Universidade de Pernambuco

Escola Politécnica de Pernambuco

Programa de Pós-Graduação em Engenharia da Computação

Probabilidade e Processos Estocásticos

Prof. Mêuser Valença

Mestrando: Marcelo Lacerda

Relatório – Previsão de Séries Temporais Utilizando Modelos de Box e Jenkins

A série escolhida para realizar modelagem e predição foi a de Itaipu, Paraná. A mesma é exibida na Figura 1. Pretende-se encontrar um modelo ARIMA que represente satisfatoriamente a série em questão, mas utilizando o menor número de coeficientes possível.

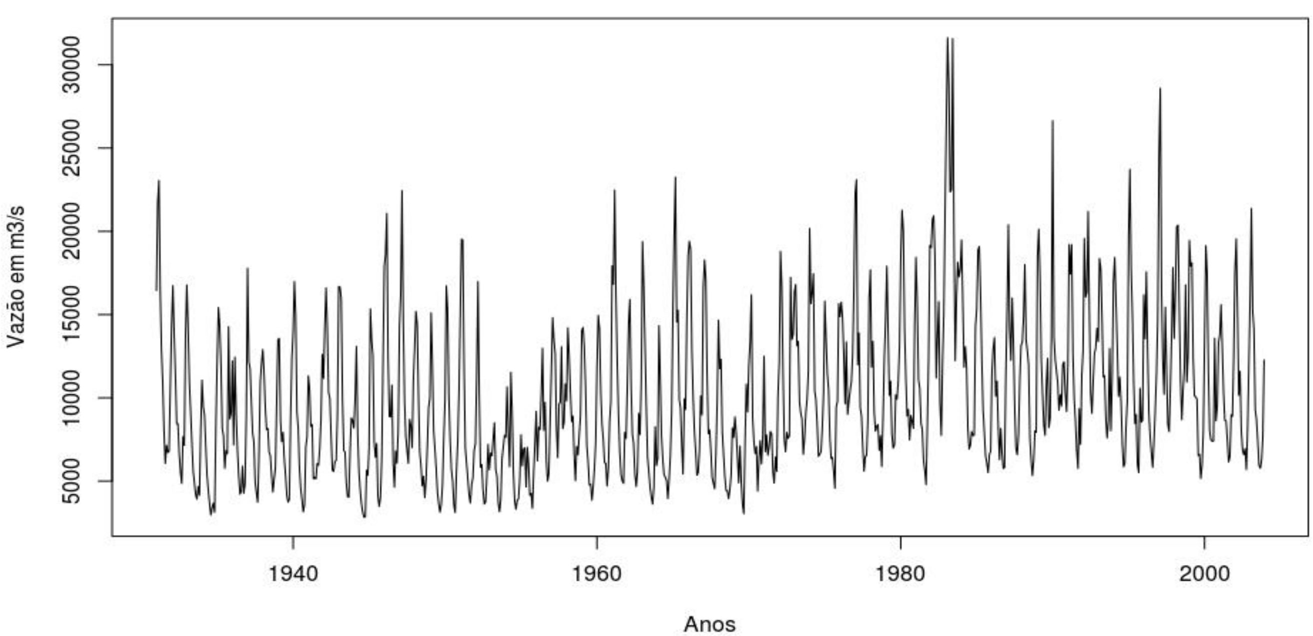

Figura 1. Série temporal de vazão em Itaipu.

O primeiro passo a ser tomado é verificar a existência de sazonalidade e tendência na série. Para verificar a existência de tendência, foi realizado o teste de Dickey-Fuller, o qual, considerando o modelo $y_y = \rho y_{t+1} + \varepsilon_t$, determina se existe raiz unitária no mesmo, ou seja, determina se o valor de ρ da série representada por tal modelo é menor do que 1. Caso seja menor do que 1, a série é estacionária, pois a cada novo valor de y, o seu valor antigo é decrescido de uma determinada taxa. Caso contrário, a série não é estacionária, pois o valor de y acumula com o passar do tempo e cresce indefinidamente. Para este teste, o seu retorno (p-value) deve ser menor do que 0,05. Nesta série, obtivemos p-value=0,01, o que significa que a mesma é estacionária. Sendo assim, não há necessidade de realizar diferenciações para torná-la não-estacionária, visto que os modelos ARIMA funcionam apenas em tais condições.

Para verificar a existência de sazonalidade, deve-se observar o correlograma (ACF) exibido na Figura 2. Como pode ser notado, picos idênticos de ocorrem nos lags 12 e 24, ou seja, com um período fixo de 12 meses. Isso significa que a série possui uma sazonalidade de período de 12. A mesma informação pode ser retirada da Figura 3 no gráfico onde é descrita a componente sazonal da série, verifica-se a repetição de um padrão anualmente. Este gráfico exibe tais dados referentes à faixa de 1994 a 2004.

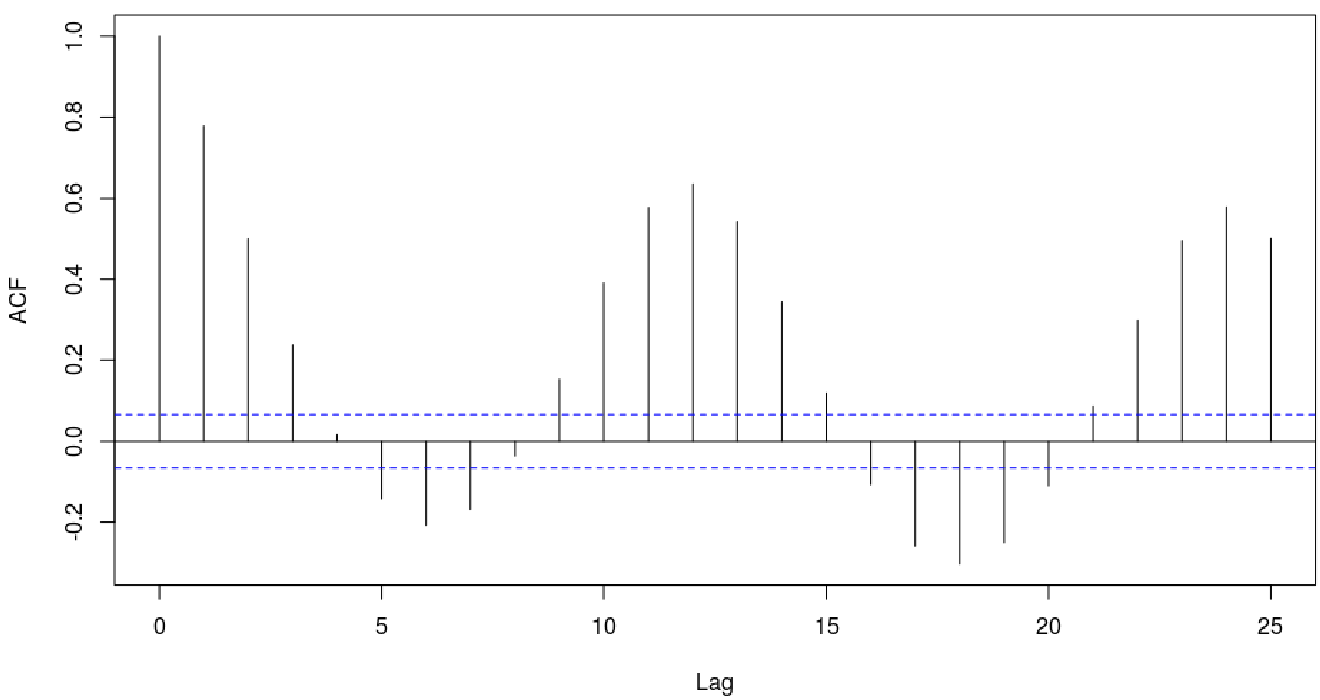

Figura 2. ACF da série original.

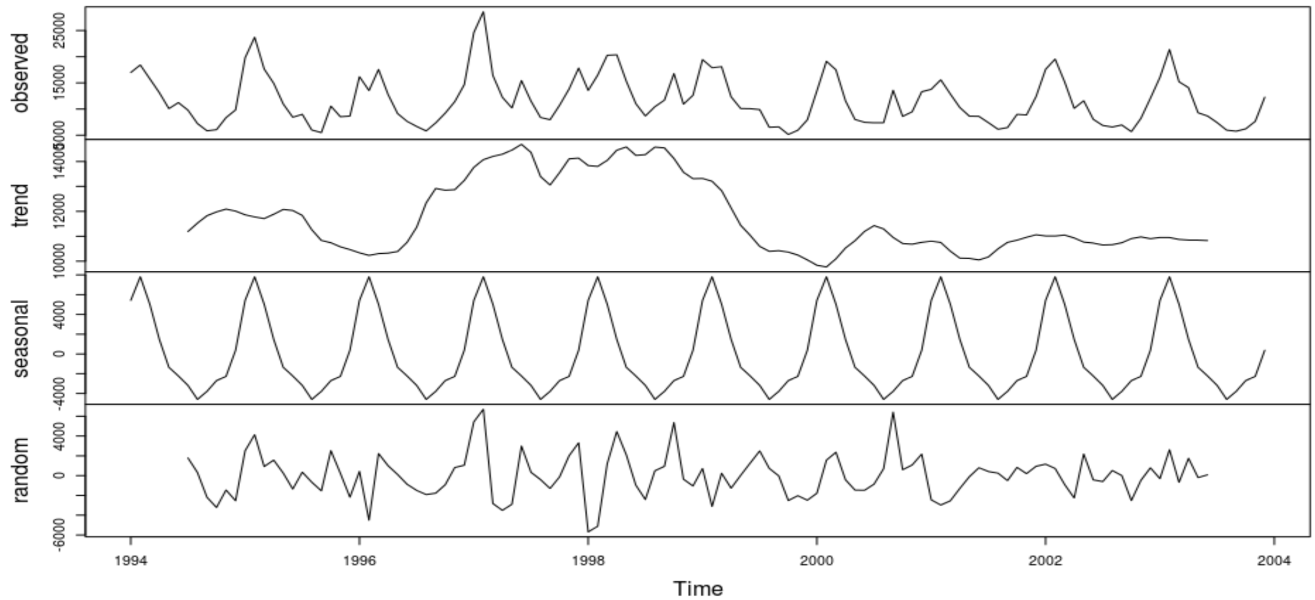

Figura 3. Decomposição da série nas componentes de tendência, sazonalidade e componente aleatória (ruido).

Verificada a existência de sazonalidade, torna-se necessária o uso de modelos SARIMA. A primeira decisão a ser tomada refere-se às diferenciações que devem ser feitas na série original. Diferenciações simples são utilizadas para remover tendência na série. Devido à não existência de tal característica na mesma, tal diferenciação não será necessária. No entanto, foi observado que a série é sazonal. Sendo assim, deve-se aplicar diferenciações sazonais, objetivando remover tal sazonalidade. O *lag* utilizado em tal diferenciação é de 12 meses, devido ao fato dos dados serem representados mensalmente e haverem 12 meses dentro de um ano. *A* Figura 4 exibe a série com 1 diferenciação sazonal. Na Figura 5, observa-se que, devido à não ocorrência de de picos periódicos no correlograma da série diferenciada, pode-se concluir que uma única diferenciação sazonal foi suficiente para remover a sazonalidade da série original.

Deve-se agora definir os parâmetros p, d, q, P, D e Q do modelo SARIMA. Os valores de d e D já foram determinados, pois referem-se às diferenciações simples e sazonais, respectivamente, ou seja, 0 e 1. Os valores de p e q podem ser extraídos através da observação dos correlogramas de autocorrelação parcial e autocorrelação. Porém, esta não é uma tarefa trivial, e depende muito da experiência do analista. Neste trabalho, determinou-se, heuristicamente, que os valores de *lag* a partir do qual os valores absolutos de ACF e PACF (Figura 6) passam a ficar abaixo de 0.5 (baixa autocorrelação, considerando que o valor absoluto máximo é 1 e o valor absoluto médio é 0.5) serão q e p, respectivamente. Neste caso, p=2 e q=3. Os valores de de P e Q foram determinados por tentativa e erro.

Porém, recomenda-se que a soma dos mesmos não seja maior ou igual a 2.

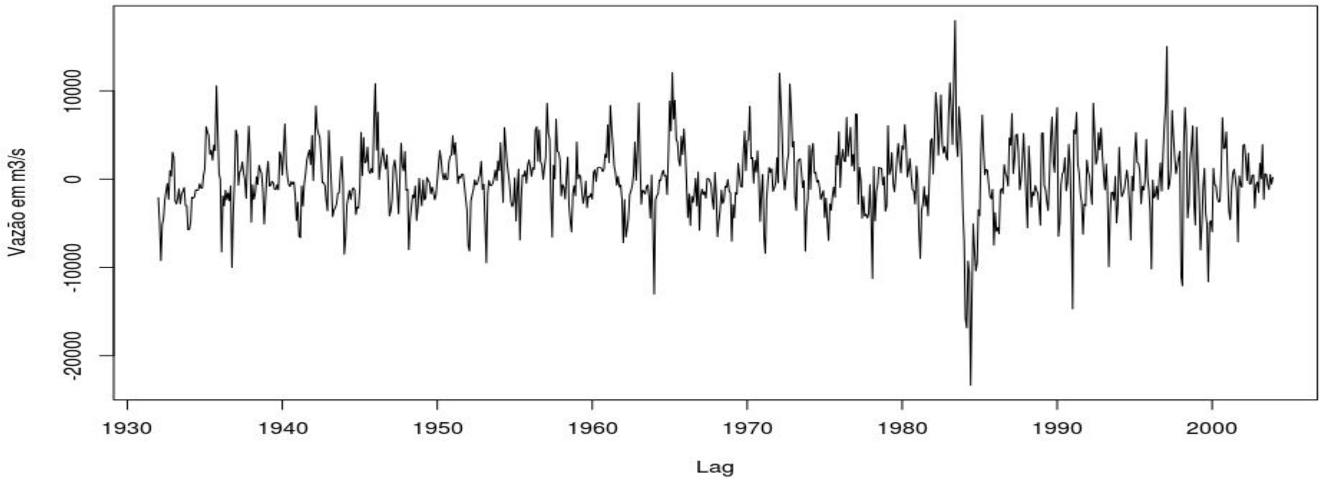

Figura 4. Série com 1 diferenciação sazonal.

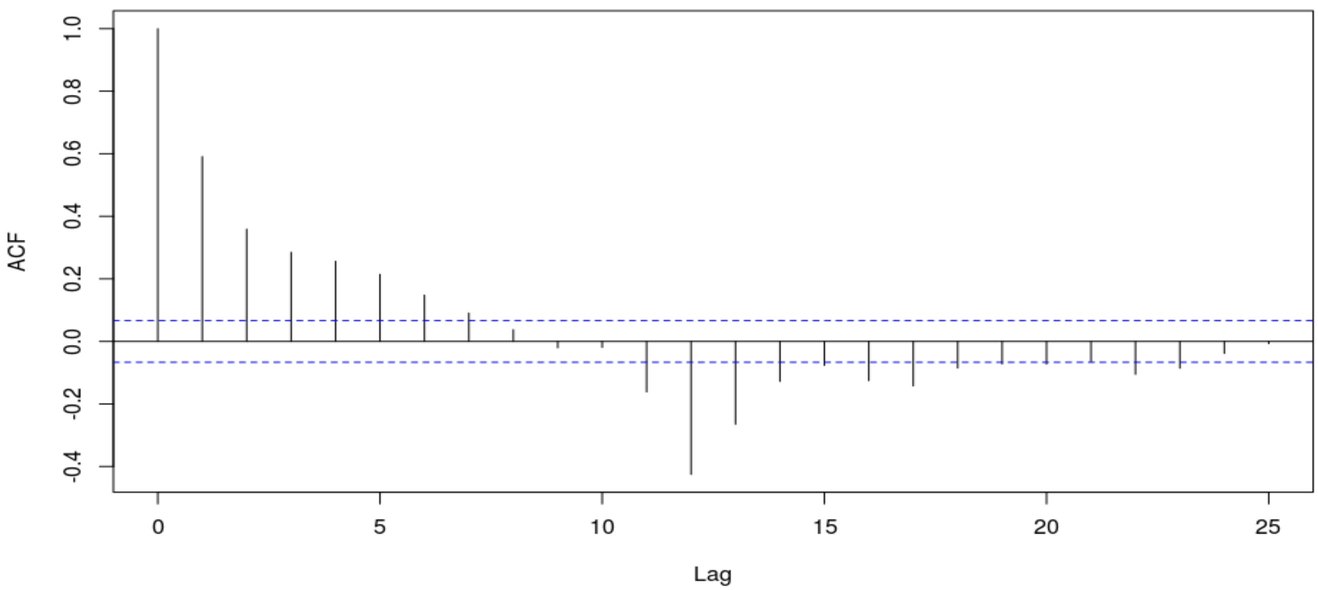

Figura 5. Correlograma (ACF) da série com 1 diferenciação sazonal.

Seguindo o que foi descrito acima, encontrou-se o modelo SARIMA(2,0,3)x(1,1,0) como o mais parcimonioso, obtendo um melhor equilíbrio entre o valor de AIC (27,75) e a quantidade de coeficientes (6). A Figura 7 exibe a série prevista a partir de 1986 em comparação com a série real.

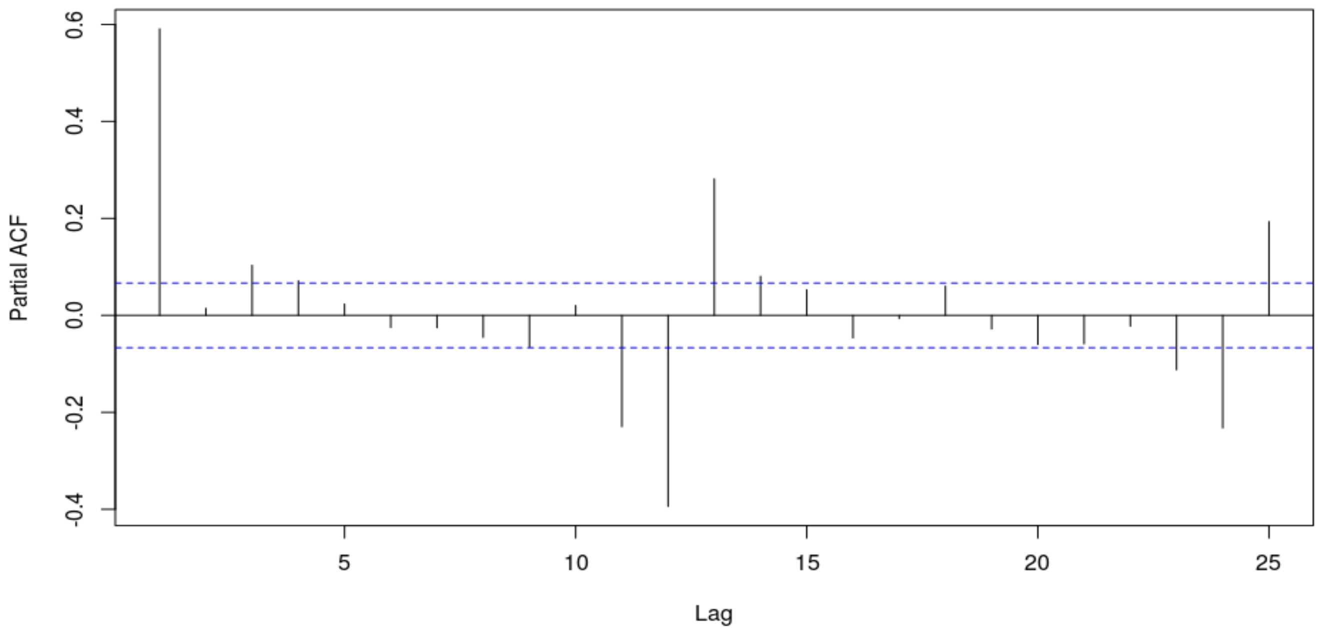

Figura 6. Correlograma de Autocorrelação parcial (PACF) da série com 1 diferenciação parcial.

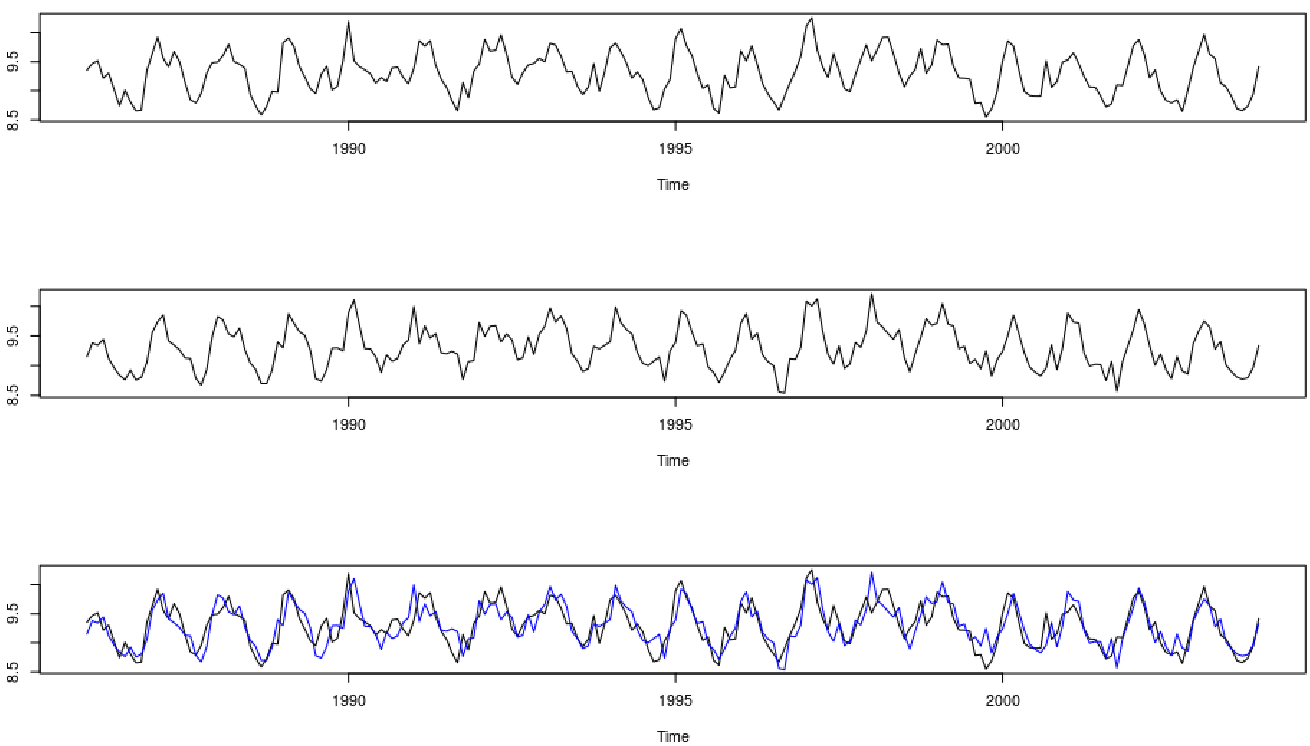

Figura 7. Série real/Série prevista/Série real (preto) e série previsa (azul).