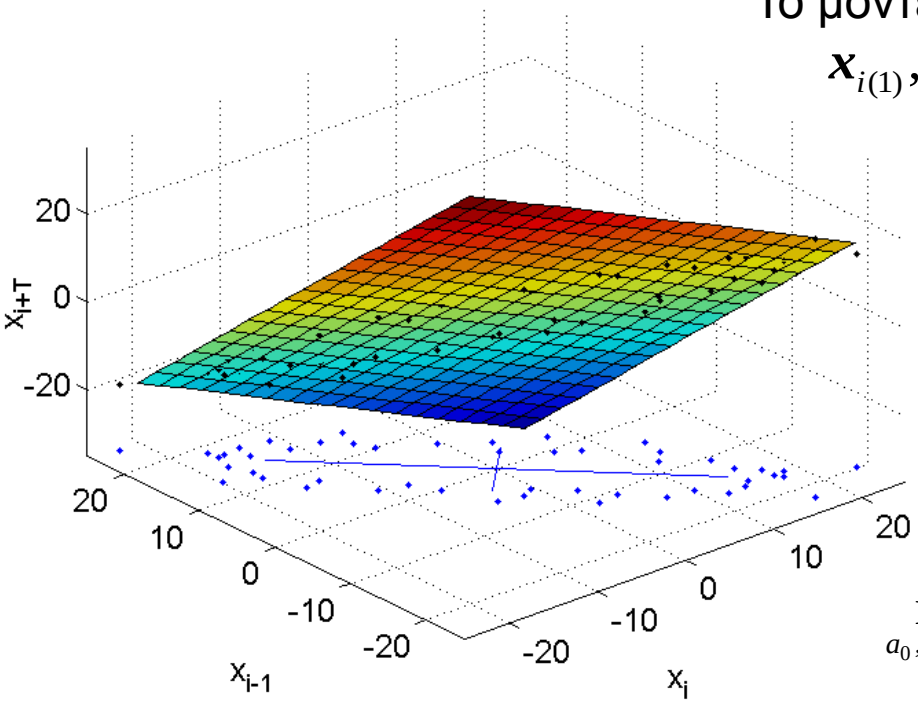
Τοπικά γραμμική πρόβλεψη

Υποθέτουμε ότι για τη «γειτονιά» του $\boldsymbol{x}_i$ ισχύει το τοπικά γραμμικό μοντέλο :

$$x_{i+1} = F(\boldsymbol{x}_i) = F(x_i, x_{i-\tau}, \ldots, x_{i-(m-1)\tau})$$
$$= a_0 + a_1 x_i + a_2 x_{i-\tau} + \cdots + a_m x_{i-(m-1)\tau}$$
$$= a_0 + \boldsymbol{a}' \boldsymbol{x}_i$$

Το μοντέλο ισχύει για τα $\boldsymbol{x}_{i(1)}, \boldsymbol{x}_{i(2)}, \ldots, \boldsymbol{x}_{i(K)}$

$$x_{i(1)+T} = a_0 + \boldsymbol{a}' \boldsymbol{x}_{i(1)}$$
$$x_{i(2)+T} = a_0 + \boldsymbol{a}' \boldsymbol{x}_{i(2)}$$
$$\vdots$$
$$x_{i(K)+T} = a_0 + \boldsymbol{a}' \boldsymbol{x}_{i(K)}$$



Εκτίμηση παραμέτρων $a_0, a_1, \ldots, a_m$ (μέθοδος ελαχίστων τετραγώνων)

$$\min_{a_0, a_1, \ldots, a_m} \sum_{j=1}^{K} (x_{i(j)+1} - a_0 - a_1 x_{i(j)} - \cdots - a_m x_{i(j)-(m-1)\tau})^2$$

Εκτίμηση λάθους πρόβλεψης

Χωρίζουμε τη χρονοσειρά σε δύο μέρη:

$$x_i(T)$$

Προβλέψεις $\hat{x}_{N_1+1}, \ldots, \hat{x}_N$

$$\underbrace{x_1, x_{2,} \ldots, x_{N_1}}_{\text{Σύνολο εκμάθησης του μοντέλου}}, \underbrace{x_{N_1+1}, \ldots, x_N}_{\text{Σύνολο για τον έλεγχο του μοντέλου}}$$

$$e_{i+T} = x_{i+T} - \hat{x}_{i+T}$$

σφάλμα πρόβλεψης

Στατιστική για το σφάλμα πρόβλεψης

$$\mathrm{NRMSE}(T) = \sqrt{\frac{\frac{1}{N-T-N_1}\sum_{t=N_1+1}^{N-T}\left(x_{t+T} - \hat{x}_{t+T}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}}$$

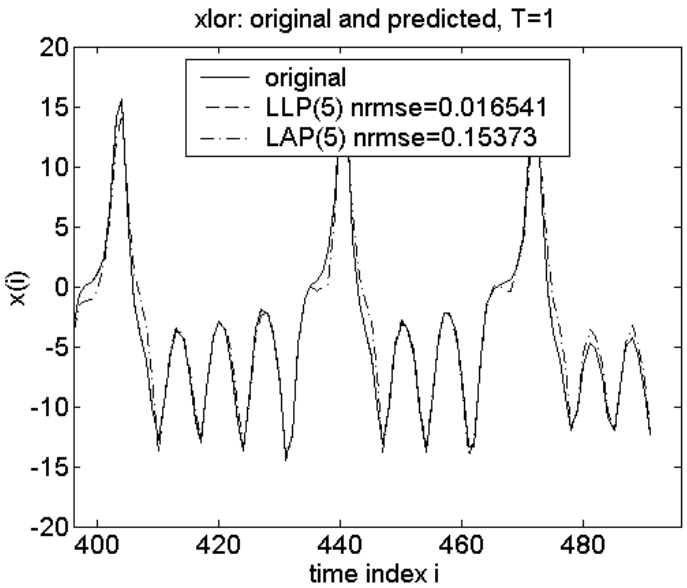
Παράδειγμα: x-Lorenz

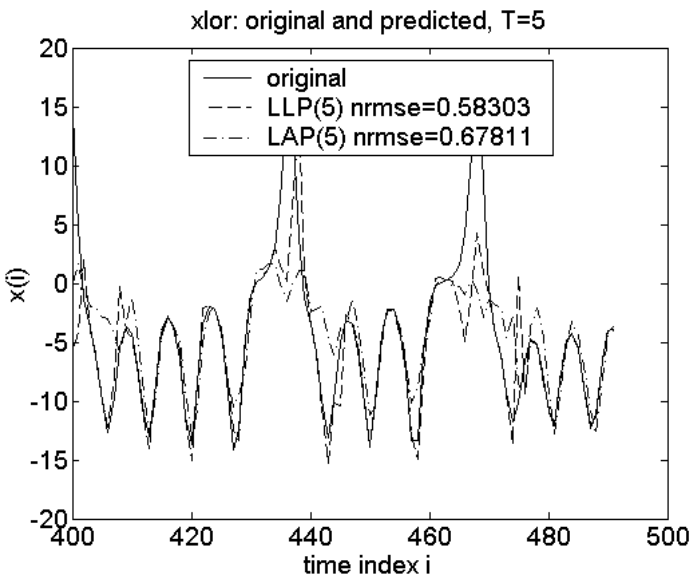
Πρόβλεψη με:

- τοπικό γραμμικό μοντέλο πρόβλεψης (LLP)
- πρόβλεψη τοπικού μέσου όρου (LAP)

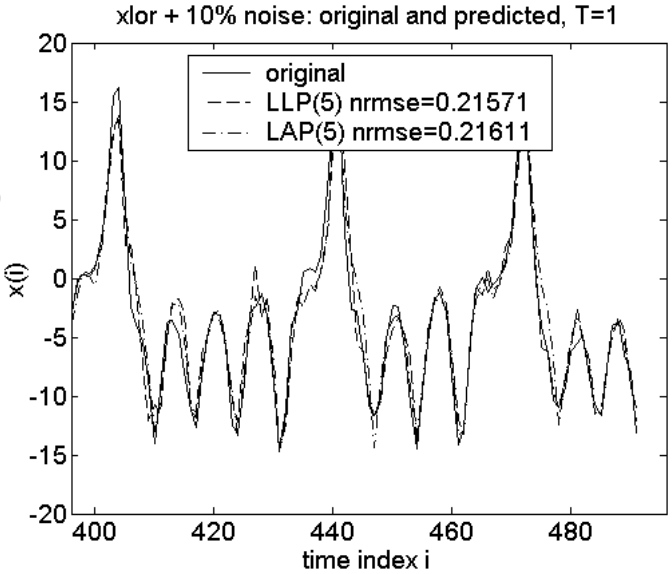
$$\tau = 1,\ m = 5,\ K = 11$$

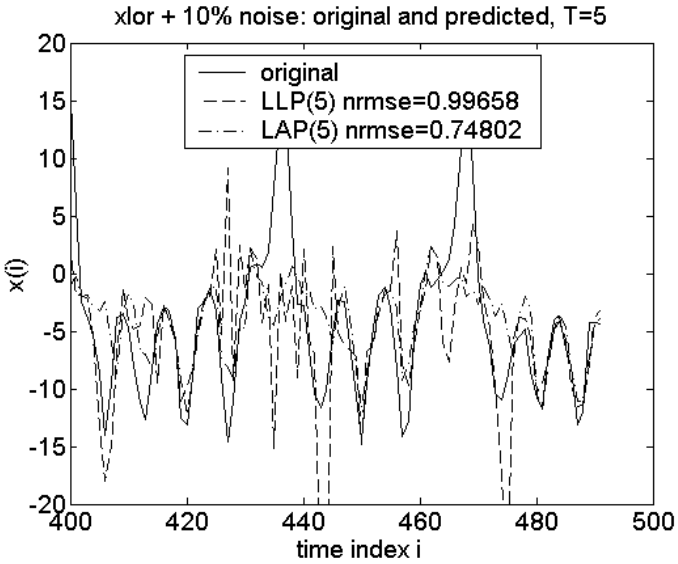
χωρίς θόρυβο





με 10%-θόρυβο

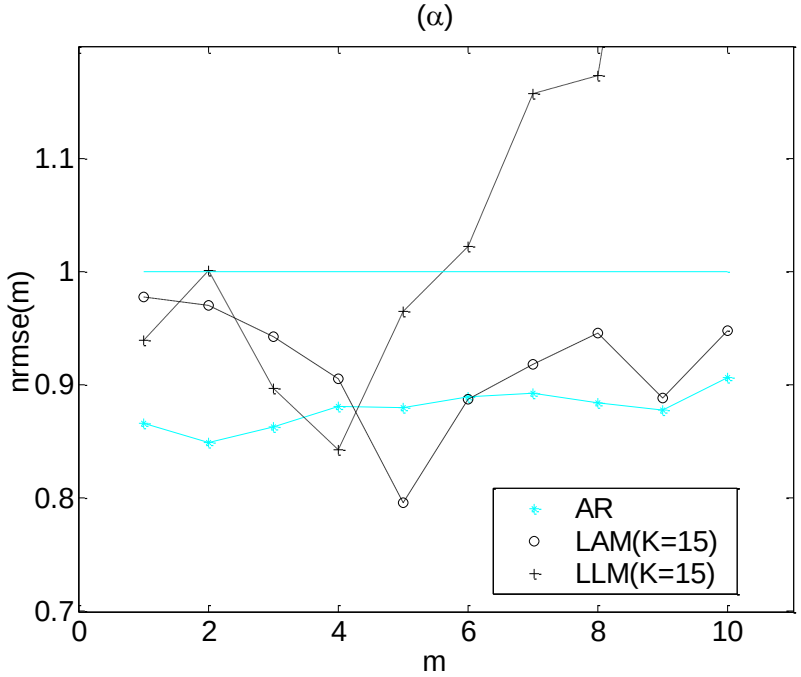




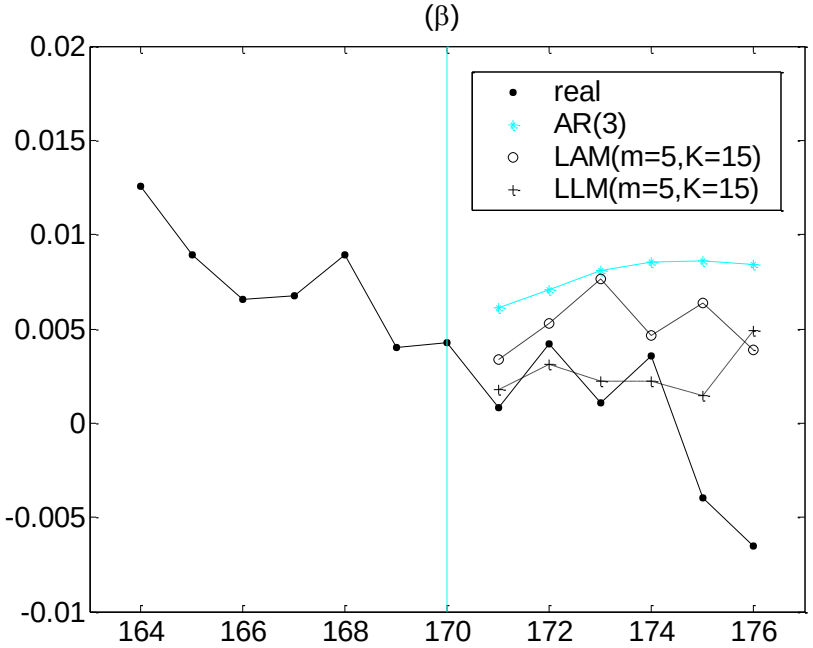
τετραμηνιαίος ρυθμός μεταβολής ΑΕΠ των ΗΠΑ την περίοδο 1947 – 1991

Πρόβλεψη με
- γραμμικό μοντέλο AR
- τοπικό μοντέλο μέσου όρου, LAM
- τοπικό γραμμικό μοντέλο, LLM

Σφάλμα πρόβλεψης (nrmse) για τα 30 τελευταία τετράμηνα



Προβλέψεις με αφετηρία το πρώτο τετράμηνο του 1989 και ορίζοντα τα 6 τελευταία τετράμηνα



Γενικός δείκτης ΧΑΑ την περίοδο 1.1.2002 – 20.9.2005

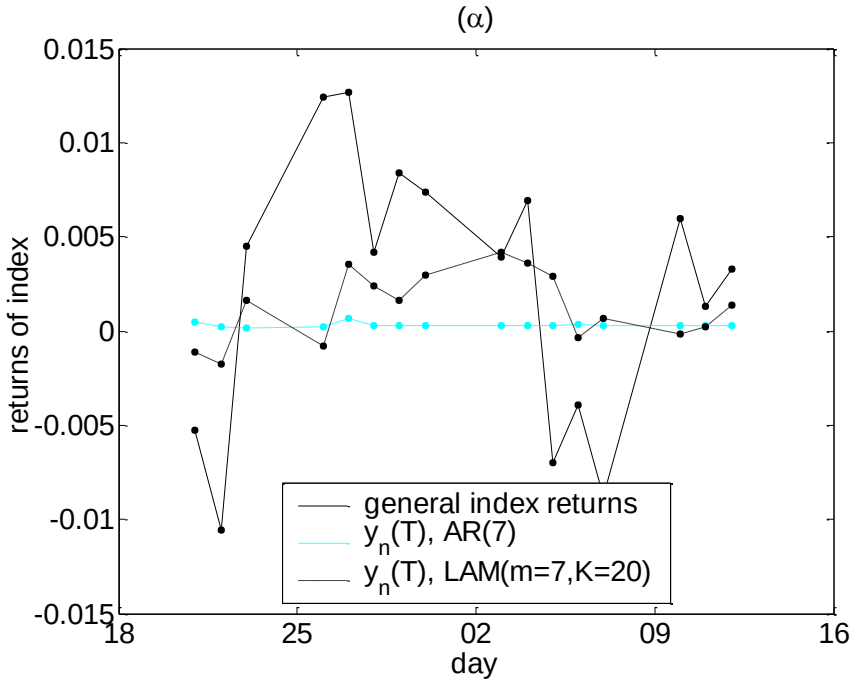
Πρόβλεψη με
- γραμμικό μοντέλο AR
- τοπικό μοντέλο μέσου όρου, LAM

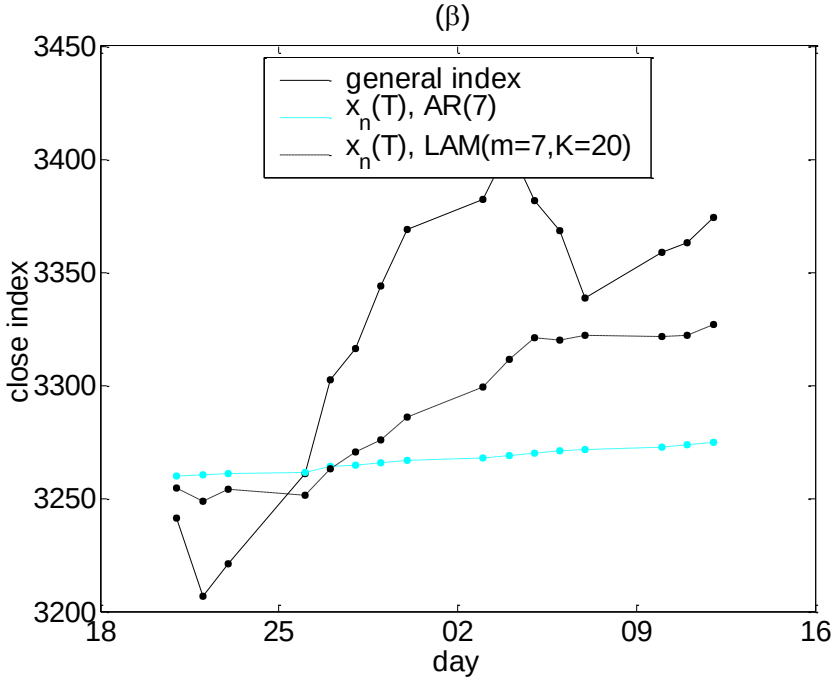
Πρόβλεψη με αφετηρία την 20.9.2005
και για 1 μέχρι 16 μέρες μπροστά

Αποδόσεις $y_t = \dfrac{x_t - x_{t-1}}{x_{t-1}}$

γενικού δείκτη

Γενικός δείκτης



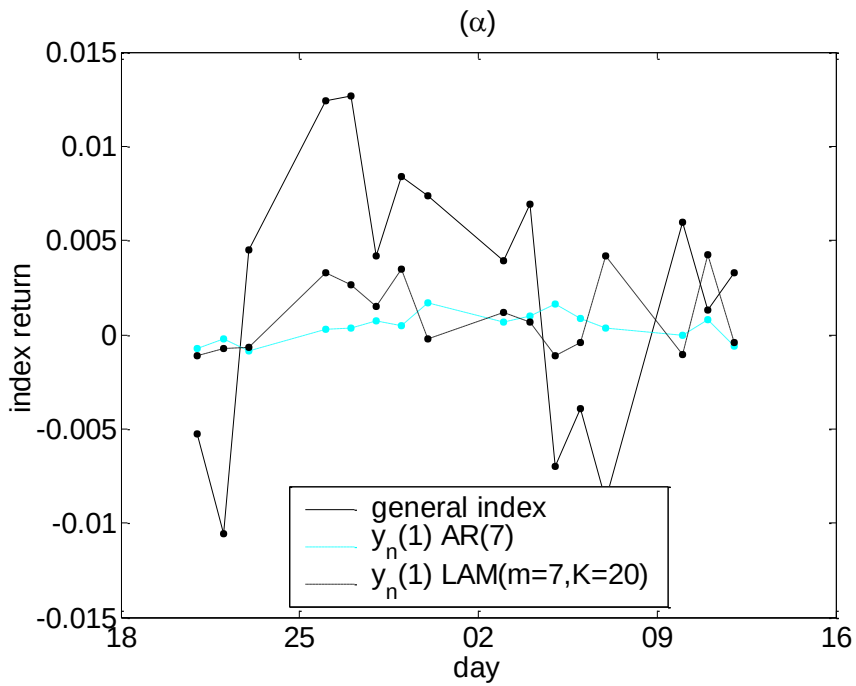
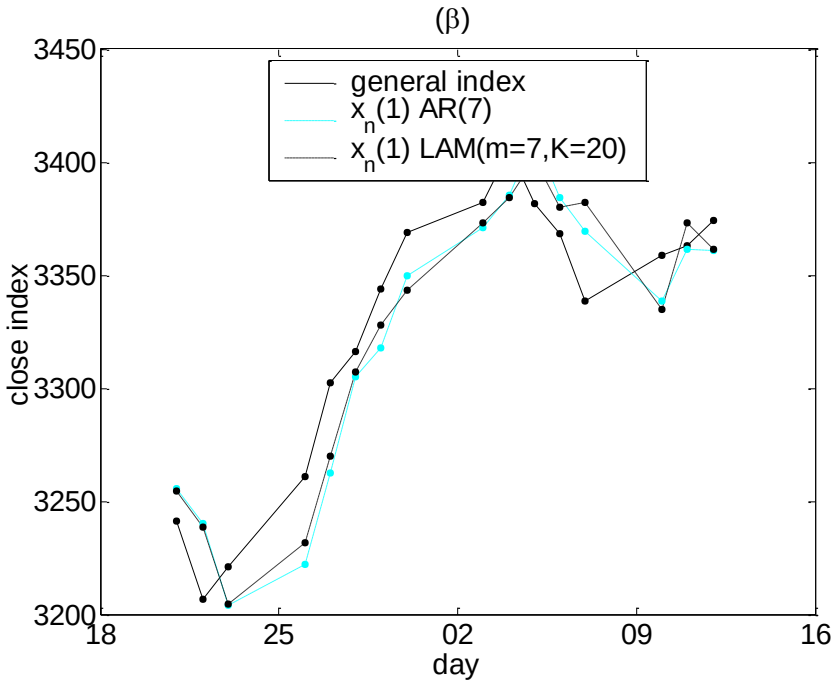


Γενικός δείκτης ΧΑΑ την περίοδο 1.1.2002 – 20.9.2005 - συνέχεια

Πρόβλεψη με
- γραμμικό μοντέλο AR
- τοπικό μοντέλο μέσου όρου, LAM

Πρόβλεψη ενός βήματος μπροστά
για την περίοδο 21.9.2005 – 12.10.2005

Αποδόσεις $y_t = \dfrac{x_t - x_{t-1}}{x_{t-1}}$

γενικού δείκτη

Γενικός δείκτης