



**Keila Mara Cassiano**

**Análise de Séries Temporais Usando Análise Espectral  
Singular (SSA) e Clusterização de Suas Componentes  
Baseada em Densidade**

**Tese de Doutorado**

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro  
Setembro de 2014



**Keila Mara Cassiano**

**Análise de Séries Temporais Usando Análise  
Espectral Singular (SSA) e Clusterização de Suas  
Componentes Baseada em Densidade**

Tese apresentada como requisito parcial para obtenção grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

**Prof. Reinaldo Castro Souza**

Orientador

Departamento de Engenharia Elétrica - PUC-Rio

**Prof.<sup>a</sup> Ana Maria Lima de Farias**

UFF

**Prof. Basílio de Bragança Pereira**

UFRJ

**Prof. José Francisco Moreira Pessanha**

UERJ

**Prof.<sup>a</sup> Karla Tereza Figueiredo Leite**

Departamento de Engenharia Elétrica - PUC – Rio

**Prof. José Eugenio Leal**

Coordenador Setorial do Centro Técnico Científico – PUC-Rio

Rio de Janeiro, 12 de setembro de 2014.

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da Universidade, do autor e do orientador.

### **Keila Mara Cassiano**

É Bacharel em Matemática pela Universidade Federal de Goiás, possui Mestrado em Estatística pela Universidade Federal de Pernambuco. Desenvolveu junto com o seu orientador uma grande quantidade de trabalhos e artigos científicos apresentados em congressos nacionais e internacionais. Também publicou artigos em revistas internacionais durante o Doutorado em Engenharia Elétrica da PUC-Rio.

#### Ficha Catalográfica

Cassiano, Keila Mara

Análise de Séries Temporais Usando Análise Espectral Singular (SSA) e Clusterização de Suas Componentes Baseada em Densidade/ Keila Mara Cassiano; orientador: Reinaldo Castro Souza. – 2014.

172 f. : il. (color.) ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2014.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Singular Spectrum Analysis. 3. Clusterização Baseada em Densidade (DBSCAN) 4. Previsão de Séries Temporais. 5. Previsão SSA. 6. Modelos ARIMA. 7. Energia Eólica. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Para meus tesouros  
Moisés Menezes,  
Pedro Arthur  
e Matheus Henrique,  
com todo meu amor.

## Agradecimentos

Agradeço a Deus pela vida e por Seu cuidado, por me fortalecer, proteger, capacitar e iluminar a cada dia, permitindo que eu chegasse até aqui.

Ao meu esposo Moisés, pelo seu nobre amor, cuidado e parceria. Pela paciência e compreensão nas minhas ausências, por ser a calma nos momentos de preocupação, por ser sempre minha fortaleza, acreditar quando eu não acreditei, e me guiar nas decisões quando eu estava desorientada.

Aos meus amados filhos Pedro Arthur e Matheus Henrique por tantas vezes serem abdicados da minha atenção, carinho e dedicação e mesmo assim iluminarem meus dias com doces momentos que equilibram as tensões dos períodos de estudos e me nutrem de prazer e felicidade plenos.

Ao professor Reinaldo Souza, pela sua orientação segura, por seu conhecimento e disponibilidade, pelo apoio e acolhimento nos momentos críticos que tive neste doutorado. Pela calma, confiança e otimismo que ele sempre me transferiu e, principalmente pela compreensão e paciência diante das minhas falhas.

Ao Professor José Francisco Pessanha, por todas as colaborações, orientações e ideias enriquecedoras, pela parceria em diversos trabalhos, disponibilidade exemplar, amizade, apoio e incentivo constantes.

Aos meus pais, Iva Cassiano e Antônio Cassiano, por sempre terem feito o melhor pra promoverem meu desenvolvimento e crescimento. Por sempre acreditarem na minha capacidade, me incentivando e apoiando em todos os momentos, especialmente nestes últimos anos problemáticos.

Às minhas doces avós Adelaide e Manoelina que estão sempre intercedendo por mim e enchendo minha vida de candura e amor.

Aos meus familiares que sempre me auxiliaram diretamente, principalmente às minha queridas Dadá e Dedé, Greice Carla e Kênia Cristina, Kátia Kelvis e Carlos Lozano, pelo apoio e ajuda que me deram quando eu precisei.

À minha amiga-filha Danielle Freitas, pela amizade, carinho e compreensão. Aos meus médicos Dr. Rodrigo e Dra. Olga, que foram muito mais que médicos, realmente amigos, me ajudando a superar e vencer os desafios dos momentos de enfermidade com profissionalismo e dedicação incomum.

Aos professores do Departamento de Engenharia Elétrica da PUC-Rio pela competência e pelo conhecimento transferido à minha formação.

Ao Professor Alexandre Street por sua orientação inicial, pois sem seu apoio e acolhimento eu não teria ingressado neste curso.

Aos colegas também alunos desta Pós-Graduação, pela convivência enriquecedora, especialmente aos colegas Luiz Albino e Rafael Moraes, pela parceria em trabalhos, e Daiane Rodrigues e Soraida Aguilar, flores neste caminho.

A todos os meus amigos pessoais que me incentivavam, torciam, compreendiam minhas ausências e estiveram sempre do meu lado, especialmente ao meu amigo Tinus van Vugt pela sua presença nesta fase da minha vida, apoio, carinho e dedicação.

À UFF e todos os colegas professores do Departamento de Estatística pela licença concedida para eu me dedicar ao doutorado.

Aos meus chefes superiores no CEDERJ e UFF, Masako Masuda, Ana Maria, Celso Costa e Regina Moreth por todo apoio e compreensão.

À PUC-Rio pelo acolhimento, seus funcionários e recursos físicos.

Aos membros da Banca Examinadora pelas importantes contribuições.

A todos os autores referenciados, que enriquecem meu trabalho com suas pesquisas.

À cidade maravilhosa do Rio de Janeiro e seu Samba, paixões da minha vida, que neutralizam todo stress e me dão tanta alegria.

Ao Povo Brasileiro, financiador direto desta conquista.

## Resumo

Cassiano, Keila Mara; Souza, Reinaldo Castro (Orientador). **Análise de Séries Temporais Usando Análise Espectral Singular (SSA) e Clusterização de Suas Componentes Baseada em Densidade**. Rio de Janeiro, 2014, 172p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esta tese propõe a utilização do DBSCAN (*Density Based Spatial Clustering of Applications with Noise*) para separar os componentes de ruído na fase de agrupamento das autotriplas da Análise Singular Espectral (SSA) de Séries Temporais. O DBSCAN é um método moderno de clusterização (revisto em 2013) e especialista em identificar ruído através de regiões de menor densidade. O método de agrupamento hierárquico até então é a última inovação na separação de ruído na abordagem SSA, implementado no pacote R- SSA. No entanto, o método de agrupamento hierárquico é muito sensível a ruído, não é capaz de separá-lo corretamente, não deve ser usado em conjuntos com diferentes densidades e não funciona bem no agrupamento de séries temporais de diferentes tendências, ao contrário dos métodos de aglomeração à base de densidade que são eficazes para separar o ruído a partir dos dados e dedicados para trabalhar bem em dados a partir de diferentes densidades. Este trabalho mostra uma melhor eficiência de DBSCAN sobre os outros métodos já utilizados nesta etapa do SSA, garantindo considerável redução de ruídos e proporcionando melhores previsões. O resultado é apoiado por avaliações experimentais realizadas para séries simuladas de modelos estacionários e não estacionários. A combinação de metodologias proposta também foi aplicada com sucesso na previsão de uma série real de velocidade do vento.

## Palavras-chave

Análise Singular Espectral; Clusterização Baseada em Densidade; DBSCAN; Mineração de Dados; Previsão; Séries Temporais; Previsão SSA; Modelos ARIMA; Energia Eólica.

## Abstract

Cassiano, Keila Mara; Souza, Reinaldo Castro (Advisor). **Time Series Analysis Using Singular Spectrum Analysis (SSA) and Based Density Clustering of The Components.** Rio de Janeiro, 2014, 172p. Doctorate Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis proposes using DBSCAN (Density Based Spatial Clustering of Applications with Noise) to separate the noise components of eigentriples in the grouping stage of the Singular Spectrum Analysis (SSA) of Time Series. The DBSCAN is a modern (revised in 2013) and expert method at identify noise through regions of lower density. The hierarchical clustering method was the last innovation in noise separation in SSA approach, implemented on package R-SSA. However, is repeated in the literature that the hierarquical clustering method is very sensitive to noise, is unable to separate it correctly, and should not be used in clusters with varying densities and neither works well in clustering time series of different trends. Unlike, the methods of density based clustering are effective in separating the noise from the data and dedicated to work well on data from different densities This work shows better efficiency of DBSCAN over the others methods already used in this stage of SSA, because it allows considerable reduction of noise and provides better forecasting. The result is supported by experimental evaluations realized for simulated stationary and non-stationary series. The proposed combination of methodologies also was applied successfully to forecasting real series of wind's speed.

## Keywords

Singular Spectrum Analysis; Density Based Clustering; DBSCAN; Data Mining; Forecast; Time Series; SSA Forecast; ARIMA Models; Wind Energy.



## Sumário

1.	Introdução	18
1.1.	Motivação	19
1.2.	Objetivo do Trabalho	21
1.3.	Relevância do Tema	21
1.4.	Organização da Tese	22
2.	Séries Temporais	23
2.1.	Definição	23
2.2.	O Operador de Diferenças	26
2.3.	Processo Ruído Branco	27
2.4.	Processo Passeio Aleatório	27
2.5.	Modelos ARIMA	28
2.6.	Medidas de Qualidade de Ajuste	32
3.	Singular Spectrum Analysis (SSA)	34
3.1.	Decomposição SSA	36
3.1.1.	Incorporação	36
3.1.2.	<i>Singular Value Decomposition (SVD)</i>	37
3.2.	Reconstrução SSA	41
3.2.1.	Agrupamento	41
3.2.2.	Média Diagonal	42
3.3.	Métodos de Agrupamento	45
3.3.1.	Agrupamento por Análise de Componentes Principais	46
3.3.2.	Agrupamento por Análise Gráfica dos Autovalores Singulares	46
3.3.2.1.	Análise do Periodograma	47
3.3.3.	Agrupamento por Clusterização	48
3.4.	Separabilidade	48

3.5.	Escolha do Parâmetro L de Defasagem	50
3.6.	Previsão SSA	53
3.7.	SSA na Análise de Séries Temporais	55
4.	Clusterização de Dados	59
4.1.	Aplicações	60
4.2.	Limitação	62
4.3.	Medidas de Similaridade	62
4.4.	Métodos de Clusterização	67
4.4.1.	Métodos Hierárquicos	69
4.4.2.	Métodos Particionais	70
4.4.3.	Métodos Baseados em Densidade	72
4.4.4.	Métodos Baseados em Grade	73
4.4.5.	Métodos Baseados em Modelos	74
4.4.6.	Métodos Baseados em Redes Neurais Artificiais	74
4.4.7.	Métodos Baseados em Lógica <i>Fuzzy</i>	75
4.4.8.	Métodos Baseados em <i>Kernel</i>	76
4.4.9.	Métodos Baseados em Grafos	76
4.4.10.	Métodos Baseados em Computação Evolucionária	77
4.4.11.	Histórico dos Métodos de Clusterização	78
5.	Método DBSCAN	97
5.1.	Algoritmo DBSCAN	104
5.2.	DBSCAN Revisado	106
5.3.	Avaliação de Performance e Aplicações do DBSCAN	110
6.	Metodologia	114
6.1.	Processos Geradores de Dados Simulados	115
6.2.	Ferramentas Computacionais	116
7.	Resultados das Avaliações Dados Simulados	117
8.	Aplicação da Metodologia à Série de Velocidade do Vento	120

8.1. Modelagem e Previsão Tradicional (sem SSA)	122
8.2. Modelagem e Previsão usando SSA	131
8.2.1. Identificação do Ruído por Verificação Visual dos Vetores Singulares	132
8.2.2. Identificação do Ruído por Análise de Componentes Principais	136
8.2.3. Identificação do Ruído pelo Método de Clusterização Hierárquica	138
8.2.4. Identificação do Ruído pelo Método de Clusterização DBSCAN	139
8.3. Previsão Mista: Filtro SSA + Previsão Clássica	141
8.3.1. Modelagem para a série aproximada obtida por verificação visual dos autovetores	142
8.3.2. Modelagem para a série aproximada obtida por ACP	143
8.3.3. Modelagem para a série aproximada obtida por Clusterização Hierárquica	144
8.3.4. Modelagem para a série aproximada obtida por Clusterização DBSCAN	144
8.4. Previsão SSA: Filtro SSA + Previsão SSA	145
8.5. Comparação dos Resultados da Previsão Clássica, Previsão Mista e Previsão SSA	147
9. Conclusões	151
10. Referências Bibliográficas	155

## Lista de Figuras

Figura 3.1. Estimaco de L para filtragem SSA	52
Figura 4.1. Conjuntos com clusters globulares, no globulares e com rudos	73
Figura 5.1. $\epsilon$ -vizinhana de q e $\epsilon$ -vizinhana de p	98
Figura 5.2. Alcance direto por densidade no mtodo DBSCAN	99
Figura 5.3. Alcance por densidade no mtodo DBSCAN	100
Figura 5.4: Conexo por densidade no mtodo DBSCAN	100
Figura 5.5. Pseudo Cdigo do Algoritmo DBSCAN - Subrotina Principal	105
Figura 5.6. Pseudo Cdigo do Algoritmo DBSCAN - Subrotina de Expanso dos Clusters	106
Figura 5.7. Pseudo Cdigo do Algoritmo DBSCAN- Revisado - Subrotina Principal	109
Figura 5.8. Pseudo Cdigo do Algoritmo DBSCAN- Revisado - Subrotina de Expanso dos Clusters	110
Figura 5.9. Desempenho de Diferentes Mtodos de Clusterizao para Dados Espaciais	111
Figura 7.1. Amostras de cada um dos processos estacionrios simulados	117
Figura 7.2: Amostras de cada um dos processos no estacionrios simulados	118
Figura 8.1. Curva de Potncia Eltrica Produzida em funo da velocidade do vento	119
Figura 8.2. Mdias mensais de velocidade do vento na Estaco Petrolina - PE	122
Figura 8.3. Box Plot e Histograma da Distribuo de velocidade do vento na Estaco Petrolina - PE	123
Figura 8.4. Boxplots das mdias de velocidade de vento na Estaco de Petrolina	124
Figura 8.5. Funo de Autocorrelao da srie de Velocidade do	

Vento de Petrolina	125
Figura 8.6. Função de Autocorrelação Parcial da série de Velocidade do Vento de Petrolina	126
Figura 8.7. Função da Autocorrelação da série $\nabla_{12}^1 Y_t$ .	127
Figura 8.8. Função da Autocorrelação Parcial da série $\nabla_{12}^1 Y_t$ .	128
Figura 8.9. Estimação do modelo SARIMA $(1,0,0) \times (1,0,1)_{12}$ para $Y_t$ pelo FPW	128
Figura 8.10. Estimação do modelo SARIMA $(1,0,0) \times (1,0,1)_{12}$ para $Y_t$ pelo FPW.	129
Figura 8.11. Série dos resíduos do modelo SARIMA $(1,0,0) \times (1,0,1)_{12}$	130
Figura 8.12. Autocorrelação dos resíduos do modelo	130
Figura 8.13. Logaritmo dos autovalores da matriz $XX^T$	131
Figura 8.14. Os seis primeiros vetores singulares em ordem de significância	132
Figura 8.15. Alguns diagramas de dispersão	133
Figura 8.16. Periodograma dos vetores singulares 2 e 3.	134
Figura 8.17. Componentes da metodologia SSA identificados por Verificação Visual do Comportamento dos Autovetores.	135
Figura 8.18. Série de Sinal Identificado por SVD	136
Figura 8.19. Série de Ruído identificado por SVD.	137
Figura 8.20. Série de Sinal Determinado por Clusterização Hierárquica.	138
Figura 8.21 . Série de Ruído Determinada por Clusterização Hierárquica.	138
Figura 8.22: Dist4 dos dados de Petrolina para a estimação de $\varepsilon$ pro algoritmo DBSCAN.	140
Figura 8.23 . Série de Sinal Identificada por DBSCAN.	140
Figura 8.24. Série de Ruído Identificada por DBSCAN.	141
Figura 8.25: Resíduos do Modelo ARIMA para séries filtradas em SSA através da Verificação Visual dos Autovetores, calculados em relação à série original	143

Figura 8.26 : Resíduos do Modelo ARIMA para séries filtradas em SSA através da Análise de Componentes Principais, calculado em relação à série original	143
Figura 8.27: Resíduos do Modelo ARIMA para séries filtradas em SSA através da Clusterização Hierárquica, calculados em relação à série original	144
Figura 8.28: Resíduos do Modelo ARIMA para séries filtradas em SSA através da DBSCAN, calculados em relação à série original	145
Figura 8.29: Resíduos da Previsão SSA para séries filtradas através da Verificação Visual do Comportamento dos Autovetores	145
Figura 8.30: Resíduos da Previsão SSA para séries filtradas através da Análise de Componentes Principais.	146
Figura 8.31 : Resíduos da Previsão SSA para séries filtradas através da Clusterização Hierárquica.	146
Figura 8.32 : Resíduos da Previsão SSA para séries filtradas através do DBSCAN	146

## Lista de Quadros e Tabelas

Quadro 4.1: Evolução dos Métodos de Clusterização	89
Tabela 7.1: Média do MAPE (in sample) da previsão de séries de Modelos Simulados, em cada uma das abordagens propostas, a partir das 100 replicações	119
Tabela 7.2: Média do RMSE (in sample) da previsão de séries de Modelos Simulados em cada uma das abordagens propostas, a partir das 100 replicações	113
Tabela 8.1. Principais Estatísticas da Série Original de Petrolina	122
Tabela 8.2: Principais estatísticas, por mês, das medidas de velocidade do vento em Petrolina/BA.	124
Tabela 8.3. Definições das componentes pelo método da Verificação Visual.	134
Tabela 8.4. Teste BDS aplicado a série de ruídos identificados por Verificação Visual.	135
Tabela 8.5. Correlação ponderada entre as componentes identificadas por Verificação Visual.	136
Tabela 8.6. Teste BDS aplicado a série de ruídos identificados por SVD.	137
Tabela 8.7. Teste BDS aplicado à série de ruídos identificados pela Clusterização Hierárquica.	139
Tabela 8.8. Definições das componentes pelo método DBSCAN.	140
Tabela 8.9. Teste BDS aplicado a série de ruídos identificados por DBSCAN	141
Tabela 8.10: Medidas de qualidade de ajuste das previsões <i>in sample</i> da série de Petrolina por todos os métodos experimentados.	147
Tabela 8.11: Medidas de qualidade de ajuste das previsões 24 passos a frente out of sample da série de Petrolina por todos os métodos experimentados.	149

Tabela 8.12. Previsões passos a frente pelo método SSA com DBSCAN e Previsão SSA para a série de velocidade do vento da estação de Petrolina

150



Não entendo.  
Isso é tão vasto que ultrapassa qualquer entender.  
Entender é sempre limitado,  
mas não entender pode não ter fronteiras.  
Sinto que sou muito mais completa quando não entendo.  
O bom é ser inteligente e não entender.  
É uma benção estranha, como ter loucura sem ser doída.  
É um desinteresse manso, é uma doçura de burrice.  
Só que de vez em quando vem a inquietação,  
quero entender um pouco...”

Clarice Lispector