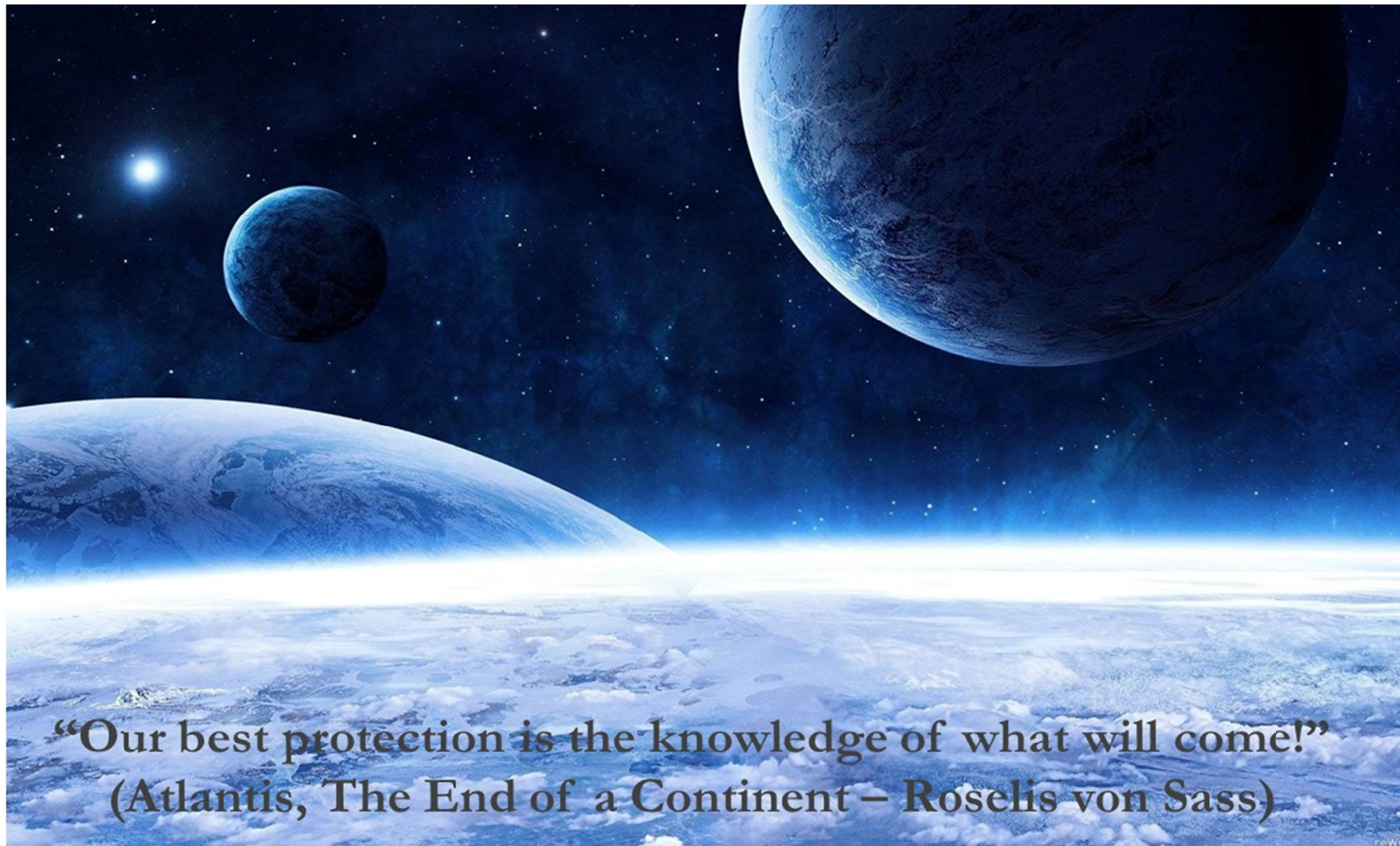


# RNAs *Radial Basis Function* no Contexto de Filtragem Preditiva Não-Linear

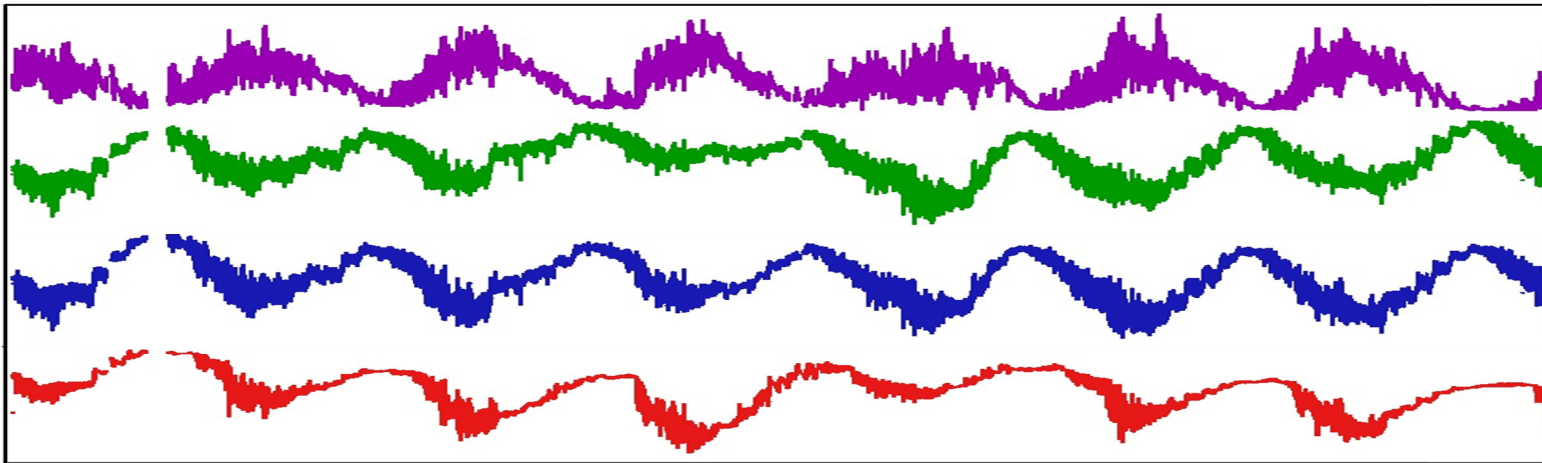


**“Our best protection is the knowledge of what will come!”  
(Atlantis, The End of a Continent – Roselis von Sass)**

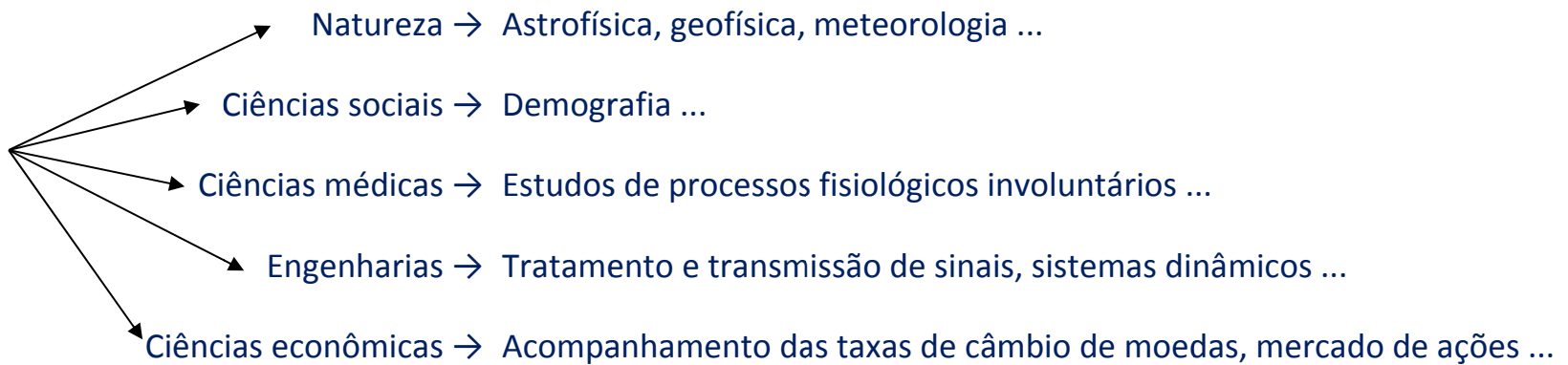
## Predição de Séries Temporais

Se conhecemos o comportamento de um sinal até um determinado ponto no passado do tempo, é possível fazer alguma inferência sobre seus valores futuros.

Tal processo de inferência é conhecido como predição.

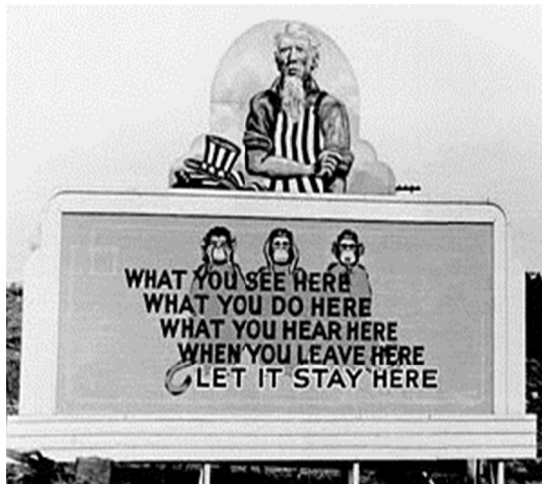
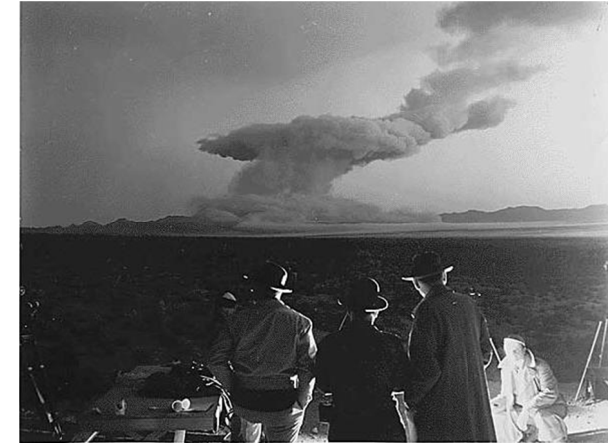
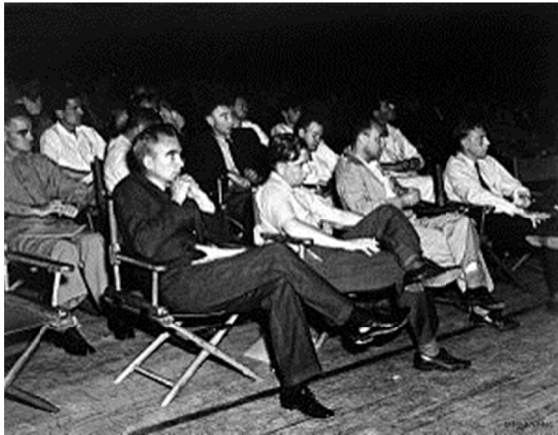


Séries  
Temporais





Projeto Manhattan em Los Alamos - New Mexico (1943) – a predição do *yield* da explosão nuclear era obtido da série temporal da onda sísmica (“A. S. Weigend and N. A. Gershenfeld. Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley Publishing Company, 1994”)



## Características de RNAs que as tornam adequadas à predição de séries temporais:

- RNAs supervisionadas têm a capacidade de aprender padrões estatísticos de ordem superior, subjacentes no conjunto de dados e, desta forma, "mimetizar" os processos estocásticos regentes das séries temporais.
- Por esta razão, quando as séries temporais a serem preditas resultam de processos desconhecidos, não-lineares e/ou não-estacionários, RNAs apresentam melhor desempenho que os métodos estatísticos de 2ª ordem tradicionais.
- Os tipos básicos de RNAs supervisionadas são as redes MLP (*Multilayer Perceptrons*) treinadas pelo algoritmo *back-propagation* e as redes RBF (*Radial Basis Function*).
- Como sabemos, ambas as redes MLPs e RBFs são aproximadoras universais. No entanto, quando se trata de aprendizado continuado, como no caso da predição de séries temporais, as redes MLPs se mostram menos adequadas porque o custo computacional de treino de uma rede MLP é muito superior ao de uma rede RBF, o que torna difícil a operação de forma dinâmica.
- RNAs RBF possuem características particulares que as capacitam a aprender rapidamente padrões complexos e tendências presentes nos dados e a se adaptar rapidamente a mudanças (especialmente adequadas à predição de séries temporais, especialmente aquelas séries regidas por processos não-lineares e/ou não-estacionários, casos em que as técnicas lineares de modelamento têm sucesso apenas limitado em seu desempenho).

## Predição Não-Linear de Séries Temporais através de RNAs *Radial Basis Functions*

A RNA RBF utilizada para predição não-linear de séries temporais é dita dinâmica, porque o aprendizado acontece de forma contínua com o desenrolar temporal da série.

Uma janela de  $N$  elementos é deslizada sobre a série temporal  $S$ , e a predição é feita a cada janela.

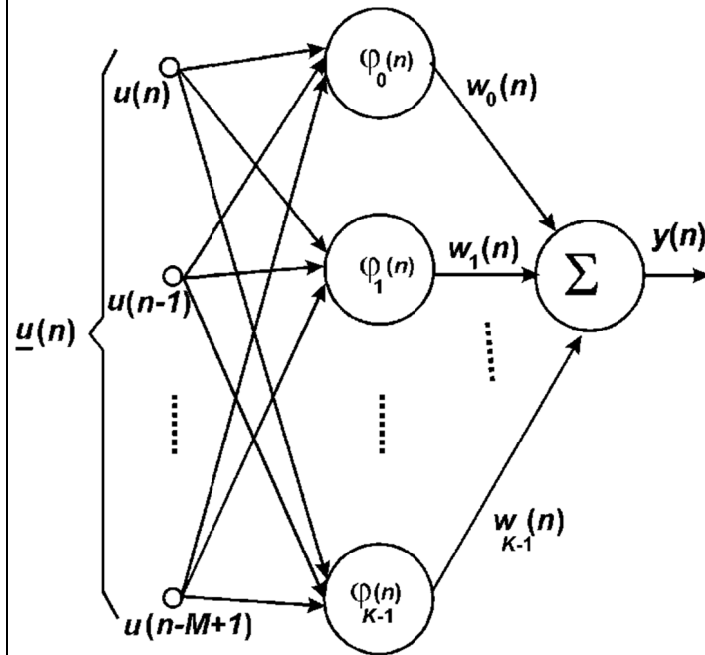
O desenrolar da série temporal é visto como um processo com matriz de transição de estados  $\Phi$ .

A matriz  $\Phi$  é representada pela RNA RBF, que é vista como um filtro não-linear com matriz de interpolação  $\Phi$ .

A matriz  $\Phi$  armazena informação sobre os estados básicos do processo a ser predito, com base no conjunto de vetores centro das funções de base radial (que são os vetores de estado do processo associado a  $S$ ), ou seja,

$\Phi = f(\underline{t}_k)$ , onde  $\underline{t}_k$ : estados básicos do processo a ser predito.

## RNA RBF utilizada na Predição Não-Linear de Séries Temporais



- $M$  (# nós de entrada) = ordem de predição;
- $K$  (# centros Gaussianos) = # de vetores de estado do processo;
- $\underline{t}_k \in \mathfrak{R}^M$  ( $k$ -ésimo vetor centro) =  $k$ -ésimo vetor de estado do processo associado à série temporal  $S$ ;
- $w_k$  =  $k$ -ésimo peso sináptico;
- $\sigma^2$  = variância comum a todos os centros Gaussianos;
- $k = 0, 1, \dots, K - 1$ .

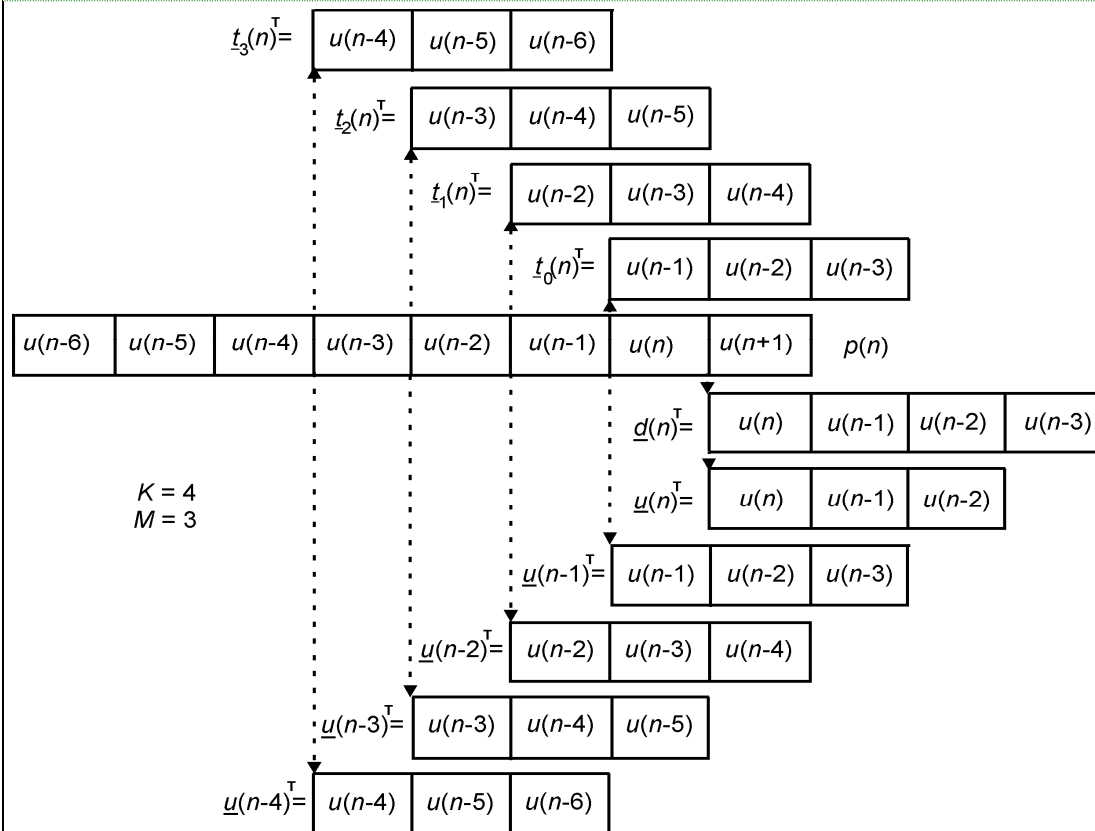
A saída da RBF quando o  $n$ -ésimo vetor  $\underline{u}(n) \in \mathfrak{R}^M$  é apresentado à sua entrada é dada por

$$y(n) = \sum_{k=0}^{K-1} w_k(n) \varphi_k(n) = \underline{\varphi}^T(n) \underline{w}(n) \quad (5.64);$$

onde  $\varphi_k(n) = \exp\left[-\frac{1}{2\sigma^2(n)} \|\underline{u}(n) - \underline{t}_k(n)\|^2\right]$  (5.65) é a saída do  $k$ -ésimo centro Gaussiano.

➤ Seja  $S = \{u(0), u(1), \dots, u(N_t - 1)\}$ , onde  $N_t$  é o número de amostras de  $S$ .

➤ A um instante  $n$  qualquer, o objetivo é prever a amostra  $u(n+1)$  de  $S$ , sendo conhecidas as  $N = K + M$  amostras prévias  $u(n), u(n-1), \dots, u(n-M-K+1)$  que compõem a janela de predição  $p(n) = \{u(n-M-K+1), \dots, u(n-1), u(n)\}$  definida sobre  $S$ .



➤ No instante  $n$ , sejam:

O processo associado ao desenrolar temporal de  $S$  representado pelo conjunto  $\mathbf{U}$  de  $K+1$  vetores  $u(n-\delta) \in \mathcal{R}^M$ ,  $\delta=0,1,\dots,K$ , definido sobre  $p(n)$ , de forma que dois vetores consecutivos de  $\mathbf{U}$  estejam deslocados entre si da distância temporal entre duas amostras subsequentes de  $S$

O vetor de entrada dado por (5.66):

$$\underline{u}(n) = [u(n) \ u(n-1) \ \dots \ u(n-M+1)]^T$$

O  $k$ -ésimo vetor de estado  $\underline{t}_k(n)$  do processo associado ao desenrolar temporal de  $S$ , dado por (5.67):

$$\underline{t}_k(n) = \underline{u}(n-k-1); k=0,1,\dots,K-1.$$

Para que se defina a variância  $\sigma^2(n)$  comum a todos os centros Gaussianos no instante  $n$ , assume-se que  $\sigma^2(n)$  seja proporcional ao quadrado da máxima distância Euclidiana entre todos os vetores de estado do processo, assim,

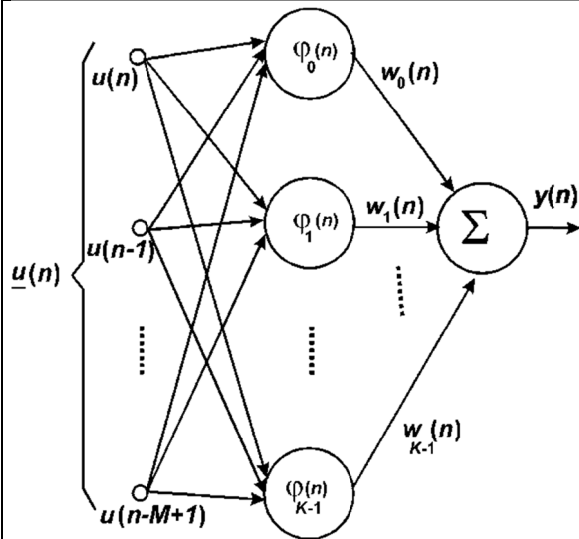
$$\sigma^2(n) = \xi \max \left\{ \left\| \underline{t}_i(n) - \underline{t}_j(n) \right\|^2 \right\}, \quad i, j = 0, 1, \dots, K-1, \quad (5.68)$$

onde  $\xi$  é a constante de proporção chamada Fator de Variância, a qual absorve a constante 2 da Equação (5.65).

Assim, a saída do  $k$ -ésimo centro Gaussiano, quando o vetor  $\underline{u}(n)$  é aplicado à entrada da rede RBF pode ser redefinida como

$$\varphi_k(n) = \exp \left\{ - \frac{\left\| \underline{u}(n) - \underline{t}_k(n) \right\|^2}{\xi \max \left\{ \left\| \underline{t}_i(n) - \underline{t}_j(n) \right\|^2 \right\}} \right\}, \quad i, j = 0, 1, \dots, K-1 \quad (5.69)$$

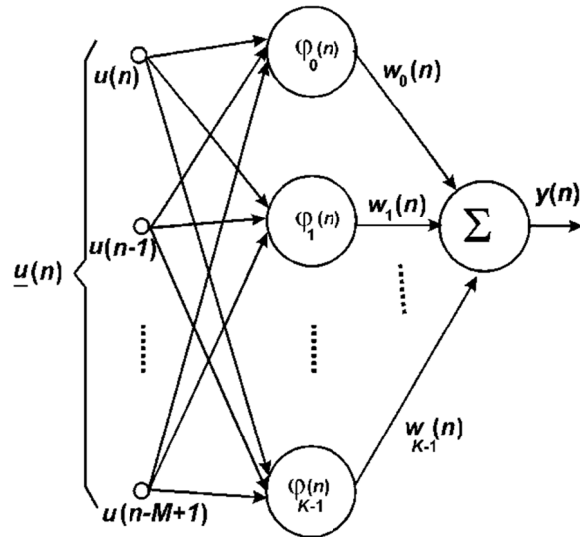
O conjunto  $\varphi_k(n)$ ,  $k = 0, 1, \dots, K-1$ , de saídas dos  $K$  centros Gaussianos, conjunto que resulta da aplicação do vetor de entrada  $\underline{u}(n)$ , pode ser colocado na forma vetorial através de  $\underline{\varphi}(n) = [\varphi_0(n) \ \varphi_1(n) \ \dots \ \varphi_{K-1}(n)]^T$  (5.70)



Por exemplo:

- $\underline{\varphi}(n-1) \in \mathfrak{R}^K$  é o vetor que resulta da aplicação do vetor  $\underline{u}(n-1) \in \mathfrak{R}^M$  à entrada da rede RBF;
- os elementos do vetor  $\underline{\varphi}(n-1) \in \mathfrak{R}^K$  são as saídas de cada centro Gaussiano ao vetor  $\underline{u}(n-1)$ ;
- o  $k$ -ésimo centro Gaussiano é definido por seu respectivo vetor  $\underline{t}_k \in \mathfrak{R}^M$  de estado do processo,  $k = 0, 1, \dots, K-1$ .





- Note que, a qualquer instante arbitrário, a transformação não-linear definida pela equação

$$\varphi_k(n) = \exp\left\{-\frac{\|\underline{u}(n) - \underline{t}_k(n)\|^2}{\xi \max\{\|\underline{t}_i(n) - \underline{t}_j(n)\|^2\}}\right\}, \quad i, j = 0, 1, \dots, K-1 \quad (5.69)$$

mapeia o vetor  $\underline{u}(n-\delta) \in \mathfrak{R}^M$  aplicado à entrada da rede RBF,  $\delta$  é um atraso arbitrário qualquer, no vetor  $\underline{\varphi}(n-\delta) \in \mathfrak{R}^K$ . Este mapeamento não-linear estabelece as condições analíticas para a representação através de estatísticas de ordem superior do processo estocástico que rege o desenrolar temporal da série.

- Portanto, a sequência de vetores de entrada  $\underline{u}(n-1), \underline{u}(n-2), \dots, \underline{u}(n-K)$  define a sequência de vetores  $\underline{\varphi}(n-1), \underline{\varphi}(n-2), \dots, \underline{\varphi}(n-K)$  obtidos pelo mapeamento não-linear  $\mathfrak{R}^M \rightarrow \mathfrak{R}^K$  definido pela Equação (5.69).

Note que, apesar de definidos em uma dimensão diferente da dimensão original dos vetores  $\underline{t}_k \in \mathfrak{R}^M$  de estados da série  $S$ , e, apesar de obtidos através de uma transformação não-linear entre as dimensões  $\mathfrak{R}^M$  e  $\mathfrak{R}^K$ , **o conjunto de vetores  $\underline{\varphi}(n-1), \underline{\varphi}(n-2), \dots, \underline{\varphi}(n-K)$  definidos em  $\mathfrak{R}^K$  permanece representando a informação sobre o desenrolar temporal da série.**

Portanto, os vetores em  $\mathfrak{R}^K$  contêm informação implícita sobre os estados de  $S$ , definidos agora em uma outra dimensão, com  $S$  sendo “vista” através de um processo não-linear.

Desta forma, o conjunto de vetores em  $\mathfrak{R}^K$  pode ser agrupado em uma matriz de transição de estados da série temporal  $S$ , agora interpretada como um processo não-linear com  $K$  estados  $\mathfrak{R}^K$  dimensionais.

A matriz de transição de estados da série temporal  $S$ , tomada como um processo não-linear com  $K$  estados  $\mathfrak{R}^K$  dimensionais no instante  $n$  é mostrada na Equação (5.71).

$$\Phi(n) = \begin{bmatrix} \underline{\varphi}(n-1)^T \\ \underline{\varphi}(n-2)^T \\ \vdots \\ \underline{\varphi}(n-K)^T \end{bmatrix} = \begin{bmatrix} \varphi_0(n-1) & \varphi_1(n-1) & \cdots & \varphi_{K-1}(n-1) \\ \varphi_0(n-2) & \varphi_1(n-2) & \cdots & \varphi_{K-1}(n-2) \\ \vdots & \vdots & \cdots & \vdots \\ \varphi_0(n-K) & \varphi_1(n-K) & \cdots & \varphi_{K-1}(n-K) \end{bmatrix} \quad (5.71)$$

Note que as linhas de  $\Phi(n)$  correspondem aos vetores de transição de estado  $\underline{\varphi}(n-1)^T, \underline{\varphi}(n-2)^T, \dots, \underline{\varphi}(n-K)^T$  do processo não-linear, os quais resultam respectivamente da aplicação dos vetores  $\underline{u}(n-1), \underline{u}(n-2), \dots, \underline{u}(n-K)$  à entrada da rede RBF.

Note também, da transformação (5.69), que o  $k$ -ésimo componente do vetor  $\underline{\varphi}(n-\delta) \in \mathfrak{R}^K$ ,  $\delta$  arbitrário, **tende para o valor máximo 1.0 à medida em que o vetor  $\underline{u}(n-\delta) \in \mathfrak{R}^M$  aplicado à entrada da RBF tende para o vetor de estado  $\underline{t}_k$  da  $k$ -ésima função de base radial.**

$$\varphi_k(n) = \exp \left\{ - \frac{\| \underline{u}(n) - \underline{t}_k(n) \|^2}{\xi \max \left\{ \| \underline{t}_i(n) - \underline{t}_j(n) \|^2 \right\}} \right\}; \quad i, j = 0, 1, \dots, K-1 \quad (5.69)$$

Portanto, (5.69) é uma transformação não-linear de  $\mathfrak{R}^M \rightarrow \mathfrak{R}^K$  que **mede o quanto o desenrolar temporal momentâneo da série  $S$  se relaciona com os  $K$  estados básicos do processo a ela associado.**

## Determinação dos pesos sinápticos $\underline{w}(n) \in \mathfrak{R}^K$ no instante $n$

Para um vetor de pesos sinápticos arbitrário  $\underline{w} = \underline{w}_a$ , o conjunto de saídas ou o vetor de saídas  $\underline{y}(n)$  para  $\Phi(n)$  dado é obtido por

$$\underline{y}(n) = \Phi(n) \underline{w}_a \quad \text{com} \quad \underline{y}(n) = [y(n-1) \ y(n-2) \ \dots \ y(n-K)]^T$$

(5.72) (5.73)

onde  $y(n-1), y(n-2), \dots, y(n-K)$  são as saídas da rede RBF com respeito aos vetores de entrada  $\underline{u}(n-1), \underline{u}(n-2), \dots, \underline{u}(n-K)$  associados ao desenrolar temporal momentâneo da série  $S$ , sendo dados o vetor de pesos sinápticos arbitrário  $\underline{w} = \underline{w}_a$  e a matriz de estados  $\Phi(n)$ .

Vamos supor que  $\underline{y}(n) = \underline{d}(n)$ , onde  $\underline{d}(n)$  é o vetor de saídas desejadas definido por  $\underline{d}(n) = [u(n) \ u(n-1) \ \dots \ u(n-K+1)]^T$ , de tal forma que

$$\underline{y}(n) = \underline{d}(n) = \begin{bmatrix} u(n) \\ u(n-1) \\ \vdots \\ u(n-K+1) \end{bmatrix} = \Phi(n) \cdot \underline{w}(n) = \begin{bmatrix} \varphi_0(n-1) & \varphi_1(n-1) & \dots & \varphi_{K-1}(n-1) \\ \varphi_0(n-2) & \varphi_1(n-2) & \dots & \varphi_{K-1}(n-2) \\ \vdots & \vdots & & \vdots \\ \varphi_0(n-K) & \varphi_1(n-K) & \dots & \varphi_{K-1}(n-K) \end{bmatrix} \begin{bmatrix} w_0(n) \\ w_1(n) \\ \vdots \\ w_{K-1}(n) \end{bmatrix} \quad (5.74)$$

Observe que cada elemento em  $\underline{d}(n)$  é o elemento que se coloca uma posição à frente na série  $S$ , com respeito ao vetor de entrada que gerou o correspondente vetor de transição de estado não-linear em  $\Phi$ .

Por exemplo, a primeira linha de  $\Phi$  na Equação (5.74) é o vetor de transição de estado não-linear  $\underline{\varphi}(n-1)^T$  que resulta da aplicação do vetor  $\underline{u}(n-1)$  à entrada da rede neural.

A saída desejada correspondente a este vetor de entrada é o elemento  $u(n)$  de  $\underline{d}(n)$ , que se encontra uma posição à frente na série  $S$ , com respeito ao vetor  $\underline{u}(n-1)$ , conforme pode ser observado na figura do slide 7.

Portanto, através da transformação linear definida pelo vetor de pesos sinápticos  $\underline{w}(n) \in \mathfrak{R}^K$  e através da transformação não-linear definida pela matriz  $\Phi(n)$ , a Equação (5.74) implicitamente relaciona cada vetor  $\underline{u}(n-k-1)$  formado da janela  $p(n)$ ,  $k=0,1,\dots,K-1$ , com o elemento  $u(n-k)$  de  $S$ , sendo  $u(n-k)$  o elemento que está localizado na série  $S$  uma posição à frente do vetor  $\underline{u}(n-k-1)$ .

$$\underline{y}(n) = \underline{d}(n) = \begin{bmatrix} u(n) \\ u(n-1) \\ \vdots \\ u(n-K+1) \end{bmatrix} = \Phi(n) \cdot \underline{w}(n) = \begin{bmatrix} \varphi_0(n-1) & \varphi_1(n-1) & \cdots & \varphi_{K-1}(n-1) \\ \varphi_0(n-2) & \varphi_1(n-2) & \cdots & \varphi_{K-1}(n-2) \\ \vdots & \vdots & & \vdots \\ \varphi_0(n-K) & \varphi_1(n-K) & \cdots & \varphi_{K-1}(n-K) \end{bmatrix} \begin{bmatrix} w_0(n) \\ w_1(n) \\ \vdots \\ w_{K-1}(n) \end{bmatrix}$$

Assim, uma vez determinado,  $\underline{w}(n)$  conterá informação de como se efetua a transição partindo dos estados prévios do processo de  $S$  até o próximo elemento de  $S$  imediatamente adiante ao respectivo vetor  $\underline{u}(n-k-1)$  representativo do desenrolar temporal momentâneo de  $S$ .

É importante reiterar que a informação de transição em  $\underline{w}(n)$  é resultante de uma transformação não-linear  $\mathfrak{R}^M \rightarrow \mathfrak{R}^K$ , e, em consequência disto, é uma informação que envolve as estatísticas de ordem superior do processo de  $S$ . Em função disto, o método de predição não-linear aqui apresentado é potencialmente mais capaz de captar as “sutilezas estatísticas” do processo estocástico subjacente em  $S$  do que o método de predição linear.

A obtenção de  $\underline{w}(n)$  é definida por  $\underline{w}(n) = \Phi^{-1}(n) \underline{d}(n)$  (5.75)

**Observação com relação à equação de obtenção do vetor de pesos,  $\underline{w}(n) = \Phi^{-1}(n) \underline{d}(n)$ :** A inversa da matriz  $\Phi$  é obtida pela pseudo-inversão matricial de Moore-Penrose, através de decomposição em valores singulares – SVD.

Embora a SVD minimize o problema da eventual singularidade de  $\Phi$  e, embora toda a série temporal seja normalizada para o intervalo  $[-1,1]$ , antes de qualquer procedimento, por precaução, adiciona-se o valor  $1 \times 10^{-9}$  à diagonal principal de  $\Phi$ , como um parâmetro de regularização, visando auxiliar o tratamento das singularidades da matriz  $\Phi$ .

## Processo de Predição

Uma vez obtido  $\underline{w}(n)$  de (5.75), aplica-se o vetor  $\underline{u}(n)=[u(n), u(n-1), \dots, u(n-M+1)]^T$  à entrada da rede neural.

O vetor peso sináptico  $\underline{w}(n)$  obtido de (5.75) armazena informação de como ocorre a transição “[estados prévios → próximo elemento](#)” dado o vetor de entrada que descreve o desenrolar temporal momentâneo da série  $S$ .

Como, por definição, uma variável de estado não sofre alteração para uma variação pequena no sistema por ela descrito, assume-se que os vetores de estado  $\underline{t}_k$  do processo de  $S$  não sofram uma mudança significativa uma posição à frente em  $S$ .

Assim, a saída da rede neural  $y(n)$  ao vetor de entrada  $\underline{u}(n)=[u(n) \ u(n-1) \ \dots \ u(n-M+1)]^T$  será uma estimativa  $\hat{u}(n+1)$  ou predição da amostra  $u(n+1)$ , dada pela Equação (5.76) com base em (5.64):

$$\hat{u}(n+1) = y(n) = \sum_{k=0}^{K-1} w_k(n) \exp \left\{ \frac{-\|\underline{u}(n) - \underline{t}_k(n)\|^2}{\xi \max \left\{ \|\underline{t}_i(n) - \underline{t}_j(n)\|^2 \right\}} \right\} = \sum_{k=0}^{K-1} w_k(n) \varphi_k(n) = \underline{w}^T(n) \underline{\varphi}(n) \quad (5.76)$$

Em outras palavras, se está implicitamente deslizando a matriz  $\Phi$  uma posição à frente ao longo de  $S$ , assumindo que os estados armazenados em  $\Phi$  permanecem inalterados e usando-se a informação de transição armazenada em  $\underline{w}$  para estimar o próximo elemento em  $S$ , a partir dos estados definidos em  $\Phi$ .

Obviamente,  $K$  deve ser grande o suficiente para que  $\Phi$  possa armazenar todos os estados significativos. Da mesma forma, a dimensão  $M$  dos vetores de estado do processo de  $S$  deve ser suficientemente grande para representar os estados significativos do processo.

A técnica de atribuição de centros definida pela Equação (5.67) para efeito de predição não-linear de séries temporais através de RNAs RBF é denominada [Atribuição Padrão dos Centros \(APC\)](#).



## Atribuição Padrão dos Centros (APC) – detalhamento

- Os neurônios da camada escondida da RBF são um conjunto de funções que constitui uma base arbitrária na qual o conjunto de entrada pode ser expandido (os dados representados através de redes neurais RBF são expandidos com referência a um conjunto finito de funções de ativação neurais, chamadas funções de base radial).
  - Cada uma destas funções é centrada em uma particular coordenada do espaço multi-dimensional dos pontos que compõem o espaço de dados de entrada. Esta coordenada define o centro  $\underline{t}_k$  de uma região de maior aglomeração de dados.
- 
- Em se tratando de predição de séries temporais, a cada instante  $n$ , uma janela de  $N$  elementos é definida sobre  $S$  e a predição é feita sobre esta janela  $\Rightarrow$  um processo dinâmico.
  - A janela é particionada em vetores  $\underline{u}(n)$  de entrada da RBF e em vetores centro  $\underline{t}_k(n)$  daquele conjunto de dados de entrada que compõe a janela de predição.
  - Os vetores  $\underline{t}_k(n)$ , portanto, se deslocam ao longo do desenrolar temporal de  $S$  (a cada nova janela, a cada instante  $n$ ) com respeito ao conjunto de dados global que compõe a série  $S$ . Desta forma, os vetores  $\underline{t}_k(n)$  caracterizam os estados do processo associado ao desenrolar temporal de  $S$ .

- A cada estado armazenado em  $\Phi$  é associada uma saída desejada da rede RBF, de tal forma a se definir um vetor de saídas desejadas  $\underline{d}$ .
- Cada elemento em  $\underline{d}$  é definido pelo elemento que está uma posição à frente na série temporal com respeito ao vetor de entrada que gerou o correspondente vetor de estados em  $\Phi$ .
- $\underline{d} = \Phi \cdot \underline{w}$ , onde  $\underline{w}$  é o vetor de pesos sinápticos da rede RBF.

- $\underline{w}$  define a transição entre os estados prévios do processo associado à série e o próximo e imediato elemento da mesma, pois:

⇒ a matriz  $\Phi$  armazena informação sobre os estados básicos do processo a ser predito e

⇒  $\underline{d}$  é o elemento que está uma posição à frente na série temporal com respeito ao vetor de entrada que gerou o correspondente vetor de estados em  $\Phi$ .

- $\underline{w}$  armazena informação sobre o modo como o próximo elemento na série é gerado a partir de seus estados prévios.

- Por inversão matricial, caso  $\Phi$  não seja singular, pode-se determinar  $\underline{w} = \Phi^{-1} \cdot \underline{d}$ .

- Deslizando a matriz  $\Phi$  uma posição à frente na janela e usando a informação de transição contida em  $\underline{w}$  pode-se estimar o próximo elemento na série.

**APC - Sequência de Operações**  
**p/ K = 4 e M = 3.**

$\mathfrak{F}: \mathfrak{R}^M \Rightarrow \mathfrak{R}^K \mid \underline{\varphi}(n) \in \mathfrak{R}^K = \mathfrak{F}\{\underline{u}(n) \in \mathfrak{R}^M\}$

$\underline{\varphi} = [\varphi_i], i = 0, 1, 2, \dots, K-1$

$\varphi_i(n) = \exp(-\|\underline{u}(n) - \underline{t}_i(n)\|^2 / 2\sigma^2(n))$

$\underline{t}_i(n) = \underline{u}(n-i-1), \underline{t}_i(n) \in \mathfrak{R}^M$   
 $i = 0, 1, 2, \dots, K-1$

$\underline{t}_3(n)^T = [u(n-4) \ u(n-5) \ u(n-6)]$   
 $\underline{t}_2(n)^T = [u(n-3) \ u(n-4) \ u(n-5)]$   
 $\underline{t}_1(n)^T = [u(n-2) \ u(n-3) \ u(n-4)]$   
 $\underline{t}_0(n)^T = [u(n-1) \ u(n-2) \ u(n-3)]$

$p(n)$   
 $N = K+M$

$[u(n-6) \ u(n-5) \ u(n-4) \ u(n-3) \ u(n-2) \ u(n-1) \ u(n) \ u(n+1)]$

$\underline{u}(n-j) \in \mathfrak{R}^M$   
 $j = 1, 2, \dots, K$

$\underline{u}(n-1)^T = [u(n-1) \ u(n-2) \ u(n-3)]$   
 $\underline{u}(n-2)^T = [u(n-2) \ u(n-3) \ u(n-4)]$   
 $\underline{u}(n-3)^T = [u(n-3) \ u(n-4) \ u(n-5)]$   
 $\underline{u}(n-4)^T = [u(n-4) \ u(n-5) \ u(n-6)]$

$\varphi_i(n) = \exp(-\|\underline{u}(n) - \underline{t}_i(n)\|^2 / 2\sigma^2(n))$

$\mathfrak{F}\{\cdot\}$

$$\begin{bmatrix} \underline{\varphi}(n-1)^T \\ \underline{\varphi}(n-2)^T \\ \underline{\varphi}(n-3)^T \\ \underline{\varphi}(n-4)^T \end{bmatrix} = \begin{bmatrix} \varphi_0(n-1) & \varphi_1(n-1) & \varphi_2(n-1) & \varphi_3(n-1) \\ \varphi_0(n-2) & \varphi_1(n-2) & \varphi_2(n-2) & \varphi_3(n-2) \\ \varphi_0(n-3) & \varphi_1(n-3) & \varphi_2(n-3) & \varphi_3(n-3) \\ \varphi_0(n-4) & \varphi_1(n-4) & \varphi_2(n-4) & \varphi_3(n-4) \end{bmatrix} =: \Phi_{K \times K}$$

$p(n)$

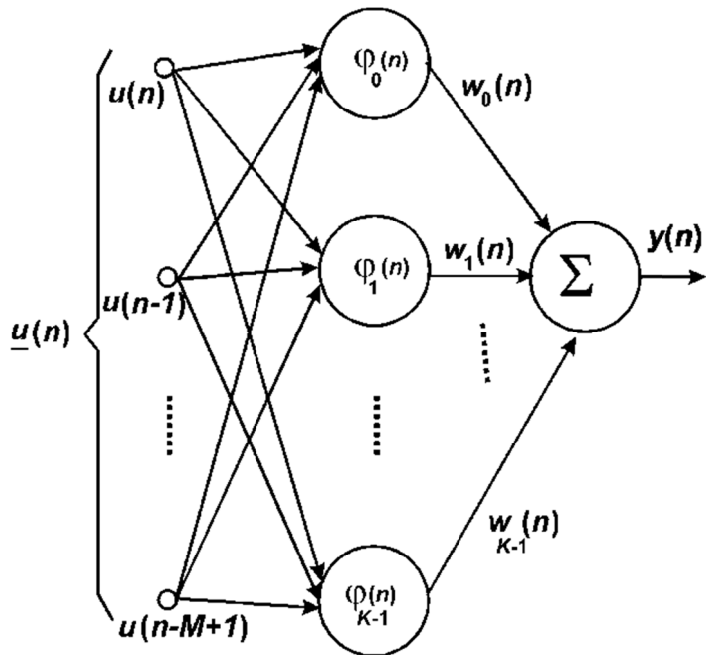
$[u(n-6) \ u(n-5) \ u(n-4) \ u(n-3) \ u(n-2) \ u(n-1) \ u(n) \ u(n+1)]$

$\underline{d}(n)^T = [u(n) \ u(n-1) \ u(n-2) \ u(n-3)]$

$$\underline{d} = \begin{bmatrix} u(n) \\ u(n-1) \\ u(n-2) \\ u(n-3) \end{bmatrix} = \Phi \underline{w} \Rightarrow \underline{w} = \Phi^+ \underline{d}$$

$$\hat{u}(n+1) = \underline{w}^T \underline{\varphi}(n), \quad \underline{\varphi}(n) = \mathfrak{F}\{\underline{u}(n)\}$$
  

$$\underline{u}(n)^T = [u(n) \ u(n-1) \ u(n-2)]$$



## Sumário da Heurística APC para Predição Não-Linear de Séries Temporais via RBFs

<b>I - Inicialização:</b>	
1.	Posicionar $p(n=0)$ de $N$ elementos, sobre $S$ .
2.	Definir $K$ e $M$ , ( $N = K + M$ ) a partir de $p(n)$ .
3.	Formar a partir de $p(n)$ os vetores-centro $\underline{t}_k(n) \in \mathfrak{R}^M$ das funções de base radial.
4.	Formar a partir de $p(n)$ os vetores $\underline{u}(n-\delta) \in \mathfrak{R}^M$ , $\delta = 0, 1, \dots, K$ do conjunto de treino.

## II - Treinamento:

1. Apresentar todos os vetores do conjunto de treino à entrada da rede RBF (exceto o vetor que contém as informações mais recentes sobre  $S$ , o vetor  $\underline{u}(n)$ ), obtendo para cada um dos demais vetores  $\underline{u}(n-\delta) \in \mathfrak{R}^M, \delta=1, \dots, K$  o correspondente vetor  $\underline{\varphi}(n-\delta) \in \mathfrak{R}^K, \delta=1, \dots, K$ .
2. Construir  $\Phi(n)$  a partir dos vetores  $\underline{\varphi}(n-\delta)$ .
3. Formar vetor saída desejada  $\underline{d}(n) = [u(n) \ u(n-1) \ \dots \ u(n-K+1)]$ , atribuindo a cada vetor  $\underline{\varphi}(n-\delta)$  uma saída desejada equivalente ao último componente do vetor  $\underline{u}(n-\delta+1)$ .

4. Determinar vetor de pesos sinápticos  $\underline{w}(n) = \Phi^{-1}(n)\underline{d}(n)$ .

5. Aplicar vetor  $\underline{u}(n)$  à entrada da RBF, obtendo na saída

$$\hat{u}(n+1) = \sum_{k=0}^{K-1} w_k(n) \exp \left[ - \frac{\|\underline{u}(n) - \underline{t}_k(n)\|^2}{\xi(n) \max_{i,j} \|\underline{t}_i(n) - \underline{t}_j(n)\|^2} \right] = \underline{w}^T(n) \underline{\varphi}(n)$$

6.  $n = n + 1$  (deslizando  $p(n)$  uma posição à frente em  $S$ ).

7. Obtida a nova janela, voltar ao passo 3 da etapa de inicialização.



## Predição Não-Linear de Séries Temporais através da Heurística APC: Exemplo

A Figura 5.8 apresenta o resultado da predição não-linear baseada na heurística APC obtido para a série *Chaotic\_LASER* (ver descrição desta série no Apêndice A).

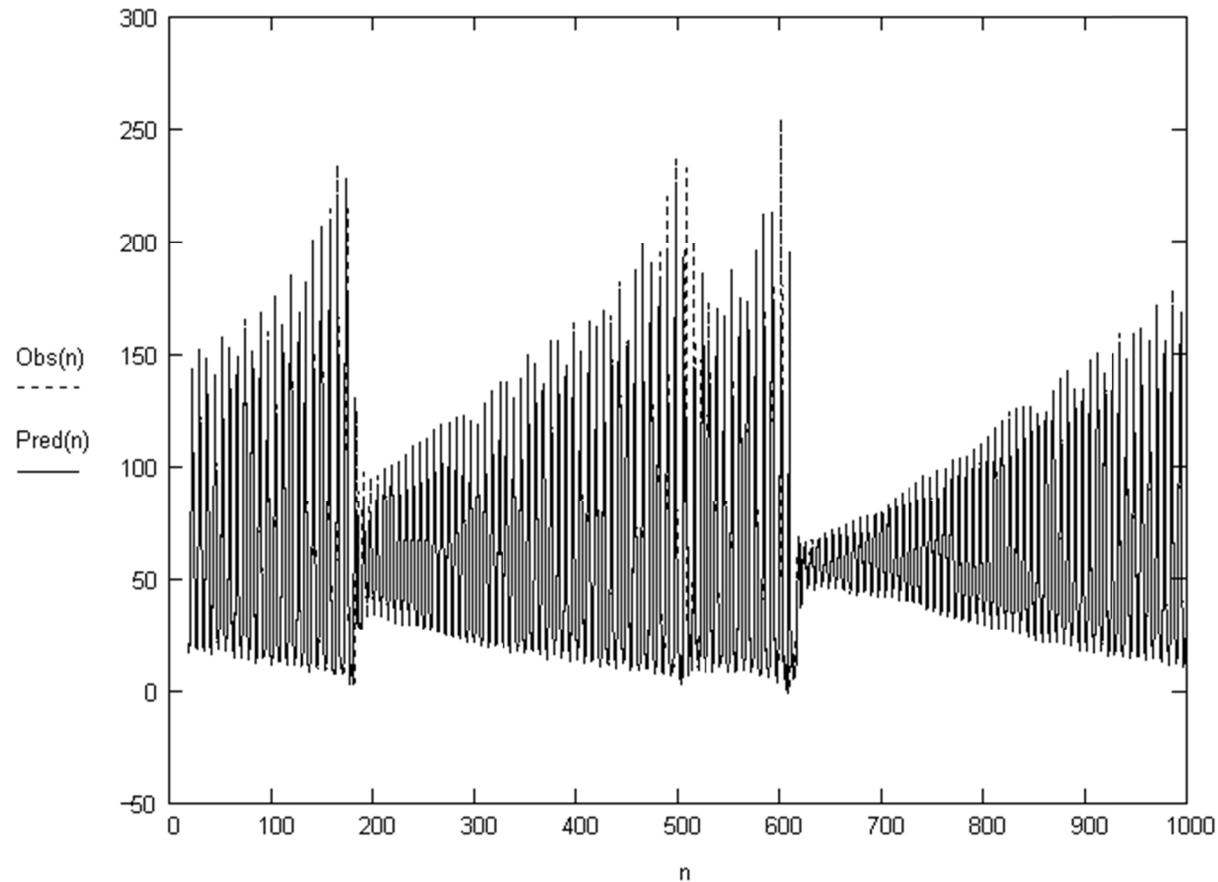


Figura 5.8: Série *Chaotic\_LASER* – Observada e Predita.

Predição não-linear através de RBFs com heurística APC.

$$M = 11 \quad K = 8 \quad N = K + M = 19 \quad \xi = 1.0 \quad \text{NMSE}(N_t - 1) = 0.096$$

## Predição Linear da série temporal Chaotic Laser ( $M = 22$ ), conforme Apêndice A

• Experimentou-se a predição linear da série *Chaotic\_LASER* (conforme Apêndice A), considerando, neste novo experimento, o dobro da ordem de predição utilizada na predição por APC, i. é,  $M = 22$  para o caso linear.

• Os valores dos coeficientes obtidos para o filtro linear, através de (5.43) são:

$W_0 = -0.670762$ ,  $W_1 = 0.545837$ ,  $W_2 = -0.253092$ ,  $W_3 = 0.198692$ ,  $W_4 = -0.0704754$ ,  $W_5 = 0.129263$ ,  $W_6 = -0.355902$ ,  $W_7 = -0.510382$ ,  
 $W_8 = 0.277562$ ,  $W_9 = -0.30334$ ,  $W_{10} = 0.217955$ ,  $W_{11} = -0.161362$ ,  $W_{12} = 0.092362$ ,  $W_{13} = -0.189575$ ,  $W_{14} = 0.0408657$ ,  $W_{15} = 0.162011$ ,  
 $W_{16} = -0.109375$ ,  $W_{17} = 0.0276988$ ,  $W_{18} = -0.0969306$ ,  $W_{19} = 0.0260158$ ,  $W_{20} = -0.0871898$  e  $W_{21} = 0.101258$ .

• O resultado da predição linear para este caso é mostrado na Figura 5.9 e o NMSE final resultante, conforme Equação (5.63), é  $\text{NMSE}(N_t - 1) = 0.185$ .

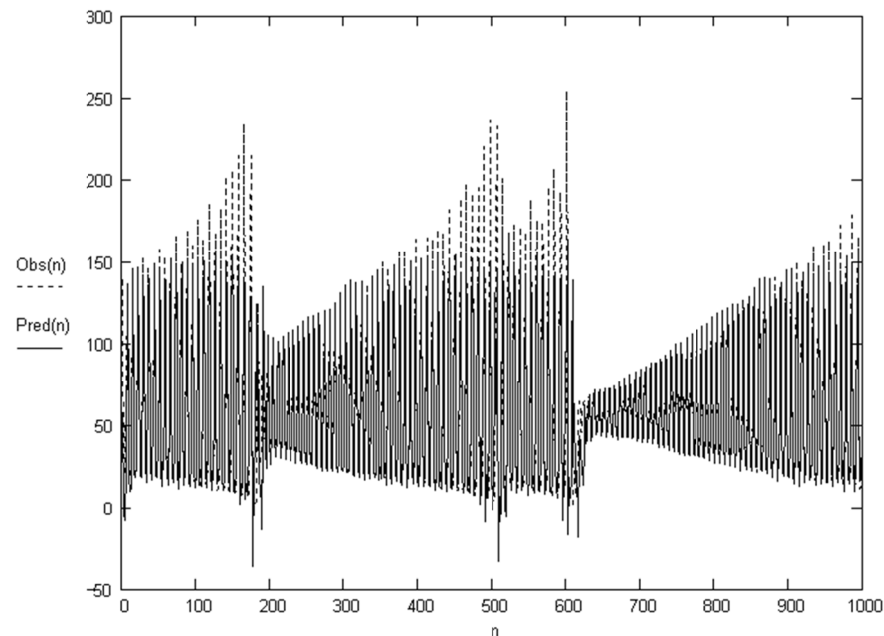


Figura 5.9: Série *Chaotic\_LASER* – Predição Linear –  $M = 22$ .  $\text{NMSE}(N_t - 1) = 0.185$ .

Heurística de Predição	Ordem de Predição	NMSE( $N_t - 1$ )	# Amostras Prévias Conhecidas
APC	11	0.096	19
Predição Linear	11	0.213	1000
Predição Linear	22	0.185	1000

A diferença de performance entre a predição linear e a heurística não-linear APC fica ainda mais evidente se lembrarmos que a heurística APC necessita, neste caso, para efetuar a predição, de apenas  $N = 19$  amostras, contidas na janela de predição  $p(n)$ .

Já a predição linear precisa conhecer todas as  $N_t = 1000$  amostras da série *Chaotic\_LASER* para montar a matriz de correlação  $\mathbf{R}$ .

## Apêndice A - Predição Linear de Séries Temporais

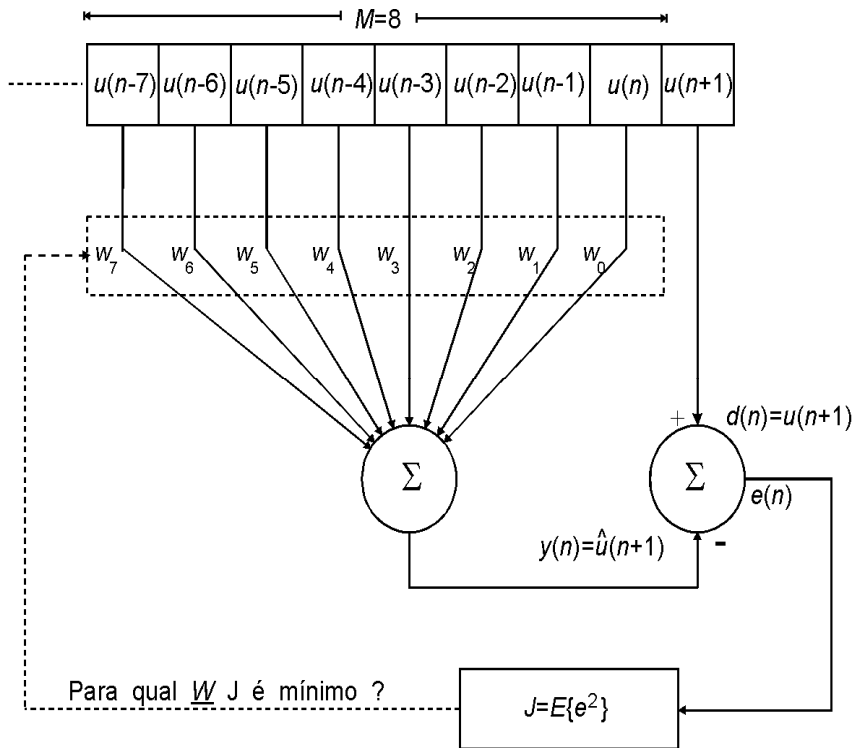


Figura 5.3: Preditor FIR de ordem  $M$ , mostrado no instante  $n$ .

Predição a um passo: predição da amostra  $u(n+1)$ , subsequente a um conjunto conhecido de amostras consecutivas prévias  $\{u(n), u(n-1), \dots\}$  pertencentes a uma série temporal discreta ( $S$ ).

$\hat{u}(n+1)$ , é expressa como uma combinação linear de  $M$  amostras prévias  $\{u(n), u(n-1), \dots, u(n-M+1)\}$ .

$$\hat{u}(n+1) = y(n) = \sum_{k=0}^{M-1} W_k u(n-k) = \underline{W}^T \underline{u} \quad (5.43)$$

Os coeficientes  $W_k, k = 0, 1, \dots, M-1$  que ponderam (5.43) definem um filtro FIR transversal.

O vetor  $\underline{W}$  que define o filtro FIR tem seus coeficientes determinados de forma a minimizar a função de custo  $J$ .

Um preditor linear de ordem  $M$  utiliza  $M$  amostras prévias conhecidas de  $S$  para estimar  $u(n+1)$ , no entanto, necessita do conhecimento de todas as amostras que compõem  $S$  para emular a matriz de correlação associada.

**O operador gradiente é aplicado com o intuito de obter o vetor de pesos  $\underline{W}$  (pesos  $W_k$ ) do filtro transversal que minimize a função de custo  $J$ .**

Como  $J = E\{e^2\}$  é uma função quadrática,  $J$  será globalmente mínimo para  $\nabla_i J(n) = 0$ . Assim:

$$\nabla_i J = \frac{\partial J}{\partial W_i} = \frac{\partial}{\partial W_i} E\{e^2\} = E\left\{2e \frac{\partial e}{\partial W_i}\right\} = E\left\{2e \frac{\partial}{\partial W_i} (d - y)\right\} = E\left\{-2e \frac{\partial y}{\partial W_i}\right\}; \quad \begin{matrix} i=0,1,\dots, \\ \dots M-1 \end{matrix}$$

$$\nabla_i J(n) = E\left\{-2e(n) \frac{\partial y(n)}{\partial W_i}\right\} = E\left\{-2e(n) \frac{\partial}{\partial W_i} \sum_{k=0}^{M-1} W_k u(n-k)\right\} = -2E\{e(n)u(n-i)\}$$

Fazendo  $\nabla_i J(n) = 0$ , substituindo  $e(n) = d(n) - y(n)$  e  $y(n)$  na equação acima, temos

$$E\left\{\left[d(n) - \sum_{k=0}^{M-1} W_k u(n-k)\right] u(n-i)\right\} = 0$$

Distribuindo os produtos e rearranjando a equação, temos

$$\sum_{k=0}^{M-1} W_k E\{u(n-k)u(n-i)\} = E\{d(n)u(n-i)\}, \text{ de onde obtemos}$$



$$\sum_{k=0}^{M-1} W_k E\{u(n-k)u(n-i)\} = E\{d(n)u(n-i)\} \rightarrow \sum_{k=0}^{M-1} W_k R_{uu}(k-i) = R_{du}(-i) \quad \text{onde:}$$

$R_{uu}$  é a função de auto-correlação do processo aleatório  $u$  para um atraso  $k-i$  entre as amostras, com  $k, i = 0, 1, \dots, M-1$ , e

$R_{du}$  é a função de correlação cruzada entre o processo aleatório que descreve a saída desejada  $d = u(n+1)$  e o processo  $u$ .

Se  $\underline{u}(n) = [u(n) \ u(n-1) \ \dots \ u(n-M+1)]^T$ , podemos escrever

$$\underbrace{E\{\underline{u}(n) \underline{u}^T(n)\}}_{\mathbf{R}} \underline{W} = \underbrace{E\{d(n)\underline{u}(n)\}}_{\underline{P}} \quad \mathbf{R}\underline{W} = \underline{P} \rightarrow \underline{W} = \mathbf{R}^{-1}\underline{P}$$

↑  
Equação de Wiener-Hopf

- A solução desta equação para  $\underline{W}$  define os coeficientes do filtro linear transversal da Figura 5.3 slide 23.
- O filtro prediz com o mínimo erro quadrático médio a amostra  $u(n+1)$  de uma série temporal que apresenta correlação entre as  $M$  prévias amostras.
- Se a matriz de correlação  $\mathbf{R}$  da série temporal é não-singular para  $M$  definido, então  $\underline{W}$  pode ser obtido por  $\underline{W} = \mathbf{R}^{-1}\underline{P}$  sendo  $\underline{P}$  o vetor que define a correlação cruzada entre o vetor  $\underline{u}(n)$  de entrada e a saída desejada  $d(n) = u(n+1)$ .

Vejamos como são expressos analiticamente a matriz  $\mathbf{R}$  e o vetor  $\underline{P}$  envolvidos no cômputo do vetor de coeficientes do filtro linear transversal através da Equação de Wiener-Hopf ( $\underline{W} = \mathbf{R}^{-1} \underline{P}$ ):

→ A matriz  $\mathbf{R} = E\{\underline{u}(n)\underline{u}^T(n)\}$  é obtida por:

$$\mathbf{R} = \begin{bmatrix} E\{u(n)u(n)\} & E\{u(n)u(n-1)\} & \cdots & E\{u(n)u(n-M+1)\} \\ E\{u(n-1)u(n)\} & E\{u(n-1)u(n-1)\} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ E\{u(n-M+1)u(n)\} & E\{u(n-M+1)u(n-1)\} & \cdots & E\{u(n-M+1)u(n-M+1)\} \end{bmatrix} \text{ ou}$$

$$\mathbf{R} = \begin{bmatrix} R_{uu}(0) & R_{uu}(1) & \cdots & R_{uu}(M-1) \\ R_{uu}(1) & R_{uu}(0) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ R_{uu}(M-1) & R_{uu}(M-2) & \cdots & R_{uu}(0) \end{bmatrix}$$

**Exemplo:**  $M=3$ ,  $\underline{u}(n) = [u(n) \ u(n-1) \ u(n-2)]^T$

$$\mathbf{R} = E \left\{ \begin{bmatrix} u(n) \\ u(n-1) \\ u(n-2) \end{bmatrix} \begin{bmatrix} u(n) & u(n-1) & u(n-2) \end{bmatrix} \right\} = \begin{bmatrix} R_{uu}(0) & R_{uu}(-1) & R_{uu}(-2) \\ R_{uu}(1) & R_{uu}(0) & R_{uu}(-1) \\ R_{uu}(2) & R_{uu}(1) & R_{uu}(0) \end{bmatrix}$$

Mas, como  $R_{uu}(x) = R_{uu}(-x)$ ,  $\mathbf{R}$  pode ser expressa como

$$\mathbf{R} = \begin{bmatrix} R_{uu}(0) & R_{uu}(1) & R_{uu}(2) \\ R_{uu}(1) & R_{uu}(0) & R_{uu}(1) \\ R_{uu}(2) & R_{uu}(1) & R_{uu}(0) \end{bmatrix}$$

→ O vetor  $\underline{P}$  é obtido por:

$$\begin{aligned} \underline{P} &= E \{d(n)\underline{u}(n)\} = \\ &= [E\{d(n)u(n)\} \ E\{d(n)u(n-1)\} \ \dots \ E\{d(n)u(n-M+1)\}]^T = \\ &= [P(0) \ P(-1) \ \dots \ P(1-M)]^T \end{aligned}$$

→ Lembrando, ainda, que o vetor de pesos que estamos buscando é expresso por

$$\underline{W} = [W_0 \ W_1 \ \dots \ W_{M-1}]^T$$

**Observação:** Objetivando reduzir a complexidade computacional envolvida no cômputo da Equação  $\underline{W} = \mathbf{R}^{-1} \underline{P}$ , como  $\mathbf{R}$  resulta em uma matriz Töeplitz, a sua inversão é, em geral, realizada pelo método de Durbin-Levinson, muito embora a pseudo-inversão de Moore-Penrose via Decomposição em Valores Singulares seja frequentemente utilizada para contornar os problemas resultantes de uma matriz  $\mathbf{R}$  quase singular.

- Para uma dada série temporal com  $N_t$  amostras totais, apresentando correlação entre  $M$  amostras consecutivas prévias ao instante a ser predito, **a precisão com que  $\mathbf{R}$  e  $\underline{P}$  representam as correlações envolvidas será tanto maior quanto maior for  $N_t$  com relação a  $M$ .**
- Isto ocorre porque, na prática, **não se conhece o processo aleatório subjacente que determina a série temporal em questão** (não são conhecidas as funções correlações que são realmente envolvidas no processo).
- Assim, o operador  $E\{\cdot\}$  é substituído pela **média dos vetores de  $M$  componentes envolvidos no cômputo de  $\mathbf{R}$  e  $\underline{P}$** , média esta realizada sobre o intervalo de  $N_t$  amostras totais conhecidas da série temporal.
- **A predição linear só tem sentido quando o processo aleatório subjacente é estacionário**, pois, em caso contrário,  $\mathbf{R}$  e  $\underline{P}$  não são univocamente definidas, mesmo para  $N_t$  suficientemente grande.  
  
Se a série temporal resulta de um processo aleatório não-estacionário,  $\mathbf{R}$  e  $\underline{P}$  variam ao longo da série, invalidando o uso da Equação de Wiener-Hopf para a obtenção do vetor de pesos  $\underline{W}$ .  
  
A solução algumas vezes adotada é assumir que a série temporal é estacionária em intervalos e adaptar  $\mathbf{R}$  e  $\underline{P}$  para cada intervalo. No entanto, o número de amostras em cada intervalo nem sempre é suficiente para expressar com fidelidade a operação  $E\{\cdot\}$ .
- Esta é a razão do uso cada vez mais disseminado de **técnicas de predição não-linear**, as quais, embora apresentem custo computacional maior, contornam a necessidade do conhecimento de um número grande de amostras passadas da série a ser predita, suficientes para que o operador  $E\{\cdot\}$  seja aproximado com fidelidade pela média temporal.

## Critério de Avaliação do Erro de Predição

Critério sugerido por Gershenfeld e Weigend em “A. S. Weigend and N. A. Gershenfeld. Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley Publishing Company, 1994”: A qualidade da predição (tanto linear quanto não-linear) é expressa em termos da razão entre as somas de erros quadráticos mostrada em (5.62).

$$\frac{\sum_t (\text{observação}_t - \text{predição}_t)^2}{\sum_t (\text{observação}_t - \text{observação}_{t-1})^2} \quad (5.62)$$

➤ Denominador: erro médio quadrático (MSE) de predição obtido para a chamada **predição pela última amostra**.

Tal método de predição considera que a melhor predição possível para a próxima amostra consiste simplesmente em repetir o valor efetivamente observado para a amostra atual.

O valor obtido por tal critério para o MSE é tomado como normalizador para o MSE resultante das diferenças entre os valores efetivamente obtidos, após a observação da amostra em questão, e os respectivos valores obtidos pelo preditor que está sendo avaliado.

Uma **razão inferior a 1.0** corresponde a uma predição melhor do que aquela obtida pela simples repetição do valor efetivamente observado para a amostra anterior àquela a ser predita – **limiar que qualifica um preditor que pretenda ser útil**.

➤ O erro obtido através do procedimento expresso em (5.62) é **chamado Erro Médio Quadrático Normalizado** (*Normalized Mean Squared Error*) e é referido na literatura por NMSE. Expressando (5.62) em forma de equação, teremos

$$\text{NMSE}(n) = \frac{\sum_{i=1}^n (o(i) - p(i))^2}{\sum_{i=1}^n (o(i) - o(i-1))^2} = \frac{\sum_{i=0}^n (u(i+1) - \hat{u}(i+1))^2}{\sum_{i=0}^n (u(i+1) - u(i))^2} \quad (5.63)$$

onde:  $o(i) \Rightarrow$  observação no instante  $i$  (o valor efetivamente observado)  
 $p(i) \Rightarrow$  predição no instante  $i$ .

➤ Para uma dada série temporal  $U$  com  $N_t$  amostras totais, o erro ao final do processo de predição de  $U$  é dado por  $\text{NMSE}_f = \text{NMSE}(N_t - 1)$ , onde  $N_t - 1$  é o índice do último elemento da série.

## Predição Linear de Séries Temporais: Estudo de Caso

- A série *Chaotic\_LASER* consta de  $N_t = 1000$  amostras.
- Cada um dos 1000 pontos que a constituem corresponde à intensidade de um *Far-Infrared-LASER* em estado caótico.
- A série foi obtida por Udo Huebner do *Phys.-Techn. Bundesanstalt*, Braunschweig, Germany.
- As medidas foram feitas a partir de um LASER NH3 não-pulsado.
- A série é descrita por Weigend e Gershenfeld em “A. S. Weigend and N. A. Gershenfeld. Time Series Prediction: Forecasting the Future and Understanding the Past. Addison-Wesley Publishing Company, 1994”.

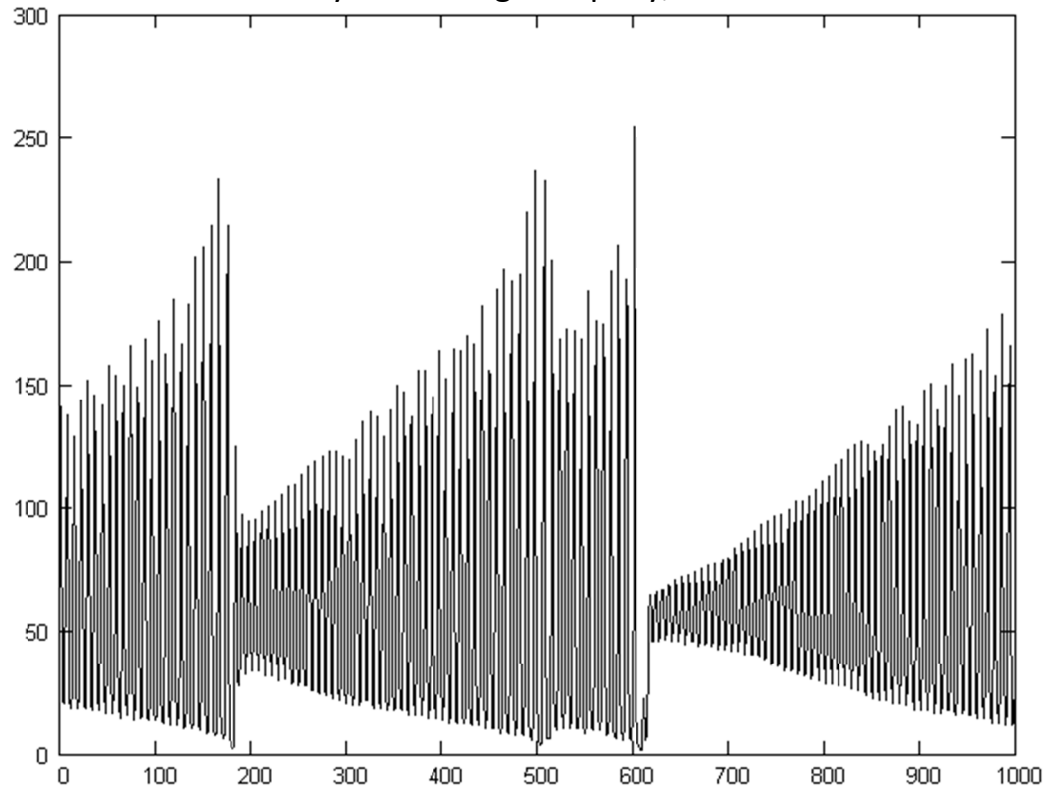
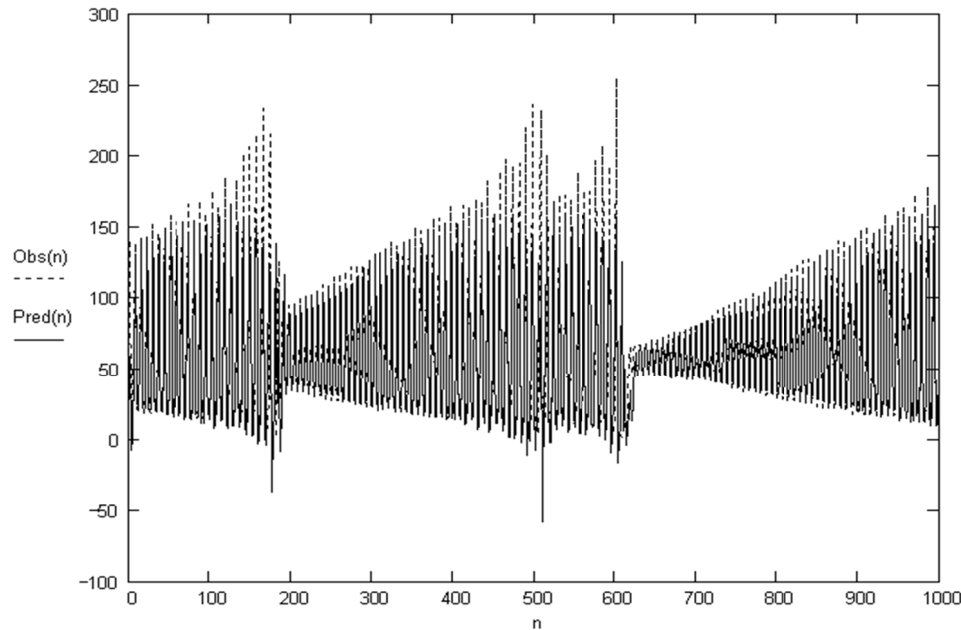


Figura 5.4: Representação gráfica da série *Chaotic\_LASER*.  
Ordenada: Intensidade de um *Far-Infrared-LASER* em estado caótico.  
Abscissa: Índice da medição.

Figura 5.5: Série *Chaotic\_LASER* - Observada e Predita (heurística de predição descrita em 5.2.1). Ordem da predição linear adotada:  $M = 11$ ;  $NMSE(N_t - 1) = 0.213$ .



Os valores para os coeficientes que ponderam tal combinação linear, obtidos através da Equação de Wiener-Hopf ( $\underline{W} = \mathbf{R}^{-1} \underline{P}$ ), são :

$$\begin{aligned}
 W_0 &= -0.689614 & W_1 &= 0.545837 \\
 W_2 &= -0.206001 & W_3 &= 0.144179 \\
 W_4 &= -0.109438 & W_5 &= 0.054155 \\
 W_6 &= -0.402716 & W_7 &= -0.334968 \\
 W_8 &= 0.221793 & W_9 &= -0.248085 \\
 W_{10} &= 0.0365866
 \end{aligned}$$

Assim, a amostra predita é dada por:  $\hat{u}(n+1) = -0.689614 u(n) + 0.545837 u(n-1) +$   
 $-0.206001 u(n-2) + 0.144179 u(n-3) - 0.109438 u(n-4) + 0.054155 u(n-5) - 0.402716 u(n-6) +$   
 $-0.334968 u(n-7) + 0.221793 u(n-8) - 0.248085 u(n-9) + 0.0365866 u(n-10)$

(Apesar de a ordem de predição ser  $M = 11$ , a predição linear necessita do conhecimento prévio de todos os  $N_t = 1000$  elementos da série temporal para montar a matriz de correlação  $\mathbf{R}$ ).

## Predição Linear de Séries Temporais: Exemplo Numérico

Dada a sequência  $m(n) = \{-2, -1.414, 0, 1.414, 2, 1.414, 0, -1.414, -2, -1.414, 0, \dots\}$ , determine  $\hat{m}(n)$  sabendo que o Preditor Linear é de ordem 2 e utiliza 11 amostras consecutivas de  $m(n)$  para a definição da matriz de correlação. Determine  $e(n)$  para  $\hat{m}(n)$ .

$$\hat{m}(n) = \hat{u}(n+1) = y(n) = \sum_{k=0}^{M-1} W_k u(n-k) = \underline{W}^T \underline{u}$$

$$\underline{W} = \mathbf{R}^{-1} \underline{P}$$

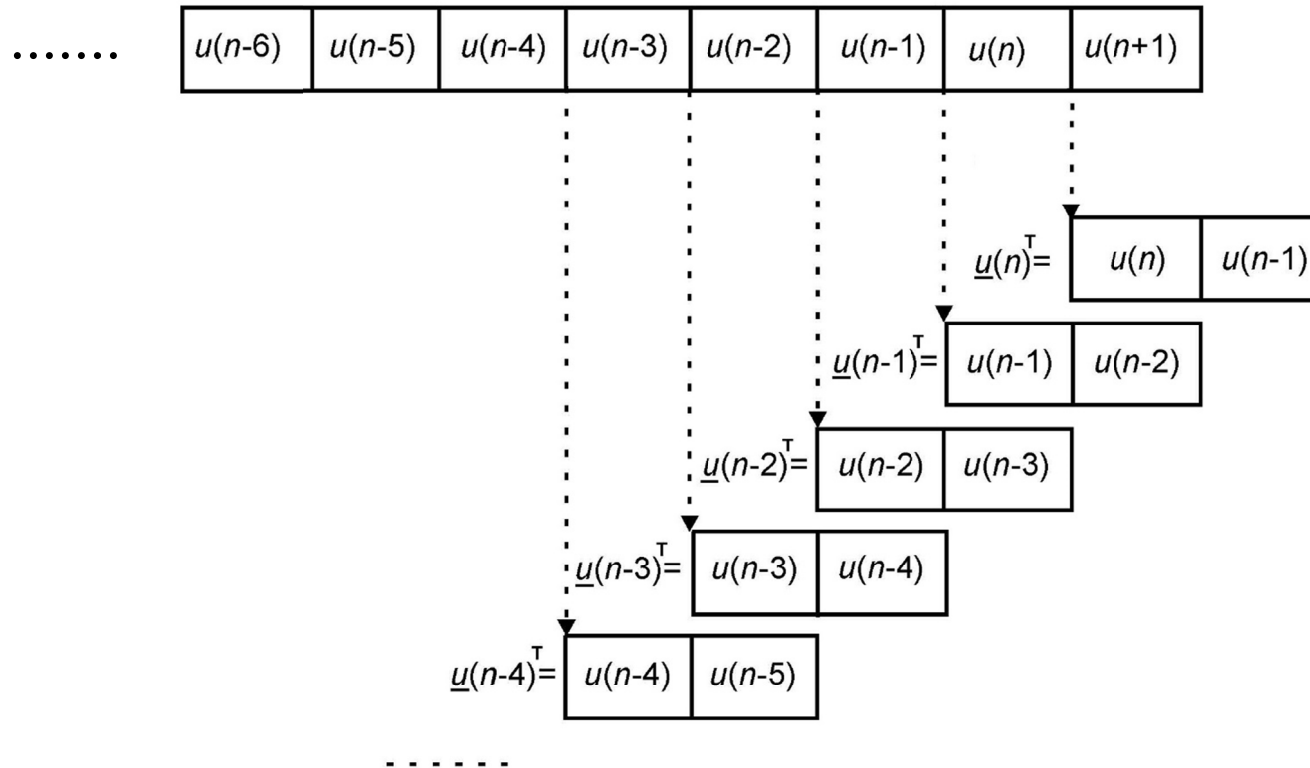
$$\mathbf{R} = \begin{bmatrix} r_{00} & r_{01} \\ r_{10} & r_{11} \end{bmatrix} \rightarrow \mathbf{R}^{-1} = \frac{1}{(r_{00} \cdot r_{11} - r_{01} \cdot r_{10})} \begin{bmatrix} r_{11} & -r_{01} \\ -r_{10} & r_{00} \end{bmatrix}$$



### Estrutura de dados para a predição:

$$m_q(n) = \{-2, -1.414, 0, 1.414, 2, 1.414, 0, -1.414, -2, -1.414, 0 \dots\} =$$

$$= \{u(n-10), u(n-9), u(n-8), u(n-7), u(n-6), u(n-5), u(n-4), u(n-3), u(n-2), u(n-1), u(n) \dots\}$$



$$m_q(n) = \{-2, -1.414, 0, 1.414, 2, 1.414, 0, -1.414, -2, -1.414, 0 \dots\} =$$

$$= \{u(n-10), u(n-9), u(n-8), u(n-7), u(n-6), u(n-5), u(n-4), u(n-3), u(n-2), u(n-1), u(n) \dots\}$$

$$\underline{u}(n) = \begin{bmatrix} u(n) \\ u(n-1) \end{bmatrix} = \begin{bmatrix} 0 \\ -1.414 \end{bmatrix}$$

$$\underline{u}(n-1) = \begin{bmatrix} u(n-1) \\ u(n-2) \end{bmatrix} = \begin{bmatrix} -1.414 \\ -2 \end{bmatrix}$$

$$\underline{u}(n-2) = \begin{bmatrix} u(n-2) \\ u(n-3) \end{bmatrix} = \begin{bmatrix} -2 \\ -1.414 \end{bmatrix}$$

...

$$\underline{u}(n-9) = \begin{bmatrix} u(n-9) \\ u(n-10) \end{bmatrix} = \begin{bmatrix} -1.414 \\ -2 \end{bmatrix}$$

## Determinação da matriz de auto-correlação $\mathbf{R}$ :

$$(m_q(n) = \{-2, -1.414, 0, 1.414, 2, 1.414, 0, -1.414, -2, -1.414, 0 \dots\})$$

$$\underline{u}(n)\underline{u}^T(n) = \begin{bmatrix} 0 \\ -1.414 \end{bmatrix} [0 \quad -1.414] = \begin{bmatrix} 0 & 0 \\ 0 & 1.9994 \end{bmatrix}_0$$

$$\underline{u}(n-5)\underline{u}^T(n-5) = \begin{bmatrix} 1.414 \\ 2 \end{bmatrix} [1.414 \quad 2] = \begin{bmatrix} 1.9994 & 2.8280 \\ 2.8280 & 4 \end{bmatrix}_5$$

$$\underline{u}(n-1)\underline{u}^T(n-1) = \begin{bmatrix} -1.414 \\ -2 \end{bmatrix} [-1.414 \quad -2] = \begin{bmatrix} 1.9994 & 2.8280 \\ 2.8280 & 4 \end{bmatrix}_1$$

$$\underline{u}(n-6)\underline{u}^T(n-6) = \begin{bmatrix} 2 \\ 1.414 \end{bmatrix} [2 \quad 1.414] = \begin{bmatrix} 4 & 2.8280 \\ 2.8280 & 1.9994 \end{bmatrix}_6$$

$$\underline{u}(n-2)\underline{u}^T(n-2) = \begin{bmatrix} -2 \\ -1.414 \end{bmatrix} [-2 \quad -1.414] = \begin{bmatrix} 4 & 2.8280 \\ 2.8280 & 1.9994 \end{bmatrix}_2$$

$$\underline{u}(n-7)\underline{u}^T(n-7) = \begin{bmatrix} 1.414 \\ 0 \end{bmatrix} [1.414 \quad 0] = \begin{bmatrix} 1.9994 & 0 \\ 0 & 0 \end{bmatrix}_7$$

$$\underline{u}(n-3)\underline{u}^T(n-3) = \begin{bmatrix} -1.414 \\ 0 \end{bmatrix} [-1.414 \quad 0] = \begin{bmatrix} 1.9994 & 0 \\ 0 & 0 \end{bmatrix}_3$$

$$\underline{u}(n-8)\underline{u}^T(n-8) = \begin{bmatrix} 0 \\ -1.414 \end{bmatrix} [0 \quad -1.414] = \begin{bmatrix} 0 & 0 \\ 0 & 1.9994 \end{bmatrix}_8$$

$$\underline{u}(n-4)\underline{u}^T(n-4) = \begin{bmatrix} 0 \\ 1.414 \end{bmatrix} [0 \quad 1.414] = \begin{bmatrix} 0 & 0 \\ 0 & 1.9994 \end{bmatrix}_4$$

$$\underline{u}(n-9)\underline{u}^T(n-9) = \begin{bmatrix} -1.414 \\ -2 \end{bmatrix} [-1.414 \quad -2] = \begin{bmatrix} 1.9994 & 2.8280 \\ 2.8280 & 4 \end{bmatrix}_9$$

$$\rightarrow \mathbf{R} = \frac{1}{10} \begin{bmatrix} 17.9970 & 14.1400 \\ 14.1400 & 21.9970 \end{bmatrix} = \begin{bmatrix} 1.7997 & 1.4140 \\ 1.4140 & 2.1997 \end{bmatrix}$$

## Determinação do vetor de correlação cruzada $\underline{P}$ :

$$(m_q(n) = \{-2, -1.414, 0, 1.414, 2, 1.414, 0, -1.414, -2, -1.414, 0 \dots\})$$

$$p / \underline{u}(n) \rightarrow d(n)\underline{u}(n) = \hat{u}(n+1)\underline{u}(n) = \hat{u}(n+1) \begin{bmatrix} 0 \\ -1.414 \end{bmatrix} = ?$$

$$p / \underline{u}(n-1) \rightarrow d(n)\underline{u}(n-1) = u(n)\underline{u}(n-1) = 0 \begin{bmatrix} -1.414 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}_0$$

$$p / \underline{u}(n-2) \rightarrow d(n)\underline{u}(n-2) = u(n-1)\underline{u}(n-2) = -1.414 \begin{bmatrix} -2 \\ -1.414 \end{bmatrix} = \begin{bmatrix} 2.8280 \\ 1.9994 \end{bmatrix}_1$$

$$p / \underline{u}(n-3) \rightarrow d(n)\underline{u}(n-3) = u(n-2)\underline{u}(n-3) = -2 \begin{bmatrix} -1.414 \\ 0 \end{bmatrix} = \begin{bmatrix} 2.8280 \\ 0 \end{bmatrix}_2$$

$$p / \underline{u}(n-4) \rightarrow d(n)\underline{u}(n-4) = u(n-3)\underline{u}(n-4) = -1.414 \begin{bmatrix} 0 \\ 1.414 \end{bmatrix} = \begin{bmatrix} 0 \\ -1.9994 \end{bmatrix}_3$$

$$p / \underline{u}(n-5) \rightarrow d(n)\underline{u}(n-5) = u(n-4)\underline{u}(n-5) = 0 \begin{bmatrix} 1.414 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}_4$$

### Determinação do vetor de correlação cruzada $\underline{P}$ (continuação):

$$(m_q(n) = \{-2, -1.414, 0, 1.414, 2, 1.414, 0, -1.414, -2, -1.414, 0 \dots\})$$

$$p / \underline{u}(n-6) \rightarrow d(n)\underline{u}(n-6) = u(n-5)\underline{u}(n-6) = 1.414 \begin{bmatrix} 2 \\ 1.414 \end{bmatrix} = \begin{bmatrix} 2.8280 \\ 1.9994 \end{bmatrix}_5$$

$$p / \underline{u}(n-7) \rightarrow d(n)\underline{u}(n-7) = u(n-6)\underline{u}(n-7) = 2 \begin{bmatrix} 1.414 \\ 0 \end{bmatrix} = \begin{bmatrix} 2.8280 \\ 0 \end{bmatrix}_6$$

$$p / \underline{u}(n-8) \rightarrow d(n)\underline{u}(n-8) = u(n-7)\underline{u}(n-8) = 1.414 \begin{bmatrix} 0 \\ -1.414 \end{bmatrix} = \begin{bmatrix} 0 \\ -1.9994 \end{bmatrix}_7$$

$$p / \underline{u}(n-9) \rightarrow d(n)\underline{u}(n-9) = u(n-8)\underline{u}(n-9) = 0 \begin{bmatrix} -1.414 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}_8$$

$$\rightarrow \underline{P} = \frac{1}{9} \begin{bmatrix} 11.3120 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.2569 \\ 0 \end{bmatrix}$$

Determinação do vetor de pesos  $\underline{W}$  do preditor linear:

$$\underline{W} = \mathbf{R}^{-1} \underline{P} = \begin{bmatrix} 1.7997 & 1.4140 \\ 1.4140 & 2.1997 \end{bmatrix}^{-1} \begin{bmatrix} 1.2569 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.1226 & -0.7216 \\ -0.7216 & 0.9185 \end{bmatrix} \begin{bmatrix} 1.2569 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.4110 \\ -0.9070 \end{bmatrix}$$

Determinação do valor estimado  $\hat{u}(n+1)$  para a 1ª. amostra subsequente à série conhecida:

$$\begin{aligned} \hat{u}(n+1) &= y(n) = \sum_{k=0}^{M-1} W_k u(n-k) = \\ &= \sum_{k=0}^1 W_k u(n-k) = W_0 u(n) + W_1 u(n-1) = (1.4110)(0) + (-0.9070)(-1.414) = 1.2825 \end{aligned}$$

Determinação do erro de predição  $e(n)$ :

$$m_q(n) = \{-2, -1.414, 0, 1.414, 2, 1.414, 0, -1.414, -2, -1.414, 0, \underline{1.414}, \dots\}$$

$$e(n) = 1.414 - 1.2825 = 0.1315$$