

## 9. Conclusões

“Forecasting in business is like sex in society, we have to have it, we cannot get along without it, everyone is doing it, one way or another, but nobody is sure he is doing it the best way. “

G. W. Plossl  
Last Frontiers for Profits

Esta tese apresentou uma proposta do uso de um método de clusterização baseado em densidade, o DBSCAN, na fase de separação do ruído da Análise Singular Espectral de Séries Temporais e mostrou, através dos resultados em simulações e aplicação da metodologia a uma série real de velocidade do vento, que a metodologia proposta é uma inovação de sucesso para Análise Singular do Espectro de Séries Temporais e também para Análise de Séries Temporais,

Quando comparada às outras quatro abordagens já tradicionais, a abordagem proposta nesta tese, que usa a combinação SSA+DBSCAN, rendeu, em todos os casos simulados, estacionários e não estacionários, e na série real de velocidade do vento, previsões de melhor qualidade quando comparadas com aquelas obtidas pelos outros métodos tradicionais, considerando como critério o MAPE, o RMSE, o RMSE normalizado e o  $R^2$ , e também para o MAD e GMRAE em análises de previsões out of sample. Tal resultado de previsão sugere afirmar que o DBSCAN quando aplicado a SSA está separando melhor o que são de fato as componentes de ruído e fornecendo uma série menos ruidosa para previsão.

A previsão SSA é uma abordagem recente que vem ganhando espaço e popularidade ao apresentar melhores resultados do que os métodos de modelagem e previsão clássicos de séries temporais, como pode ser visto em Hassani (2007) ; Hassani

et al. (2009), Esquivel (2012), Hassani et al. (2013)-b e Hassani et al (2013)-b e isto foi confirmado neste trabalho: a previsão SSA apresentou para as quatro abordagens de separação de ruído, as melhores previsões para a série.

Importante notar também que em todos os casos, a previsão sem utilizar SSA é a de menor eficiência em todas as avaliações, confirmando o sucesso do uso da Análise Singular Espectral na Análise de Séries Temporais antes da modelagem e previsão. Até a abordagem SSA com separação de ruído de menor qualidade (a que usa clusterização hierárquica) supera a qualidade de previsão do Modelo Box & Jenkins. Ressalta-se também que, mesmo usando abordagem Clássica de Previsão, é interessante filtrar a série temporal por SSA, por qualquer método de separação de ruído, pois a modelagem seguida de filtragem SSA também tem melhor qualidade de ajuste do que a modelagem com a série original.

Neste trabalho foi verificado que o método de Clusterização Hierárquico, implementado no R no pacote RSSA para separação de ruído entre as autotriplas da SVA deve ser usado com cautela, pois ele tende a superestimar as séries de ruído, deixando o sinal sem informações importantes para modelagem da série. Para tal, recomenda-se o uso do DBSCAN que é um método mais robusto e especializado a separar componentes de ruído.

A série de velocidade do vento em Petrolina analisada aqui possui uma sazonalidade bem definida apresentada nos correlogramas. E pela abordagem clássica pôde-se verificar que a velocidade do vento num momento atual pode ser descrita por velocidade do vento dos mesmos meses dos dois anos anteriores e também carrega informação da velocidade do vento no mês anterior, por isso a série foi bem modelada por um SARIMA(1,0,0)x(3,1,2). Mesmo quando filtrada por SSA a série continua com esta estrutura SARIMA(1,0,0)x(3,1,2). Mas as melhores previsões da série foram obtidas por previsão SSA usando DBSCAN para separar as componentes de ruído.

No setor elétrico, métodos de previsão em séries temporais estão sempre em discussão, pois a área demanda controle e previsão de várias variáveis, como a velocidade do vento na produção de energia eólica. Por isso o interesse em descobrir métodos e abordagens novas ou melhorar métodos já existentes é crescente e atual. Não se tem a pretensão de dizer que o proposto nesta tese seja ideal e eficiente em todas as análises de quaisquer séries temporais. Entretanto este trabalho deixa uma contribuição, robusta, não paramétrica, de seleção e identificação automática, que uma vez tendo as ferramentas computacionais necessárias, não depende de palpite subjetivo do pesquisador, por isso é uma alternativa útil a ser testada.

A proposta de combinação do método DBSCAN associado a SSA para análise de séries temporais é inédita na literatura e provou ser eficiente em melhorar os resultados de previsão SSA e ainda mais os resultados de modelagem clássica de Séries Temporais, com grandes vantagens: não requer análise prévia expert do pesquisador, pois não precisa de inspeção especialista para detecção das componentes de ruído, como é requisitada na Análise Visual do comportamento Gráfico dos Autovetores; não requer pré-análise para a identificação de modelo, como é necessário para a modelagem Box & Jenkins; não é tão restritiva para classificar ruídos como a determinação feita pela Análise de Componentes Principais; e oferece eficiente classificação de ruído e clusterização mais robusta e automática do que o método de clusterização hierárquica.

O fato de o DBSCAN associado a SSA não precisar de inspeção especialista para detecção das componentes de ruído, como é requisitada na Análise Visual do comportamento Gráfico dos Autovetores é uma grande vantagem pois tal análise pictórica de um conjunto de gráficos pode não ser prática, quando o número de componentes é elevado como, por exemplo, em séries horárias.

Uma idéia para trabalhos futuros seria usar o DBSCAN com Distância de Mahalanobis, ao invés da distância euclidiana. A Distância de Mahalanobis leva em consideração a matriz de covariâncias calculada com todos os objetos e é amplamente utilizada em [análise de clusters](#), mas no caso específico desta abordagem, em que os objetos a serem clusterizados são séries temporais, considerar a matriz de covariâncias

pode ser mais interessante ainda, capturando esta informação de covariância entre as séries e deixar isso influenciar na formação dos clusters que apresentem a máxima verossimilhança.

Outra idéia seria utilizar DBSCAN usando distâncias adaptativas tal como foi usada com sucesso por Cavalcanti Junior (2006) para o método de Clusterização Fuzzy c-Means e por SOUZA *et al.* (2003) no método de clusterização k-means. Nesta abordagem deseja-se clusterizar séries temporais. Sabe-se que uma série pode ter muita similaridade a outra em um período e nem tanta similaridade em outro; as distâncias adaptativas podem ajustar bem essa característica.

E, enfim, outra abordagem que pode ser analisada é a eficiência de um algoritmo genético para definir os parâmetros  $\varepsilon$  e *MinPts* do DBSCAN.