



INSTITUTO SUPERIOR TÉCNICO
PROCESSAMENTO DIGITAL DE SINAIS

Deteção de Anomalias em Dados de Produção de Energia

Mestrado Integrado em Engenharia Aeroespacial

Renato Loureiro, 89708
Miguel Félix, 87083

2020/21

1 Modelos de predição

R1.a)

Implementou-se em *MATLAB* o algoritmo de cálculo do coeficiente a , com o nome `r1a.m`.

R1.b)

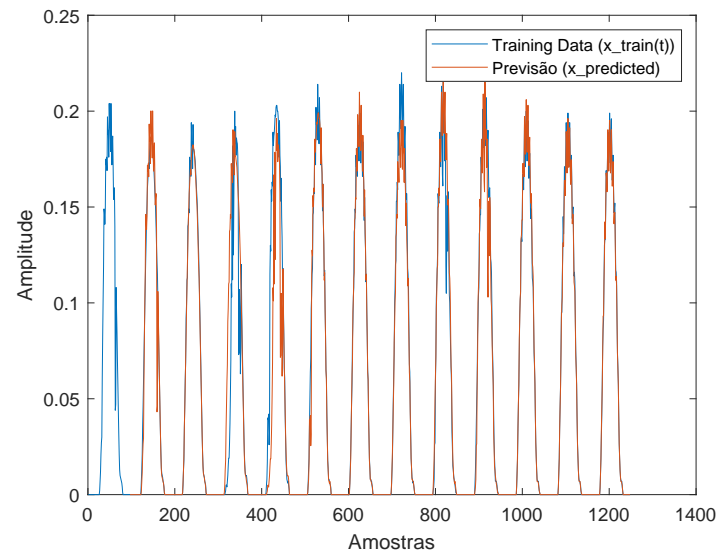


Figura 1: Gráficos dos dados fornecidos x_{train} versus previsão efetuada pelo modelo de longo prazo.

De notar que a previsão apenas tem valores apresentados no gráfico a partir da amostra 96 ($N = 96$), pois os dados das primeiras 96 amostras do x_{train} são usados para prever o segundo dia, não sendo possível prever o primeiro.

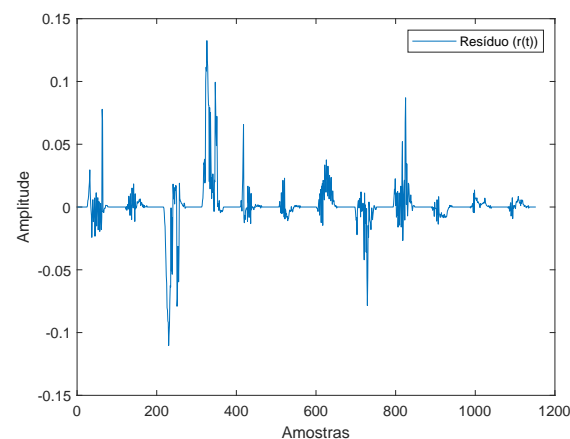


Figura 2: Gráfico do resíduo obtido pelo modelo de previsão de longo prazo.

Verifica-se através da análise do gráfico que o resíduo (obtido subtraindo a previsão à *training data*) apresenta valores que podem chegar a picos de amplitude de aproximadamente 0.13. Apesar de serem bons resultados, através da implementação do proposto na seguinte parte do enunciado esperam-se ainda melhores resultados.

R1.c)

O valor calculado para o coeficiente foi $a = 0.9810$.

Como seria de esperar, o valor do coeficiente é aproximadamente 1 pois as previsões são todas realizadas para a mesma altura do dia, o que implica que não haja grandes variações nas leituras feitas pelos painéis solares (o sol encontra-se aproximadamente sempre nas mesmas posições).

Para calcular a energia de um sinal discreto basta somar os quadrados de cada valor desse sinal, como implementado na função `energy.m`. O resultado obtido foi $energia_r = 0.3478$. O valor da energia do resíduo é baixo o que indica uma boa aproximação entre a previsão e os dados reais.

R1.d)

Implementou-se em *MATLAB* o algoritmo de cálculo dos coeficientes a_k , com o nome `calc_coef.m`.

R1.e)

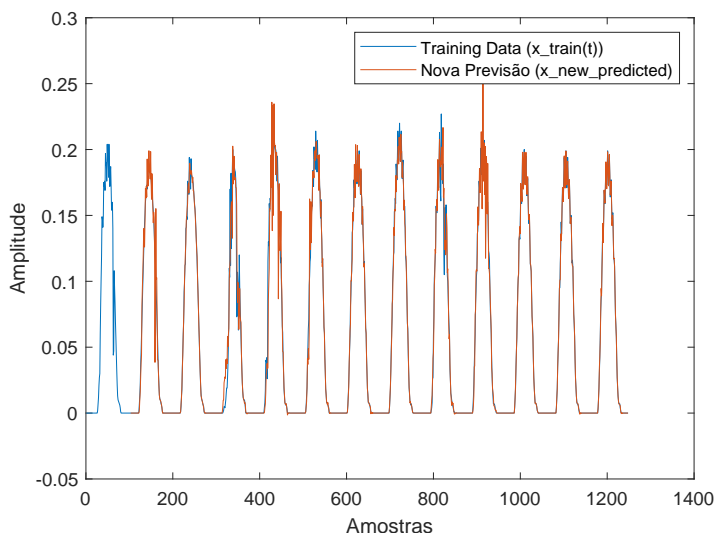


Figura 3: Gráficos dos dados fornecidos x_{train} versus previsão efetuada pelo modelo de longo prazo.

Novamente, e seguindo a mesma lógica da questão anterior (**R1.b**), a nova previsão apenas tem valores apresentados no gráfico a partir da amostra $96 + 6$ ($P = 6$). Os resultados aparentam ser ligeiramente melhores, como seria expectável.

Analisando o gráfico e comparando com o obtido anteriormente na seção **R1.b**) é possível verificar que as amplitudes, de uma forma geral, diminuíram, nunca chegando ao valor de amplitude de 0.1.

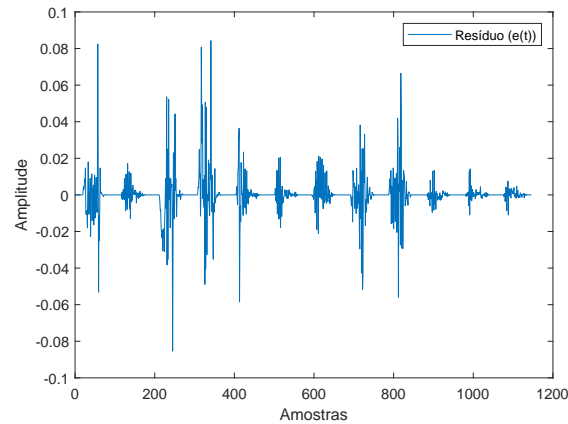


Figura 4: Gráfico do resíduo obtido pelo modelo de previsão de curto prazo.

R1.f)

Os valores calculados para os coeficientes foram $a_1 = 0.5993$, $a_2 = 0.1496$, $a_3 = -0.0031$, $a_4 = 0.2730$, $a_5 = -0.1552$ e $a_6 = -0.0279$.

O resultado obtido foi $energia_e = 0.1248$. Pode constatar-se que o valor da energia do resíduo é ainda mais baixo do que na computação anterior o que indica uma melhor previsão efetuada. Isto seria expectável pois esta previsão tem por base a previsão feita anteriormente e funciona como que um complemento, obtendo valores ainda mais próximos dos reais.

2 Detecção de anomalias

R2.a) A seguinte função determina o conjunto de amostras consideradas como anomalias considerando um limite de diferença entre o valor real e a predição de 0.11.

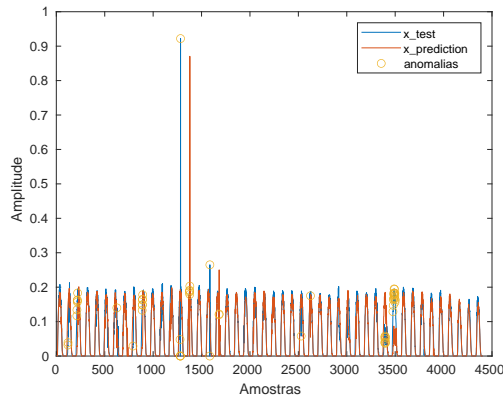
```

1 function [time, x, good] = anomaly_detection(x_real,x_prediction)
2     thr=0.11;
3     j=1;
4     for i=1:length(x_real)
5         if(abs(x_real(i)-x_prediction(i))>thr)
6             time(j)=i;
7             x(j)=x_real(i);
8             j=j+1;
9         end
10    end
11    good=x_real;
12    for i=1:length(time)
13        good(time(i))=x_prediction(time(i));
14    end
15 end

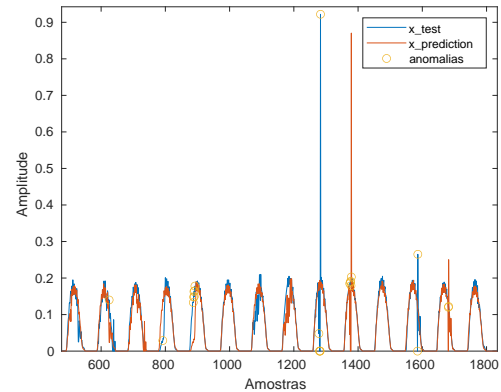
```

R2.b)

Nas seguintes figuras observa-se os resultados da aplicação do modelo de predição 1 e 2 (longo e curto) e as respectivas anomalias identificadas usando a função apresentada anteriormente.

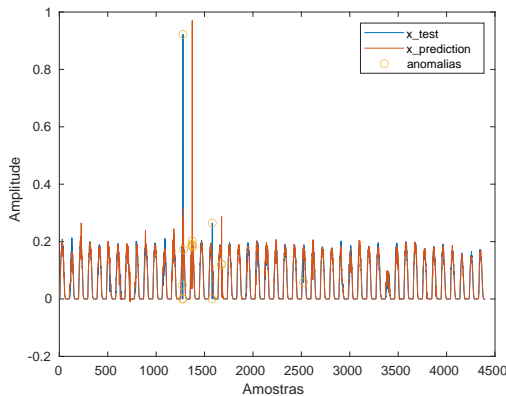


(a) Comparação de dados reais e de predição e anomalias

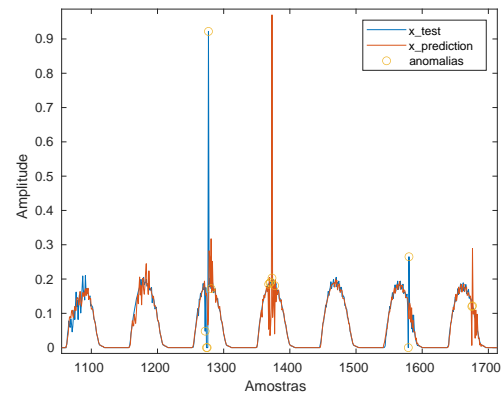


(b) Zoom da figura 5a

Figura 5: Modelo 1 - predição a longo prazo



(a) Comparação de dados reais e de predição e anomalias



(b) Zoom da figura 6a

Figura 6: Modelo 2 - predição a curto prazo

Verifica-se que com ambos os modelos é possível determinar as anomalias existentes, sendo que para o mesmo *threshold* o modelo de longo prazo determina mais possíveis anomalias, pois este como apenas considera relações a longo prazo não é capaz de adaptar a algumas variações ao longo do tempo, i.e. o modelo de longo prazo pode levar à consideração de falsas anomalias. Relativamente ao modelo de curto prazo, este é capaz de selecionar melhor as anomalias mais agressivas e as pequenas variações ao longo do tempo ele é capaz de as prever.

Se se considerar como anomalia o sucedido nas amostras que rondam a amostra 3400, o modelo 2 não é capaz de detetar, pois é uma variação generalizada da produção de energia desse dia. Mas o modelo de longo prazo como considera apenas amostras do dia anterior, consegue identificar este tipo de fenómenos.

R2.c)

Na secção anterior verifica-se que quando ocorre uma anomalia num certo período, as previsões do período seguinte são corrompidas pela existência da anomalia do período anterior dado que o modelo de previsão utilizado recorre a dados do dia anterior - isto leva a existirem falsas anomalias no dia seguinte.

Nesta secção propõe-se duas soluções, que funcionam quer na utilização do modelo longo prazo quer no curto prazo. A primeira abordagem é utilizar o modelo para prever anomalias do dia seguinte e caso estas existam, substitui-se o instante da ocorrência da anomalia pela previsão feita ou utiliza-se as amostras dos instantes anteriores para determinar o valor no instante da anomalia.