



Universidades Lusíada

Duarte, Hercílio Rui Dinis, 1987-

Estudo da detecção de anomalias no consumo de energia eléctrica da cidade de Luanda

<http://hdl.handle.net/11067/2526>

Metadados

Data de Publicação	2016-07-13
Resumo	O consumo de energia eléctrica da cidade de Luanda de momento apresenta muitas transformações e facilmente poderá desencadear muitos interesses para a comunidade académica e científica. Uma das questões que poderá ser fundamental a partir deste estudo é o facto de poder associar as técnicas de computação utilizadas a ciências económicas, estatística, meteorológicas, etc. Acabando por demonstrar o grande interesse e demonstrar o poder de suas aplicações. A Computação Científica é uma área actua...
Palavras Chave	Consumo de energia - Angola - Luanda, Consumo de energia - Previsões - Angola - Luanda, Algoritmos computacionais
Tipo	masterThesis
Revisão de Pares	Não
Coleções	[ULL-FCEE] Dissertações

Esta página foi gerada automaticamente em 2024-04-29T18:13:45Z com informação proveniente do Repositório



UNIVERSIDADE LUSÍADA DE LISBOA

Faculdade de Ciências da Economia e da Empresa

Mestrado em Ciências da Computação

Estudo da detecção de anomalias no consumo de energia eléctrica da cidade de Luanda

Realizado por:
Hercílio Rui Dinis Duarte

Orientado por:
Prof. Doutor João José Amaro Amador Leitão

Constituição do Júri:

Presidente:	Prof. Doutor Mário Caldeira Dias
Orientador:	Prof. Doutor João José Amaro Amador Leitão
Arguente:	Prof. Doutor Paulo Jorge Gonçalves Pinto
Vogal:	Prof. Doutor Eng. Joaquim Mesquita da Cunha Viana

Dissertação aprovada em: 3 Fevereiro de 2015

Lisboa

2014



U N I V E R S I D A D E L U S Í A D A D E L I S B O A
Faculdade de Ciências da Economia e da Empresa
Mestrado em Ciências da Computação

Estudo da detecção de anomalias no consumo de
energia eléctrica da cidade de Luanda

Hercílio Rui Dinis Duarte

Lisboa

Outubro 2014



UNIVERSIDADE LUSÍADA DE LISBOA
Faculdade de Ciências da Economia e da Empresa
Mestrado em Ciências da Computação

**Estudo da detecção de anomalias no consumo de energia
eléctrica da cidade de Luanda**

Hercílio Rui Dinis Duarte

Lisboa

Outubro 2014

Hercílio Rui Dinis Duarte

Estudo da detecção de anomalias no consumo de energia eléctrica da cidade de Luanda

Dissertação apresentada à Faculdade de Ciências da Economia e da Empresa da Universidade Lusíada de Lisboa para a obtenção do grau de Mestre em Ciências da Computação.

Orientador: Prof. Doutor João José Amaro Amador Leitão

Lisboa

Outubro 2014

Ficha Técnica

Autor(a) Hercílio Rui Dinis Duarte
Orientador(a) Prof. Doutor Eng. João José Amaro Amador Leitão
Título Estudo da detecção de anomalias no consumo de energia eléctrica da cidade de Luanda
Local Lisboa
Ano 2014

Mediateca da Universidade Lusíada de Lisboa - Catalogação na Publicação

DUARTE, Hercílio Rui Dinis, 1987-

Estudo da detecção de anomalias no consumo de energia eléctrica da cidade de Luanda / Hercílio Rui Dinis Duarte ; orientado por João José Amaro Amador Leitão. - Lisboa : [s.n.], 2014. - Dissertação de Mestrado em Ciências da Computação, Faculdade de Ciências da Economia e da Empresa da Universidade Lusíada de Lisboa.

I – LEITÃO, João José Amaro Amador, 1967-

LCSH

1. Consumo de energia - Angola – Luanda
2. Consumo de energia – Previsões - Angola – Luanda
3. Algoritmos computacionais
4. Universidade Lusíada de Lisboa. Faculdade de Ciências da Economia e da Empresa - Teses
5. Teses - Portugal - Lisboa

1. Energy consumption - Angola – Luanda
2. Energy consumption - Forecasting - Angola – Luanda
3. Computer algorithms
4. Universidade Lusíada de Lisboa. Faculdade de Ciências da Economia e da Empresa - Dissertations
5. Dissertations, Academic - Portugal - Lisbon

LCC

1. HD9502.D83 2014

DEDICATÓRIA

Não devia existir um dia que justificasse o que és, pois o que és ultrapassa as barreiras da imaginação.

Prefiro às vezes calar, sem palavras a sublinhar, sem grandeza a descrever sem saber o que dizer.

Não há palavras que expressem o sentimento de seres minha mãe.

Dedico a ti!

AGRADECIMENTOS

À Deus, a explicação de todas as ciências e do inimaginável... O senhor do poder e de todas as crenças. O lógico da vida em espírito. A explicação da existência dos céus e dos mares.

Ao Professor Doutor João José Amaro Amador Leitão, o meu orientador da Dissertação pela paciência e muita dedicação, graça as suas persistências fui capaz de ter maiores noções em relação aos caminhos a seguir para o alcance do objectivo.

Ao Professor Eng. Gilberto Capessa, que desde a minha Licenciatura até cá tem sabido orientar e acompanhar nas minhas grandes tomadas de decisões.

Ao Dr. Hélder da Rocha Machado, responsável da Mediateca da Universidade Lusíada de Lisboa, sempre esteve pronto em ajudar e dar muito bons conselhos.

Aos Professores Doutor Paulo Pinto, Doutora Isabel Alvarez e Doutor Marco Costa.

À Dra. Anabela Neto, responsável pelo Gabinete de Estágios, Saídas Profissionais e Empreendedorismo da ULL, pelo profissionalismo, disponibilidade e acompanhamento em todo percurso.

À minha mãe, luz da minha vida, por acreditar que um dia eu conseguiria concretizar boa parte dos meus projectos, por passar várias noites sentindo a minha falta mas com força suficiente por acreditar que a cada dia que passa os nossos sonhos iriam aproximar-se cada vez mais da realidade. Às minhas irmãs, pela força constante e incentivo.

Ao meu pai, o arquitecto do meu incentivo... que Deus o tenha! Sempre foi a voz mais alta que me levasse a enfrentar vários tipos de realidades por mais que parecessem impossíveis. A ti devo muito mais do que um dia irei conseguir pagar, pois nada justifica a tua bondade. És tudo e muito mais, obrigado meu grande Pai.

À minha esposa, Ledna Duarte, pela paciência que teve ao longo destes anos, por longos conselhos durante as noites enquanto duravam meus estudos, por ser a maravilhosa conselheira e também uma boa orientadora.

Aos meus colegas de Luta, Adjah da Cruz, Helder Francisco e Zenilda Gongga pela colaboração e conclusão de vários projectos académicos, e pelo grande incentivo em jamais desistir quando tudo é possível de se conquistar.

“A coisa mais bela que podemos experimentar é o mistério. Essa é a fonte de toda a arte e ciências verdadeiras.”
Albert Einstein.”

APRESENTAÇÃO

Estudo na detecção de anomalias no consumo de energia eléctrica da cidade de Luanda

Hercílio Rui Dinis Duarte

O consumo de energia eléctrica da cidade de Luanda de momento apresenta muitas transformações e facilmente poderá desencadear muitos interesses para a comunidade académica e científica.

A Computação Científica é uma área actualmente pouco explorada, mas que fornece muitas soluções que poderão ser aproveitadas para a compreensão de fenómenos que envolveriam muitos passos sem o auxílio de seus recursos.

No caso deste estudo, viemos apresentar um conjunto de métodos que visam a detecção de mudanças ou anomalias e se for necessário a sua melhoria pela readaptação dos filtros aplicados.

Será demonstrado o percurso que um algoritmo poderá tomar para prever e detectar mudanças em dados de consumo. Para tal, é necessário caracterizar e conhecer-se bem a natureza destes dados de modo a optar-se por uma metodologia correcta para a sua estruturação. Assim sendo, a escolha dos dados de consumo pressupõe haver algum método eficaz que facilitará a sua manipulação, interpretação e projecção futura. Os dados de consumo de energia eléctrica poderão ser interpretados como sendo uma estrutura em série temporal que facilmente poderá ser manipulada e que permite a extracção de padrões interessantes. Neste caso, importará fazer-se um estudo sobre as séries temporais, seus principais propósitos e achar um modelo óptimo capaz de permitir que sejam fornecidas respostas no decorrer da dissertação.

Para que o objectivo possa ser cumprido, iremos basear-nos na descrição comportamental e suas mudanças em intervalos de tempo previamente determinados, através do algoritmo que poderá ser capaz de apresentar os resultados em formas

diferentes mediante um filtro que se baseia nos métodos *Moving Average* e Periodicidade admitindo reajustes.

Optou-se por trabalhar com dados de energia eléctrica uma vez que apresentam características muito próprias para se demonstrar o poder e vantagem dos algoritmos.

A cidade de Luanda actualmente, por viver um crescimento significativo na sua população, poderá implicar ou não algum tipo de adaptação nos modelos e métodos a serem estudados que por sua vez poderão originar resultados interessantes.

A aplicação do CUSUM é importante por se tratar de um método clássico aplicável que poderá descrever mudanças no processo e a aplicação dos métodos *Moving Average* e Periodicidade poderão filtrar melhor a obtenção dos resultados, assim como ajudar com maior precisão na detecção e estudo de suas tendências.

Palavras-chave: Algoritmos, Detecção de mudanças, anomalias, CUSUM, Média Móvel, Séries temporais, Sazonalidade, Tendência, Periodicidade.

PRESENTATION

Study on detection of anomalies in electricity consumption of Luanda

Hercílio Rui Dinis Duarte

The electricity consumption of the city of Luanda currently presents many changes and can easily trigger many concerns for the academic and scientific community.

The Scientific Computing is an area currently underexplored, but it provides many solutions that can be utilized for understanding phenomena that involve many steps without the aid of their resources.

In this study, we came to present a paradigm to detect whether any change or abnormality on historical consumption data was necessary to resort to methods that can solve any case upon the rehabilitation of applied filters.

We shall demonstrate the way that an algorithm can take to detect and predict changes in consumption data, for such is necessary characterize and know well the nature of these data to opt for a proper methodology for structuring. Therefore, the choice of consumption data assumes there is some effective method which will facilitate their manipulation, interpretation and future projection. The data of electrical energy consumption can be interpreted as a structure of time series that can easily be manipulated and which permits the extraction of interesting patterns. In this case, is important make up a study of the time series, its main purpose and find an optimal model to allow responses to be provided during the dissertation.

We chose to work with data of electricity since they have very specific features to demonstrate the power and advantages of the algorithms on which we will apply.

For the purpose can be fulfilled, we will build on behavioral description and their changes in time intervals previously determined by the algorithm that will be able to present the

results in different ways through a filter that bases itself on Moving Average methods or Periodicity admitting adjustments.

The city of Luanda by currently experiencing a significant growth in their population may or may not involve some sort of adjustment in models and methods to be studied which in turn may lead to interesting results.

Application of CUSUM is important because it is a classic method that can apply to describe changes in the process and the application of Moving Average method and Periodicity can be used as filter to return better results, as well as help with more accurate detection and study of their trends.

Keywords: Algorithms, anomaly detection, CUSUM, Moving Average, time series, seasonality, trend.

LISTA DE ILUSTRAÇÕES

Ilustração I – Total de clientes activos 2009 – 2011 [EDEL – Empresa de Distribuição de Electricidade, 2011]	20
Ilustração II – Um exemplo sobre anomalias “A simple exemple of anomalies in a 2-dimensional data set” [Chandola et al. 2009]	27
Ilustração III – Exemplo de identificação de anomalias contextuais em series [Chandola et al. 2009]	29
Ilustração IV – Representação gráfica de uma Tendência	35
Ilustração V – Representação gráfica da Sazonalidade	36
Ilustração VI – Observação de uma série temporal com previsões de origem t e horizontes de previsão iguais a um, dois e h [Moretin & Toloí. 2009].....	37
Ilustração VII – Exemplo de uma janela de previsão com cinco entradas e horizonte de previsão igual a um [Moretin & Toloí. 2009].....	38
Ilustração VIII – População diferente pela análise de variância [Acarini & Amaral, 1992]	50
Ilustração IX – Populações distinguidas pela análise multivariada (apenas duas variáveis) [Acarini & Amaral, 1992]	50
Ilustração X – Representação do agrupamento de dados com a identificação de pontos fora de seus domínios de actuação [Scikits, 2010].....	51
Ilustração XI – Gráfico comparativo entre o consumo BT, MT e BTE	54
Ilustração XII – CUSUM da BT relativamente aos seus valores de consumo ..	59
Ilustração XIII – CUSUM da MT reactivamente aos seus valores de consumo	61
Ilustração XIV– CUSUM da BTE relativamente aos seus valores de consumo	63
Ilustração XV – Aplicação do MA de janela 12 para o CUSUM da BT	65
Ilustração XVI – Aplicação do MA de janela 12 para o CUSUM da BT	67
Ilustração XVII – Aplicação do MA de janela 3 para o CUSUM da BT	68
Ilustração XVIII – Aplicação do MA de janela 12 para o CUSUM da MT	70
Ilustração XIX – Aplicação do MA de janela 6 para o CUSUM da MT	71
Ilustração XX – Aplicação do MA de janela 3 para o CUSUM da MT	72
Ilustração XXI – Aplicação do MA de janela 12 para o CUSUM da BTE.....	73
Ilustração XXII – Aplicação do MA de janela 6 para o CUSUM da BTE.....	73
Ilustração XXIII – Aplicação do MA de janela 3 para o CUSUM da BTE.....	74
Ilustração XXIV – Aplicação da Periodicidade para o CUSUM da BT.....	77
Ilustração XXV – Aplicação da Periodicidade para o CUSUM da MT	78
Ilustração XXVI – Aplicação da Periodicidade para o CUSUM da BTE	80

LISTA DE TABELAS

Tabela 1 - Consistência interna dos questionários [EDEL – Empresa de Distribuição de Electricidade, 2011]	19
Tabela 2 - Consistência interna dos questionários [EDEL – Empresa de Distribuição de Electricidade, 2011]	20
Tabela 3 - Consistência interna dos questionários [EDEL – Empresa de Distribuição de Electricidade, 2011]	20
Tabela 4 – Tabela comparativa entre Séries temporais e não temporais	32
Tabela 5 – Valores assumidos por uma única variável em série temporal.....	32
Tabela 6 – Tipos de Intervalos básicos como unidade das Séries Temporais [SAS Institute Inc, 2002]	40

LISTA DE ABREVIATURAS, SIGLAS, SÍMBOLOS E ACRÓNIMOS

BT	Baixa Tensão
BTE	Baixa Tensão Especial
Ct	Componente cíclica no tempo t
CUSUM	Cumulative Sum
EDEL	- Empresa de Distribuição de Electricidade
GWh	Giga Watt/Hora
H	Horizonte de previsão
ISEP	Instituto Para o Sector Empresarial Público
KWh	Kilo Watt/Hora
Kz	Kwanza
Kz/kWh	Kwanza por Kilo Watt/hora
MA	- Moving Average
MM	- Média Móvel
MT	Média Tensão
MW	Mega Watt
NumPy	Numerical Python
Rt	Componente aleatória no tempo t
St	Componente sazonal no tempo t
T	Índice temporal
Tt	Componente de tendência no tempo t
$\hat{Z}_t(h)$	Previsão no instante t + h
$\sum f(y)$	Somatório para uma função em y
$\sum e$	Somatório de erros

SUMÁRIO

1. Introdução	16
1.1. Enquadramento teórico.....	18
1.2. Problemas e motivações.....	22
1.3. Objectivos	23
1.4. Metodologia de investigação	24
1.5. Estrutura do trabalho	24
2. Detecção de eventos e anomalias	25
2.1. Conceitos gerais	25
2.2. Séries temporais.....	30
2.2.1. Classificação de séries temporais	33
2.2.2. Decomposição de series temporais.....	33
2.2.3. Previsão de series temporais	36
2.2.4. Periodicidade.....	38
2.3. Métodos de controlo estatístico para a detecção de mudanças	41
2.3.1. Algoritmo de soma acumulada	43
2.3.2. Algoritmo moving average	46
2.3.3. Interpretação de um resultado baseado em agrupamento dos dados	48
3. Resultados dos métodos para a detecção de mudanças e anomalias no consumo de energia eléctrica da cidade de Luanda	54
3.1. Análise dos gráficos de consumo	54
3.2. Descrição da variação dos dados.....	55
3.3. Detecção de mudanças e anomalias com o CUSUM	57
3.3.1. Descrição do CUSUM relativamente às mudanças no consumo	58
3.3.2. Aplicação do CUSUM no consumo de baixa tensão	58
3.3.3. Aplicação do CUSUM no consumo de média tensão	60
3.3.4. Aplicação do CUSUM no consumo de baixa tensão especial	62
3.4. Aplicação do moving average para filtrar o CUSUM.....	64
3.4.1. Moving average para filtrar o cusum na baixa tensão	65
3.4.2. Moving average para filtrar o CUSUM na média tensão	69
3.4.3. Moving average para filtrar o CUSUM na baixa tensão especial.....	72
3.5. Aplicação da periodicidade para o CUSUM.....	75
3.5.1. Periodicidade para o CUSUM de baixa tensão	76
3.5.2. Periodicidade de média tensão	78

3.5.3. Periodicidade para a baixa tensão especial	79
4. Conclusões	82
4.1. Conclusões gerais	82
4.2. Contribuição para o conhecimento científico	85
4.3. Limitações do estudo	86
4.4. Recomendações futuras	86
Anexos	94
Lista de anexos	95
Anexo A.....	96
Anexo B.....	98
Anexo C	100

1. INTRODUÇÃO

As técnicas para a análise de dados possibilitam determinar o comportamento em qualquer tipo de ambiente, desde que os dados sejam devidamente classificados. Assim sendo, para se detectar uma anomalia há sempre a necessidade de lidar com eventos separados e a natureza de um dado evento pode depender de algum outro evento. Entretanto é necessário perceber-se quando é que um determinado dado pertence a um determinado grupo e quais as principais características que o possibilitam pertencer. Para tal, é necessário no estudo a levar-se em causa, considerar os limites que separam um grupo do outro de modo a definirem-se critérios para a validação deste processo de segmentação. A verificação de anomalias pode não ser uma tarefa fácil dependendo das causas envolvidas, mas com as técnicas a serem usadas nesta dissertação, quando bem aplicadas a interpretação dos resultados acaba por ser simples.

A necessidade de se aprimorar métodos para a análise de dados tem crescido com o aumento no nível de complexidade entre as organizações. O que poderá ser considerado como base para a interpretação de uma ou mais informações tem a ver com a organização dos dados, logo é importante ter o domínio sobre os tipos de dados e como estruturá-los. As séries temporais acabam por ser uma metodologia que já possibilita ter uma estrutura de dados de consumo devidamente preparada para que a qualquer momento estes dados sejam inferidos por métodos especialistas.

O estudo do comportamento de series temporais acaba por ser muito importante, uma vez que se trata de um tipo de estrutura de dados que facilita o agrupamento de vários dados e a forma como estes se comportam a reacções naturais do tempo. Nos últimos anos tem havido um interesse crescente no uso de técnicas para se extrair interessantes padrões de sequências temporais com vista a inovação de vários sistemas [Guralnik & Jaideep, 1999]. De forma geral, ter sempre o controlo do tempo em qualquer actividade, permite alcançar com precisão diversos e interessantes objectivos. Como pudemos afirmar que os sistemas de monitorização estão sempre directamente relacionados ao controlo de um ou mais intervalos temporais.

Um exemplo evidente de uma série temporal poderá ser o controlo da temperatura de um corpo qualquer. Se a temperatura for acima de um certo limite, então poderá registar-se aquele instante de tempo como sendo um caso especial a ser estudado e o porquê das causas. Entretanto, foi escolhido efectuar um estudo destes comportamentos

através de algoritmos estatísticos de controlo, sendo que poderá detectar-se mudanças sempre que algum ponto se deslocar alcançando um determinado limite estabelecido. Para a aplicação deste algoritmo, a aplicação das séries temporais neste trabalho, permitirá optar-se por um modelo que poderá adequar-se de modo a dar-nos a realidade estruturada relacionada com o consumo de energia, tendo em conta o domínio das suas principais características e propriedades. A segmentação destas características e propriedades poderá determinar um método que nos permita solucionarmos o problema.

O Algoritmo CUSUM, proposto por E. S. Page em 1954, com o principal propósito de detecção de mudanças, tem seu potencial no facto de agregar desvios em várias etapas de tempo para posteriormente detectar mudanças na média do processo em causa, enquanto o *Moving average* poderá ser aplicado com o intuito de ser possível a criação de um registo de médias geradas à medida que se percorre uma sequência de dados, ou série temporal, possibilitando a eliminação de componentes como a Sazonalidade, Ciclo, etc. Atenção que dependendo do problema existe a necessidade de haver indicações dos vários limites para o *Moving Average*. A forma como encaramos o *Moving Average*, possibilita que possamos resolver problemas de acordo com a especificidade, uma vez que este algoritmo pode facilmente adaptar-se a diferentes realidades, sendo que series temporais aparentemente complexas, facilmente podem ser agrupadas em vários subgrupos. Já a aplicação da Periodicidade, poderá ser relevante pelo facto dos dados de consumo ao longo dos anos terem sempre uma característica periódica, e ao longo da dissertação poderá interpretar-se melhor estas características e suas possíveis causas. De acordo com [Hunter & John, 2007], teremos como suporte uma linguagem de programação muito usada para soluções em Computação Científica, o Python. Pois tem-se desenvolvido muito com o surgimento de bibliotecas bastante poderosas no intuito de haver a apresentação de soluções completas para o mundo científico. Alguns dos métodos aplicados para a análise dos dados e interpretação de suas mudanças têm como dependência bibliotecas e extensões como:

Numpy

É um pacote para a linguagem Python que suporta arrays e matrizes multidimensionais, possuindo uma larga colecção de funções matemáticas para trabalhar com estas estruturas.

Matplotlib

É uma biblioteca de “*plotagem*” para a linguagem de programação Python e sua extensão matemática numérica NumPy.

Math

É uma biblioteca que possibilita o acesso a funções matemáticas tais como, Aritméticas, Trigonométricas, Exponenciais, Logarítmicas, etc.

1.1. ENQUADRAMENTO TEÓRICO

Para o enquadramento deste estudo, primeiramente será apresentada a empresa de distribuição de electricidade de Luanda, de modo a perceber-se alguns objectivos que têm sido cumpridos.

A EDEL-EP é uma unidade económica de interesse público e de grande dimensão, dotada de personalidade jurídica e de autonomia administrativa, financeira, patrimonial e de gestão. Como pessoa colectiva se rege pelos princípios definidos no seu estatuto, nas normas complementares de execução, nas demais disposições consagradas na Lei e, no que não estiver especialmente regulado, pelas normas de direito vigentes em Angola. O objecto social da EDEL-EP consubstancia-se na distribuição e comercialização de energia eléctrica, tendo como mercado principal a província de Luanda e, recentemente, alguns municípios da província do Bengo. Neste contexto e através dos principais indicadores de gestão, são balanceadas neste documento as actividades mais relevantes desenvolvidas pela empresa durante o ano em análise.

Contrariamente ao verificado no primeiro ano do mandato do Conselho de Administração, nomeado em Novembro de 2009, em que se registou a ausência de orientações estratégicas, a Empresa consolidou no período em análise o seu funcionamento com base no Plano Estratégico-2010/2012.

Neste âmbito, o valor pecuniário do consumo da Energia Facturada foi de Kz 9.241.408.504,62, enquanto o valor pecuniário do consumo da Energia Cobrada foi de Kz 7.227.197.668,35, representando 78% do facturado no ano em análise. Relativamente aos serviços de exploração da rede, continuaram as restrições na

distribuição, com incidência particular para os meses mais quentes, devido às limitações na rede de transporte e ao aumento das cargas em função das necessidades de refrigeração. A quantidade de Energia Adquirida foi de 2.825,96 GWh, apesar dos 15% de Perdas Técnicas admitidas, entretanto as Energias Facturada e Cobrada representam 85% e 67%, respectivamente, da Energia Adquirida no ano em análise.

No domínio dos investimentos, destacam-se a recepção dos 89 postos de transformação, que possibilitaram a ligação à rede eléctrica de 68.240 agregados familiares.

A gestão dos recursos financeiros da Empresa beneficiou ao longo do ano da recepção dos Subsídios a Preços, atribuídos pelo ISEP (Instituto para o Sector Empresarial Público), recebendo um total de Kz 8.910.331.185,68. Entretanto, apesar do alcance de um resultado de exploração positivo, verificou-se no ano em análise que a falta de actualização da tarifa, i.e. 3,40 Kz/kWh, tem provocado um agravamento do hiato entre o seu valor nominal e o seu custo real.

A execução das acções atrás resumidas, integradas nas linhas de orientação do plano de actividades permite fazer um balanço positivo dos resultados obtidos pela Empresa.

A seguir o quadro, referente ao processo comparativo de contratação de novos clientes ao longo do triénio:

Tabela 1 - Consistência interna dos questionários [EDEL – Empresa de Distribuição de Electricidade, 2011]

Leituras	Ano 2009	Ano 2010	Ano 2011
Previstas	854.644	725.051	771.270
Inseridas	683.369	468.578	545.967
Rácio <small>Inser/Prev</small>	80%	65%	71%

Na tabela seguinte constata-se que no ano em análise foram ligados mais 68.240 novos clientes, dos quais 99,1% a nível da BT e 0,9% a nível da MT. Portanto, este aumento resulta das acções desenvolvidas, principalmente, nas zonas que beneficiaram de melhoria e reforço da capacidade instalada.

Tabela 2 - Consistência interna dos questionários [EDEL – Empresa de Distribuição de Electricidade, 2011]

Cientes Novos	Ano 2009	Ano 2010	Ano 2011
Baixa Tensão	33.034	53.225	67.630
Media Tensão	66	110	610
Total	33.100	53.335	68.240

Destaca-se ainda que neste ano os 204.302 clientes que estavam sob a gestão dos Agentes Autorizados, representando 42,4% da clientela da EDEL em BT, enquanto no ano anterior foram 155.929, representando 49,8%, e no ano precedente foram 134.132, representando 51,4% destes. A seguir o gráfico com o total de clientes activos, controlados no triénio:

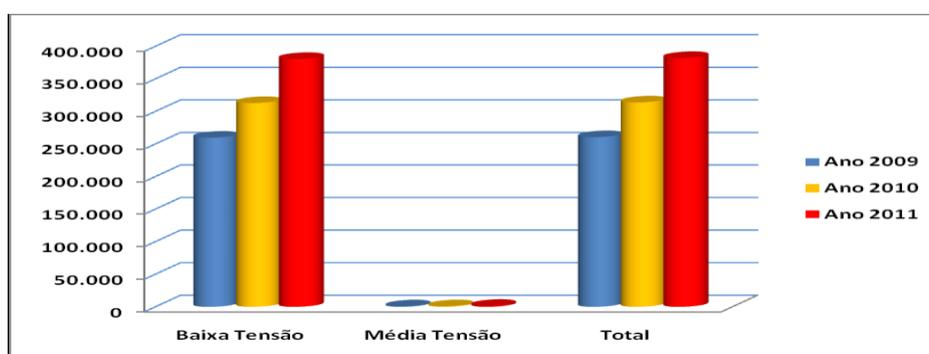


Ilustração I – Total de clientes activos 2009 – 2011 [EDEL – Empresa de Distribuição de Electricidade, 2011]

Tabela 3 - Consistência interna dos questionários [EDEL – Empresa de Distribuição de Electricidade, 2011]

Centros de Distribuição	Clientela por Tipo Contrato			Análise Vertical
	Normais	Avençados	Total	
Ingombotas	22.183	1.561	23.744	6,2%
Maianga	53.898	7.143	61.041	16,0%
Sambizanga	28.720	5.922	34.642	9,1%
Cazenga	59.863	0	59.863	15,7%
Kilamba Kiáxi	67.011	16.596	83.607	21,9%
Viana	46.733	0	46.733	12,2%
Rangel	12.404	3.059	15.463	4,0%
Samba	20.477	764	21.241	5,6%
Cacuaco	20.862	1	20.863	5,5%
Bengo	3.462	0	3.462	0,9%
DCMT	1.743	117	1.860	0,5%
PEPP	9.629	0	9.629	2,5%
Total	346.985	35.163	382.148	100,0%

No quadro anterior, onde se pode observar a segmentação da clientela por tipo de contracto em cada uma das unidades de distribuição e comercialização, constata-se

que 90,8% dos clientes têm contractos em estado Normal, enquanto 9,2% têm contractos em estado de Avença, i.e. sem contador. Destacam-se os Centros de Distribuição do Cazenga, Viana, e Bengo que não têm clientes com contratos em estado de Avença.

A distribuição do consumo de energia eléctrica para a cidade de Luanda é uma área bastante crítica, pois que vive grandes problemas na satisfação dos clientes e as infra-estruturas muitas das vezes não apresentam muitas condições dificultando as mudanças inovadoras. As ferramentas que possibilitam este tipo de análises podem ser propostas estrategicamente em várias áreas que poderão possibilitar o desenvolvimento para a satisfação de clientes ou consumidores poupando muito esforço de recolhas constantes de dados. Com esta prática poderá ser possível segmentar vários tipos de consumidores e perceber-se os seus comportamentos em diferentes intervalos temporais, permitindo descrever a origem destes efeitos sazonais e procurar mantê-los sob controlo com o intuito de sugerir melhorias à medida que possam verificar-se falhas. Em [Guralnik e Jaideep, 1999], afirma-se que nos últimos anos tem havido um interesse crescente no uso de técnicas para se extrair padrões de sequências temporais com vista à inovação de vários sistemas.

Para a realização desta tese de dissertação, os dados analisados foram disponibilizados pela EDEL, uma vez ter-se constatado que o crescimento no interesse de métodos de controlo de estatístico está em alta com o surgimento de bibliotecas específicas em linguagens de programação viradas a computação científica. Neste sentido foram aplicados algoritmos de detecção de anomalias com uma forte base proveniente das ciências estatísticas.

O estudo deste trabalho surge numa fase interessante uma vez que os dados do consumo de energia eléctrica de Luanda poderão apresentar variações incontroláveis, entretanto o tratamento devido dos dados poderá merecer maior atenção de modos a que os métodos aplicados poderão tornar-se mais precisos. Luanda regista um crescimento acelerado em algumas áreas que podem estar a dificultar algumas automatizações.

1.2. PROBLEMAS E MOTIVAÇÕES

Acaba em ser uma grande motivação o facto de a economia estar a crescer e o consumo de energia poder ser visto como uma variável que poderá fornecer boas indicações e que varia bastante no modo como é distribuída para a realidade de Luanda.

A projecção de uma série temporal resulta ser uma metodologia interessante de modo a que métodos computacionais tenham acesso e possam extrair inteligentemente padrões que podem ser estudados com o intuito de prever e evitar diversas situações.

Este estudo tem relação com uma variável que poderá estar a sofrer muitas influências com as transformações que Luanda tem registado. Esta variável pode ser observada num modelo tratado como a função básica do consumo proposta por Keynes, assim sendo, a renda poderá estar a sofrer muitas mudanças e de acordo a esta teoria, esta é directamente proporcional ao consumo. O facto de a renda ser um valor que varia bastante na economia em Luanda, excelentes resultados poder-se-ão obter e haverá um grande contributo com as conclusões retiradas nesta dissertação.

Keynes afirmou que os homens estão dispostos, como regra e em média, a aumentar seu consumo conforme sua renda aumenta, mas não na mesma proporção que o aumento na renda, logo poderá ser interessante espelhar-se o comportamento do consumo de dados aplicados relativamente ao modo de vida populacional e perceber-se até que ponto o seu comportamento poderá derivar certas características algorítmicas a serem observadas.

Os processos que envolvem sistemas reais são sempre tidos como tão complexos que mesmo havendo uma forma analítica de os resolver, poderá ser mais produtivo optar pelo uso de teorias probabilísticas, como tal, este projecto poderá fornecer bases para uma relação entre automatismo e formalismo probabilísticos numa base dedutiva.

Em computação, qualquer sistema formal representa um sistema computável, entretanto iremos verificar que os formalismos probabilísticos terão um papel importante na melhoria do desempenho do sistema que se pretende construir. Irá observar-se que a probabilidade da ocorrência de um intervalo de tempo poderá implicar uma certa mudança ou não no sistema.

A distribuição de energia eléctrica poderá atingir situações negativas críticas pelo facto de não serem identificadas situações críticas que aconteceram num determinado instante de tempo. Assim sendo, este estudo vem relacionar os consumidores, interpretar resultados provenientes de seus comportamentos e possibilitar que sejam aplicados em resultado de suas interacções.

1.3. OBJECTIVOS

Este estudo tem como objectivo a análise de dados de consumo de energia eléctrica da cidade de Luanda do ano de 2009 a 2012, com vista à observação do comportamento das variações relativamente as médias de consumo e perceber as consequências de deixar passar as grandes variações que poderão ser consideradas como anomalias. Perceber a necessidade de tratar-se dados a parte (por grupos), suas vantagens para uma determinada população, visto que poderão existir fenómenos como a Sazonalidade, logo a necessidade de se pôr em prática o reajustamento a medida que os eventos necessitam de se adaptar a determinadas realidades.

Entretanto, vale resumir os passos para este estudo no seguinte:

- Estudo de conceitos sobre anomalia segundo a abordagem de vários autores;
- Estudo das Series temporais, necessidade de sua utilização, identificação de suas principais características e necessidade de sua inclusão para o consumo de energia eléctrica;
- Estudo de métodos de controlo estatísticos e aplicação destes ao consumo de energia denotado como uma série temporal para a verificação de mudanças e anomalias;
- Concluir os métodos que melhor poderão adequar-se na detecção de anomalias no consumo da energia eléctrica da cidade de Luanda.

1.4. METODOLOGIA DE INVESTIGAÇÃO

Para esta Dissertação, os dados são de consumo de energia eléctrica da cidade de Luanda e a partir da empresa EDEL (Empresa de Distribuição de Electricidade de Luanda) e com uma certa aprovação.

O algoritmo CUSUM é o método principal para gerar-se as mudanças e anomalias, pois trata-se de um conjunto de procedimentos que acha os resultados de acordo a uma análise estatística. Esta metodologia é específica para detectar-se mudanças em dados e pode ser aplicada em muitas soluções em Ciências da Computação

Uma vez os dados recolhidos e observado o seu tratamento, para sua a organização, foram estudados métodos para avaliar qual deles seria o certo para a realidade do problema, entretanto notou-se que o CUSUM (Soma acumulada), tendências negativas e positivas ao longo de um período, mas este resultado pode ser difícil de ser interpretado ou mal interpretado, assim sendo, o *Moving Average* é um método específico de filtragem dentro do estudo de interpretação de Series Temporais de modos a evitar que as médias não desloquem-se muito de seus valores. Desta forma o CUSUM em concordância com o *Moving Average* poderá apresentar um resultado mais legível. Também foi estudado a periodicidade como uma característica que os dados de séries temporais transportam, e aproveitando-se desta decidiu-se percorrer os dados recorrendo a intervalos periódicos de modos a retirar-se alguns factores desnecessários.

1.5. ESTRUTURA DO TRABALHO

Este estudo está constituído por quatro capítulos e encontra-se estruturado da seguinte forma:

Capítulo I

No primeiro capítulo é feita uma introdução sobre o crescimento no interesse de extrair-se conhecimento e padrões em series temporais para a contribuição no processo de inovação em diversas áreas da ciência. A relevância de controlar-se o tempo para a garantia de determinados objectivos e de inserção do conceito de algoritmos para o controlo estatístico para detecção de mudanças em séries temporais. Assim sendo, introduziram-se alguns caminhos sobre a forma como os algoritmos poderão relacionar-

se para a solução do problema. Por último a necessidade de aplicar-se esta metodologia para os dados de consumo de energia eléctrica.

Capítulo II

Para o segundo capítulo foi reservado o estudo sobre o estado da arte e revisão de literatura sobre detecção de eventos e pontos anômalos, “*Change-point and Event-Detection*”, assim como a apresentação dos métodos a serem utilizados para se chegar aos resultados. Aqui a preocupação é com as boas práticas para a verificação e validação de métodos específicos nas diversas áreas e a proposta de aplicação dessas abordagens para a análise de dados no seu sentido global.

Capítulo III

O capítulo três foi reservado somente para a interpretação dos resultados gerados da metodologia aplicada aos dados, assim como observar se estes realmente têm alguma relação com a realidade da cidade de Luanda.

Capítulo IV

Capítulo dedicado à apresentação das conclusões do trabalho, as limitações com as quais nos deparámos ao longo do estudo, contribuição do trabalho para o mundo científico e apresentação como proposta para investigações futuras.

2. DETECÇÃO DE EVENTOS E ANOMALIAS

2.1. CONCEITOS GERAIS

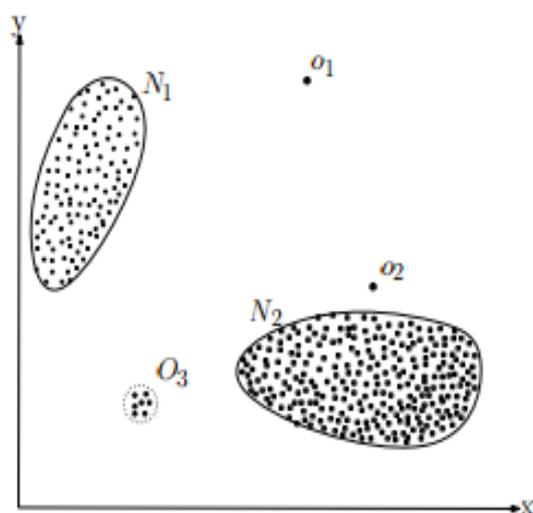
Uma tarefa importante na exploração de dados sobre fenómenos e processos que se desenvolvem ao longo do tempo é a detecção de mudanças significativas, que também podem ser consideradas como anomalias, que acontecem num dado fenómeno estudado.

De acordo com [Hwang, 2011], um grande foco no estudo em detecção de eventos é o facto de permitir o desenvolvimento de novos métodos estatísticos e computacionais para a detecção precoce e precisa dos eventos e padrões emergentes, pois esta prática

viabiliza descobertas e novas práticas de acção. Entretanto, pode-se afirmar que o estado actual desta área apresenta-se num ponto alto com o desenvolvimento de novas ferramentas da computação.

Para [Chandola et. al., 2009], a detecção de anomalias refere-se ao problema de encontrar padrões em dados que não se encontram em conformidade em relação ao comportamento esperado. Entretanto estes demonstram também que as anomalias são padrões em dados que não se conformam com uma noção bem definida de comportamento normal, logo graficamente podem ser vistos como casos isolados que requerem um acompanhamento de modos a prever melhorias a um sistema. Assim sendo, poderá entender-se como detecção de anomalia, um conjunto de métodos que podem servir de maneira propositada o acompanhamento de um processo de modos a que a trajectória de seus dados possam ser devidamente descrita e caracterizada.

O grande objectivo de detectar anomalias é perceber a necessidade de tratar dados a parte, suas vantagens para um determinado grupo de dados *Cluster*¹ ou agrupamento, visto que poderão existir fenómenos como a Sazonalidade. Deste modo dever-se-á pôr em prática o reajustamento a medida que os eventos necessitarem adaptar a diferentes realidades.



¹Neste contexto, é visto como áreas a serem detectadas consideradas como regiões anómalas ou de interesse para pesquisa sendo que os componentes representantes cooperam entre si na realização de uma tarefa.

Ilustração II – Um exemplo sobre anomalias “A simple exemple of anomalies in a 2-dimensional data set” [Chandola et al. 2009]

Em [Wong & Neill, 2009], diz-se que a área de detecção de anomalias permite monitorar e evitar situações críticas, assim em muitas tarefas do mundo real tais como vigilância, descoberta científica e limpeza de dados envolvem monitorização de dados recolhidos rotineiramente.

Em [Kawahara & Masashi, 2009], relacionou-se o conceito de anomalias às séries temporais, e considerou que neste caso se trata de um problema que consiste na descoberta de pontos de tempo nas propriedades de mudanças de dados em series temporais.

“Change-point detection is the problem of discovering time points at which properties of time-series data change” [Kawahara & Masashi, 2009].

Em [Das & Schneider, 2007], a forma tradicional de detecção de anomalia está no controlo de processos industriais. Dados de séries temporais de vários sensores são monitorados para a detecção de processos fora de controlo.

Em [Kawahara et.al., 2007] afirma-se que mudanças abruptas das propriedades de séries temporais, muitas vezes contêm informações extremamente importantes de várias perspectivas, e, portanto, o problema de descobrir os momentos em que as mudanças ocorrem, são chamados de pontos de mudança ou situação ou situações anormais, e esta área tem recebido muita atenção em estudos científicos para a percepção de diversos fenómenos.

Em [Guralnik et. al., 1999] afirma-se que o problema da detecção de eventos passa a ser o reconhecimento da mudança dos parâmetros do modelo, ou talvez mesmo a alteração do próprio modelo, num momento desconhecido, entretanto a identificação deste momento poderá ser visto como um caso especial de estudo. Mas um aspecto importante da técnica de detecção de uma anomalia é a natureza da anomalia.

Em [Zaperlão, 2009], estudou-se sobre detecção de anomalias e aplicabilidade em redes de computadores, entretanto definiu-se como a presença de desvios súbitos e acentuados que ocorrem no tráfego em consequência de diversas situações como

defeitos em *softwares*, uso abusivo de recursos da rede, falhas em equipamentos, erros em configurações e ataques. A detecção por anomalia ou por comportamento (*anomaly detection*), diferente da detecção por abuso, compara os eventos correntes do sistema ou do tráfego de rede a perfis de actividades normais, ou seja, aos modelos de detecção por anomalia. Nesta abordagem, as anomalias são consideradas como possíveis intrusões.

Em [Chandola, 2009], anomalias podem ser classificadas em quatro categorias seguintes:

- **Anomalias em pontos:** Se um exemplo individual de dados pode ser considerada como anormal, com respeito ao resto dos dados, então o exemplo é denominado como um ponto de anomalia. Este é o tipo mais simples de anomalia e é o foco da maioria das pesquisas sobre detecção de anomalias. Por exemplo, na Ilustração II, os pontos O1 e O2 assim como pontos do O3 região encontram-se fora dos limites das regiões normais e, portanto, são anomalias do ponto, uma vez que são diferentes a partir de pontos de dados normais;
- **Anomalias contextuais:** Se uma instância de dados é normal num contexto específico (mas não de outro modo), em seguida, ele é designado como uma anomalia contextual (também referida como anomalia condicional [Song *et al.* 2007]). A noção de um contexto é induzida pela estrutura do conjunto de dados e tem de ser especificada como parte da formulação do problema. Cada instância de dados é definida usando dois conjuntos de atributos a seguir;
- **Atributos contextuais:** Os atributos do contexto são utilizados para determinar o contexto dessa instância. Por exemplo, em conjuntos de dados geográficos, a longitude e latitude de um local são os atributos do contexto. Em dados de séries temporais, o tempo é um atributo contextual que determina a posição de uma instância em toda a sequência;
- **Atributos comportamentais:** Os atributos comportamentais definem as características não contextuais de uma instância.

O comportamento anómalo é determinado usando os valores para os atributos comportamentais dentro de um contexto específico. Esta propriedade é fundamental na

identificação de atributos contextuais e comportamentais para uma técnica de detecção de anomalia contextual.

Anomalias contextuais têm sido mais comumente exploradas em dados de séries temporais [Weigend et al. 1995; Salvador & Chan 2003] e dados espaciais [Kou et al. 2006; Shekhar et al. 2001].

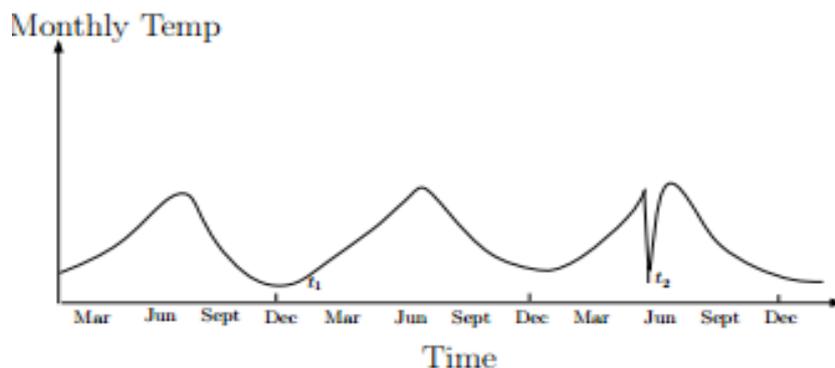


Ilustração III – Exemplo de identificação de anomalias contextuais em series [Chandola et al. 2009]

"Note that the temperature at time t_1 is same as that at time t_2 but occurs in a different context and hence is not considered as an anomaly" [Chandola et al. 2009]

A média móvel (*Moving Average*) para as séries temporais para a detecção de variações significantes. A variação local dos sinais de alta e média banda é calculada ao longo de um determinado intervalo de tempo por uma janela deslizante. Uma métrica chamada desvio do arranjo, que é a relação entre a variação local dentro da janela e a variância global, é calculada para cada uma dessas janelas. Desta forma a média deslizante ao ser aplicada para um conjunto dados em serie temporal, facilmente adapta-se para cada realidade e pode dar resposta esperada para cada caso.

Para o consumo de energia eléctrica, padrões a serem observados poderão contribuir de alguma forma para processos de inovação dos sistemas envolvidos, uma vez que o processo de monitoração envolve várias amostras e a observação da sequência de seus registos possibilita verificar e validar o estado de suas variações mediante métodos de controlo. Podemos realçar que as várias fases a serem desenvolvidas na apuração dos dados de consumo permitirão a obtenção de certas conclusões com precisão e criação de várias possibilidades na melhoria da qualidade e modelagem dos vários sistemas integrados.

Adquirir informações relevantes sobre eventos de um determinado ambiente pode ser uma actividade fundamental e necessária para a criação de vantagens competitivas. De entre os factores ambientais que podem vir a merecer especial atenção destacam-se as inovações tecnológicas devido ao elevado impacto que podem proporcionar em diferentes organismos. De forma geral, um processo de monitoração de dados com vista a inovação entre organismos sugere que os menos relevantes podem enfrentar desafios específicos com vista a alcançar também grandes objectivos.

Assim, os objectivos que se poderão alcançar com esta abordagem podem ser resumidos no seguinte:

- A análise de algum impacto que situações fora de comum de um processo de monitoração poderão vir a causar internamente em determinados organismos ou níveis sociais;
- O aprimoramento de métodos específicos que possibilitam detectar os pontos específicos causadores de tais situações fora de comum, tais como, os métodos estatísticos a serem estudados na análise do processo deste estudo;
- A descrição rigorosa dos algoritmos que contêm em sua estrutura tais métodos, suas funcionalidades e a aplicação para o objectivo deste estudo;
- A percepção da possibilidade de melhoria dos processos e o impacto geral de inovação para diferentes sistemas;
- A sugestão de padrões específicos com vista a contribuir em processos de inovação.

2.2. SÉRIES TEMPORAIS

Uma série temporal representa uma colecção de valores obtidos a partir de medições sequenciais ao longo do tempo [Esling & Carlos, 2012].

Para [Hwang, 2011], uma série temporal é muitas vezes o resultado da observação de um processo subjacente no decurso da qual os valores são recolhidos a partir de medições feitas em instantes espaçados uniformemente e de acordo com uma taxa de

amostragem de dados. De acordo a definição, os dados de uma série temporal podem ser colectados em intervalos regulares, como diariamente (preços das acções, relatórios meteorológicos), semanalmente (informações como oferta de moeda), mensalmente (taxa de desemprego), etc. Para, [Hwang, e Seokyon 2011] em estatística, econometria, matemática aplicada e processamento de sinais, uma série temporal é uma colecção de observações feitas sequencialmente ao longo do tempo e entretanto as séries temporais existem nas mais variadas áreas de aplicação, como: Ciências económicas, ciências médicas, ciências naturais e engenharias. Graças ao advento da computação, actualmente os dados podem ser recolhidos a intervalos extremamente curtos, obtidos de forma prática e contínua.

Formalmente as séries temporais, são muitas vezes representadas por meio de funções matemáticas, ou seja, assume-se que o valor obtido é função de alguma outra variável (ou de diversas variáveis), ou, o que é a mesma coisa, que existe uma lei de formação que determina esta série temporal, em que a sua forma T é uma sequência de valores n , em que:

$$T = (t_1, t_2, t_3 \dots t_n)$$

A função que determina a série temporal não precisa ser sempre linear. Na verdade, ela pode ter qualquer formato (quadrática, exponencial...) e pode depender de mais de uma variável. Dependentemente da função, poderemos chegar ao objectivo de detectar-se interferências denominadas como anomalias. Estas interferências, poderão ser detectadas pelo funcionamento de um determinado algoritmo para a realidade em causa.

A detecção de anomalias em análise de séries temporais é um problema importante, pois a presença de até mesmo alguns dados anómalos podem levar à má especificação do modelo, a estimativa de parâmetros tendenciosa, e a previsões pobres. [Galeano et. al. 2006]

Uma característica muito importante deste tipo de dados é que as observações vizinhas são dependentes e o interesse é analisar e modelar esta dependência.

Mediante as definições seguintes pode-se distinguir uma série temporal de uma série não temporal:

Tabela 4 – Tabela comparativa entre Séries temporais e não temporais

Séries temporais	Séries não temporais
Uma série que mostra a temperatura diária de uma cidade, sendo que o primeiro dado representa o primeiro dia e assim por diante. $T = \{14, 17, 18, 17, 16\}$	A temperatura diária de várias cidades.
Números de homicídios anuais em determinado país. $H = \{190, 123, 160, 201\}$	Números de homicídios anuais em diferentes países
As médias mensais do consumo de água de uma determinada região.	As médias mensais do consumo de água de regiões diferentes regiões.

De acordo com [Hwang, 2011] e tal como pode ser verificado na tabela acima, uma série temporal é geralmente representada por uma variável que assume diferentes valores em diferentes momentos do tempo, e por isso utiliza-se um subscrito junto à letra para denotar o período a que o valor específico (realização) se refere. Por exemplo, se a variável "consumo de electricidade de Luanda de todos meses de Maio entre 2009 a 2012" for representada por C , podemos denotar como os valores mensais de consumo $c_1, c_2, c_3 \dots c_n$. Uma vez modelados os dados em serie temporal, será possível a identificação de existência do factor de sazonalidade, ciclo ou tendência. Mediante este conceito poderemos obter a seguinte tabela como resultado:

Tabela 5 – Valores assumidos por uma única variável em série temporal

Período em análise	Representação do ano na série temporal	Valor assumido pela variável C em cada período
2009	2009 = 0 (ano inicial)	102.053.841,90
2010	2010 = 1	102.158.433,70
2011	2011 = 2	102.576.142,50
2012	2012 = 3	124.427.839,10

Assim sendo, C representa a série temporal que armazena os valores médios de consumo do ano de 2009 à 2012.

Embora as series temporais sejam uma prática muito usada, a maioria dos estudos com este tipo de dado pressupõem que a serie seja estacionária. Implica que uma série temporal será bem interpretada se seus valores andarem em torno de uma média.

A análise de séries temporais fornece ferramentas para a selecção de um modelo que pode ser usado para previsão de eventos futuros. A modelagem da série temporal é um problema estatístico. As previsões são utilizadas em procedimentos computacionais para estimar os parâmetros de um modelo a ser usado para os recursos limitados atribuídos ou ao descrever processos aleatórios. Modelos de séries temporais assumem que as observações variam de acordo com alguma distribuição de probabilidade sobre uma função subjacente de tempo.

2.2.1. CLASSIFICAÇÃO DE SÉRIES TEMPORAIS

Série estacionária

Uma série estacionária é o que em matemática costuma chamar-se série convergente, ou seja, aquela que flutua em torno de uma mesma média ao longo do tempo. Para [Gujarati & Porter 2008], uma serie é estacionária se a sua média e variância não variam substancialmente ao longo do tempo.

O processo estacionário é o mais comum e basicamente isso significa que o comportamento da série não se altera com o passar do tempo.

Série não estacionária

A série não estacionária, ou função divergente em matemática, é aquela que tem raiz unitária. A tendência de uma serie não estacionaria é infinita, ou seja, que não existe um valor único para média, pois este encontra-se em constante oscilação. Um exemplo real para este comportamento, tendo em conta a realidade desse trabalho, poderá dar-se na mudança constante do comportamento do consumidor no consumo de energia e uma técnica de controlo poderá ser a implementação do controlo móvel da média.

2.2.2. DECOMPOSIÇÃO DE SERIES TEMPORAIS

[Makridakis et al. 1989] afirma que vários procedimentos para a decomposição de séries temporais foram desenvolvidos, cada qual tentando isolar as componentes não-

observáveis da série o mais fielmente possível. O objectivo desses procedimentos consiste em remover cada uma das componentes, permitindo que o comportamento da série temporal seja melhor compreendido e, conseqüentemente, prognosticar valores futuros mais apropriados.

De acordo as evidências apresentadas, a exclusão de uma das componentes poderá implicar a exclusão de uma ou de um conjunto de anomalias que dados poderão transportar.

Em [Cardoso, 2001], é afirmado que no acto de analisar-se uma série temporal, primeiramente deve-se modelar o fenómeno estudado para, a partir daí, descrever-se o comportamento da série, fazer estimativas, e por último, avaliar quais os factores que influenciaram o comportamento da série, buscando definir relações de causa e efeito entre duas ou mais séries.

No acto de decomposição, uma variável Y denominada série temporal no tempo t pode ser modelada como sendo o produto de ou segmento de quatro variáveis, ou seja, $Yt = Tt \times Ct \times St \times Rt$. Segundo [Box & Reinsel, 1970], Tt significa componente de tendência no tempo t , Ct significa componente cíclica no tempo t , St significa componente sazonal no período t e Rt significa componente aleatório no tempo t . Como uma alternativa, o modelo de aditivo [Balestra & Nerlove, 1966] também pode ser utilizado, no qual $Yt = Tt + Ct + St + Rt$ onde Tt, Ct, St, Rt têm o mesmo significado como descrito acima.

Assim sendo e de acordo as evidências apresentadas em [Box & Reinsel, 1970], segundo [Cardoso, 2001], uma serie temporal pode ser geralmente decomposta em:

Tendência

Capta elementos de longo prazo relacionados com a série de tempo. Tendência, no contexto de séries temporais, pode ser definida como uma mudança constante de nível. Uma dificuldade na definição de tendência consiste em encontrar um período ideal que prove existência de uma tendência e não outro efeito como sazonalidade [Chatfield, 1982]. Por exemplo, variações climáticas, às vezes, exibem variações cíclicas durante

um longo período, como 50 anos. Se para essas variações for disponibilizada uma série que a represente por um período de apenas 20 anos, essa oscilação poderia parecer uma tendência, porém se várias centenas de anos forem disponíveis, a oscilação seria visível. A figura 5 mostra um exemplo de tendência.

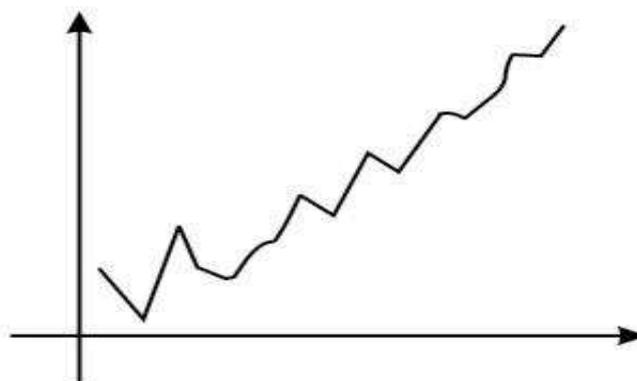


Ilustração IV – Representação gráfica de uma Tendência

Ciclo

Longas ondas, mais ou menos regulares, em torno de uma linha de tendência, ou seja, os ciclos são caracterizados pelas oscilações de subida e de queda nas séries, de forma suave e repetida, ao longo da componente de tendência. Por exemplo, ciclos relacionados à actividade económica ou ciclos meteorológicos.

Sazonalidade

Capta os padrões regulares da série de tempo. Muitos exemplos de séries temporais como, séries de vendas e leituras de temperaturas, exibem variações que são periódicas. Por exemplo, a venda de protector solar cresce consideravelmente no nordeste durante os meses de Janeiro e Fevereiro. Esse tipo de variação é de fácil entendimento e pode ser facilmente removida para a obtenção de uma série livre de sazonalidade. A figura 4 mostra um exemplo de série sazonal.

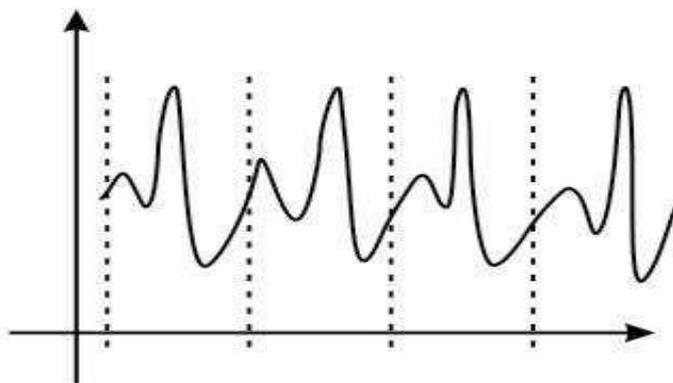


Ilustração V – Representação gráfica da Sazonalidade

Aleatória

Capta todos os efeitos que não foram incorporados pela série de tempo via os três componentes anteriormente citados, ou seja, é o resíduo.

Segundo [Morettin, 1981], a componente sazonal representa as flutuações da série de acordo com algum factor de sazonalidade. O ciclo apresenta um comportamento similar à componente sazonal, embora tenha normalmente comprimento maior que aquela. Justamente pelo facto de não apresentar duração uniforme, a identificação da componente ciclo é mais problemática. A tendência representa o aumento ou declínio gradual nos valores das observações de uma série temporal. Com a remoção das componentes de sazonalidade, ciclo e tendência, a componente aleatória fica determinada [Makridakis et. al. 1989].

Para [Hwang e Seokyon, 2011] se uma série contém uma sazonalidade ou uma tendência, então a série deve ser transformada de não-estacionário para uma série estacionária, tratando-se de remover as componentes. Por outro lado, [Martins, 2001] também apresenta evidências que ao analisar uma série temporal, deve-se estudar cada um destes componentes separadamente, retirando-se o efeito que outros componentes transportam. Esta afirmação poderá e então permitir a liberdade de analisar-se uma única componente e derivadamente surgirem conclusões para o domínio a ser estudado.

2.2.3. PREVISÃO DE SERIES TEMPORAIS

Para a previsão é necessário perceber que uma série temporal pode ser conceituada simplesmente como qualquer conjunto de observações ordenadas no tempo [Morettin & Toloi, 1981].

Para estas observações o grande foco será o facto de caracterizar-se futuros objectos dependentemente de resultados presentes. Neste caso poder-se-á expressar uma série temporal por:

$$Z_t = \{Z_t \in \mathbb{R} \mid t = 1, 2, 3 \dots N\}$$

Onde t é um índice temporal, e N é o número de observações. Considerando a existência de observações de uma série temporal até o instante t , a previsão no instante $t + h$ é denotada por $\hat{Z}_t(h)$, cuja origem é t e o horizonte de previsão é h . As previsões em $\hat{Z}_t(t + 1)$, $Z(t + 2)$, ... $\hat{Z}_t(t + h)$ $Z(t+1)$, $Z(t+2)$, podem ser vistas mediante a ilustração VI.

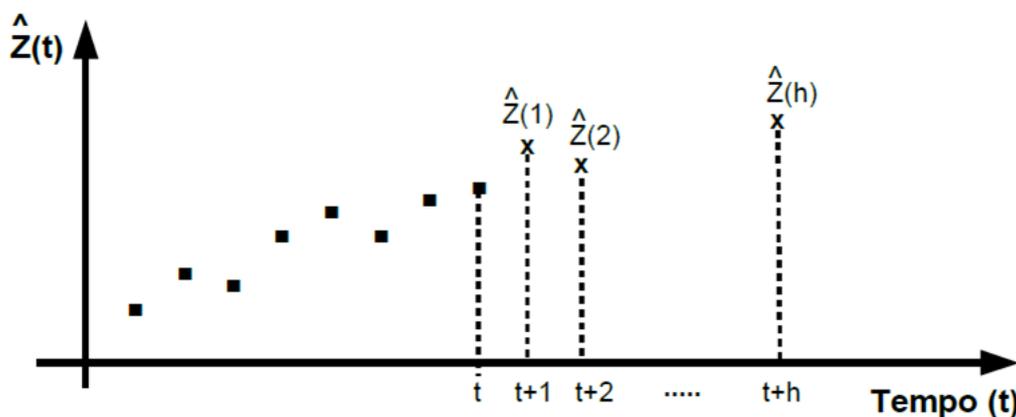


Ilustração VI – Observação de uma série temporal com previsões de origem t e horizontes de previsão iguais a um, dois e h [Moretin & Toloi. 2009]

Além do horizonte de previsão, outro parâmetro usado pelo processo de previsão é o número de elementos históricos anteriores ao horizonte de previsão. Ele é chamado de janela de previsão e está presente em boa parte dos métodos de previsão de séries temporais. A janela de previsão é utilizada para formar o exemplo (padrões) nos quais alguns métodos de previsão realizam a extracção do conhecimento (aprendizado) para aplicação na previsão de valores futuros. O elemento que segue imediatamente à janela de previsão constitui o alvo, ou seja, o elemento que se deseja prever. A série temporal

é normalmente dividida em dois conjuntos de elementos: o primeiro é destinado ao método de previsão para a obtenção do aprendizado (conjunto de treino) e o segundo é usado para verificação de seu desempenho na previsão de valores futuros (conjunto de teste). A Figura 03 mostra um exemplo da definição desses componentes para previsão de uma série temporal: a divisão da série em conjuntos para treinamento e teste, uma janela de previsão que possui cinco elementos e o horizonte de previsão de um elemento (o alvo).



Ilustração VII – Exemplo de uma janela de previsão com cinco entradas e horizonte de previsão igual a um [Moretin & Tolo. 2009]

2.2.4. PERIODICIDADE

As séries temporais transportam sempre a periodicidade, sendo que estas são um conjunto de características que se repetem ao longo do tempo, ou seja, de tempo em tempo poderá ser observado fenômenos com mesmas características e propriedades.

Em princípio importa apresentar que a periodicidade é válida apenas para toda a variável que se relaciona com o tempo, ou seja, para todo número real, o período é o tempo gasto para que um movimento circular se repita com as mesmas características.

Numa série temporal T , um símbolo S é dito ser periódico com um período p se S existe a cada recorrência de p . Por exemplo, na série temporal $T = abcabbabcb$, o símbolo b é periódico com período 5 desde que b seja representado 4 vezes ao longo da série [Elfeky et al. 2005].

Eis aqui a apresentação do conceito de uma função periódica, de modos a interpretar-se a periodicidade segundo [Guedes, 1992]:

Uma função $f: \mathcal{R} \rightarrow \mathcal{R}$ é dita periódica de período T (ou apenas T -periódica) se existe um número real T tal que $f(x + T) = f(x)$ para todo x real.

Observe que se uma função tem período T então $f(x + nT) = f(x)$ para todo n inteiro, ou seja, é também periódica de período nT .

Ou Seja,

$f(x + T) = f(x)$ até $f(x + nT)$, e a frequência da grandeza é $f = 1/T$.

Uma característica fundamental de dados de séries temporais é a frequência com que as observações são espaçadas no tempo. O número de vezes que as observações de uma série temporal ocorrem, são chamadas de frequência de amostragem ou a periodicidade da série. Por exemplo, uma série temporal com uma observação a cada mês tem uma frequência de amostragem mensal ou periodicidade mensal.

Segundo [SAS Institute Inc, 2002], poderá tirar-se maior proveito dos dados estruturados em séries temporais pela interpretação da periodicidade de ocorrência de eventos e aplicar algum método capaz de fazer a devida leitura destas ocorrências. Assim poderá supor-se que os dados de series temporais estão sempre envolvidos a algum ou alguns periodos. Simplesmente poderão ser traçadas estratégias de modo a se perceber em que períodos certos fenómenos envolvidos decorrem, e sendo assim, mediante esta aplicação poderá conseguir-se métodos eficazes que trabalham com base nesta abordagem.

A observação ou aplicação desta característica poderá ser útil para a observação de mudanças em dados numa estrutura de séries temporais ou aqueles que possam ter algum outro tipo de relação com o tempo.

Em [Elfeky et al. 2005] afirma-se que a periodicidade é usada para prever tendências em dados de séries temporais.

O problema da periodicidade é de encontrar padrões periódicos ou, ciclicidade ocorrendo em um conjunto de dados relacionados com o tempo. Segundo [Lin et al.

2002], O problema relaciona dois conceitos: padrão e intervalo. O interesse é encontrar padrões que se repetem ao longo do tempo e seus intervalos recorrentes (período), ou encontrar os padrões de repetição de uma sequência, bem como o intervalo que corresponde ao período padrão.

De acordo a [Berberidis, Christos et al. 2002], a periodicidade é uma característica particularmente interessante que pode ser utilizada para compreender dados de séries temporais e prever as tendências futuras. Em [Ergun, 2004], também analisou-se algoritmos para análise de tendências pela obtenção da periodicidade. Mas nas suas observações, este afirmou que nenhuma série em tempo real é exactamente periódica, pois existem sempre diferenças.

Especificando intervalos de tempo

Para a periodicidade, existem sempre intervalos, e como método poderá ser importante que estes possam ser facilmente controlados de modo a adaptarem-se a serie temporal a ser estudada.

Em [SAS Institute Inc, 2002], foram especificados intervalos com seguintes objectivos:

- Validar a periodicidade em dados;
- Verificar meses de ocorrência de periodicidades;
- Prever observações de acordo a dados de saída.

Tabela 6 – Tipos de Intervalos básicos como unidade das Séries Temporais [SAS Institute Inc, 2002]

Name	Periodicity
Ano	Anual
Semi - anual	Semi-anual
Quaternal	Quadrimestral
Mensal	Mensalmente
Semi - mensal	1º e 16º de cada mês

Semana	Semanalmente
Dias de Semana	Diária com excepção a finais de semanas
Dia	Diária
Hora	De hora a hora
MINUTE	Para cada uma hora
Segundo	Todos os segundos

A periodicidade é uma característica dos dados de séries temporais, embora o conceito não ser na sua íntegra defendido neste domínio de estudo, uma vez que os dados de series temporais nunca são exactamente os mesmos em suas repetições. Na realidade, essa característica é uma definição aproximada e que pode ajudar para uma melhor a caracterização desse tipo de dados na detecção de outras características como a sazonalidade, tendência e os ciclos.

2.3. MÉTODOS DE CONTROLO ESTATÍSTICO PARA A DETECÇÃO DE MUDANÇAS

O Controlo estatístico é uma metodologia que se baseia na análise estatística e gráfica dos dados de um processo, com o propósito de entender, monitorar e aperfeiçoar seu desempenho.

Segundo [Montgomery, 2009], um método de controlo estatístico típico, tratado por gráfico de controlo exibe três linhas paralelas: a central, que representa o valor médio do característico de qualidade; a superior, que representa o limite superior de controlo (LSC); e a inferior, que representa o limite inferior de controlo (LIC).

Os gráficos de controlo estatísticos devem ser implementados para monitorar as variáveis que influenciam a qualidade de itens produzidos e também para se obter deferentes estados intervalares para o controle estatístico. De acordo com esta afirmação, em [Montgomery, 2009], um processo estatístico estará sob controlo se as causas especiais de variação forem eliminadas do processo, e os pontos plotados do gráfico de controlo permanecerem dentro dos limites de controlo pré-definidos. Para este princípio, uma determinada variável pode ser monitorada por meio de diferentes estados representados no tempo, pois esta é a razão da interpretação de variáveis como em séries temporais. As causas comuns são originárias da variabilidade natural do processo, sendo, portanto, consideradas aleatórias, sinalizando que o processo está

sob controlo estatístico. Já as causas especiais são originárias da formação de padrões, sinalizando que algo de especial está influenciando o processo, sendo, portanto necessário identificá-la e removê-la para que toda a produção não seja afectada. Importante aqui perceber-se que quando estamos diante de causas especiais, diz-se que o processo está fora de controlo estatístico.

Segundo [Vieira, 1999], para planear um gráfico de controlo, é preciso estabelecer o tamanho da amostra e a frequência da amostragem. Não existem fórmulas para determinar esses valores, mas em geral, é possível escolher uma das duas estratégias:

- Tomar amostras pequenas e frequentes;
- Tomar amostras grandes e pouco frequentes.

Segundo [Montgomery, 2009], há pelo menos cinco boas razões para a popularidade dos gráficos de controlo:

- Gráfico de controlo é uma técnica comprovada para a melhoria da produtividade, reduzindo sucatas e retrabalho e, conseqüentemente, aumento de produtividade e baixo custo do produto;
- São eficazes na prevenção de defeitos, pois, com o processo sob controlo, a organização produz certo da primeira vez;
- Com os gráficos de controlo, o operador intervém sobre o processo com a certeza, de um fundamento teórico e provável, de que o processo está fora de controlo, ou seja, “se não está quebrado, não conserte”;
- Fornecem informações de diagnóstico. Para um operador ou engenheiro experiente, pode-se tirar, por meio dos pontos do gráfico de controlo, uma possível mudança no processo que melhore seu desempenho;
- Fornecem informações sobre a capacidade do processo, informam os valores de seus parâmetros mais importantes e também a estabilidade ao longo do tempo, permitindo que se estime a capacidade do processo em atender às especificações dos clientes.

Princípios para gráficos de controlo eficazes

- Os gráficos sempre utilizam limites de controlo localizados à distância de três desvios-padrões da linha média;
- O desvio-padrão utilizado deve ser estimado com base na variação dentro da amostra;
- Os dados devem ser obtidos e organizados em amostras (ou subgrupos) segundo um critério racional;
- O conhecimento obtido através dos gráficos de controlo deve ser empregado para tomada de acções necessárias.

2.3.1. ALGORITMO DE SOMA ACUMULADA

O Algoritmo CUSUM directamente incorpora todas as informações na sequência de valores de amostra, *plotando* as somas acumuladas dos desvios dos valores da amostra a partir de um valor alvo [Montgomery, 2009]. Segundo [Vargas et. al. 2004], este algoritmo valoriza os desvios, considerando-os pontos necessários para a compreensão do comportamento global de um sistema, visando a sua melhoria.

[Perry & Joseph, 2011], também sustentam a teoria afirmando que, um esquema CUSUM acumula a diferença entre um valor observado x_i e um valor de referência k . Se esta soma excede um intervalo de decisão h , o sistema em causa indica que existe uma perturbação.

Para [Wong & Neill, 2009], o objectivo do CUSUM ao agregar os desvios em várias etapas de tempo é de detectar pequenas mudanças em dados relativamente aos seus valores médio. De acordo, [Eftimie, 2009] afirma que o CUSUM é uma importante ferramenta que pode ser usada para detectar pequenas mudanças na média de um processo e numa segunda fase permite fazer o controlo. Uma vez construído o gráfico que permite visualizar as mudanças através de desvios, tal como afirma-se em [Montgomery, 2009], os pontos plotados do gráfico estarão sob controlo estatístico, ou seja, dentro dos limites de controlo se não houver uma ou mais causas especiais no processo. Estas causas especiais são tidas como anomalias que poderão ser analisadas e caracterizadas.

Especialmente para a detecção de mudanças, ou de um conjunto de valores anómalos, [Nenes et. al., 2008] afirmam que, o controle de processo através do algoritmo CUSUM apresenta uma abordagem que é a utilização de dois métodos estatísticos em separado, uma para a detecção de mudanças positivas e outra para mudanças negativas. Está em causa a apresentação de uma abordagem com dois objectivos distintos a considerar. A detecção do crescimento e decrescimento relativamente a média, ou seja, quanto maior for o número de valores positivos a serem adicionados, a função poderá deslocar-se mais para cima, caso contrário poderá deslocar-se para baixo.

Períodos em que o gráfico CUSUM segue um caminho relativamente simples indicam um período em que a média não se alterou, ou seja, caso não haja alterações na média do processo, os desvios positivos serão compensados pelos desvios negativos, sendo que existirá uma oscilação de forma aleatória em torno da média. Neste caso, uma inclinação ascendente indica um período em que os valores tendem a ser acima da média global. Da mesma forma um segmento com uma inclinação descendente indica um período de tempo em que os valores tendem a ser mais abaixo da média global. A mudança brusca na direcção do CUSUM indica uma mudança súbita, ou alteração na média.

Para a interpretação do gráfico CUSUM requer-se alguma prática. Suponha-se que, durante um período de tempo os valores tendem a ser mais acima da média global. A maioria dos valores adicionados à soma cumulativa vai ser a soma positiva e irá aumentar de forma constante, em [Morgenstern, 1988], a aplicação do método CUSUM na detecção de mudanças permite a identificação do instante e da duração de cada mudança.

A função objectiva do CUSUM é um valor obtido da soma entre uma função e o erro desta função com uma tendência a zero. Entretanto, o CUSUM deverá tender a uma função objectiva em que para a soma acumulativa obtida dos valores de erros adicionados a função, deverão anular-se uns aos outros.

Mediante as evidências apresentadas, o método consiste em detectar pequenas mudanças na média ou em processos analisados de determinadas amostras:

- Uma função $f(y)$, irá achar a média de uma nova função que irá depender da soma da média dos dados de uma amostra com a soma dos erros em relação a estes dados.
- $\sum f(y) + \sum e$, para esta função, teremos sempre o controlo dos erros, de maneira a que sempre que ocorrer alguma mudança, facilmente acharemos a diferença que houve em relação ao valor 0 denotada pela fórmula do CUSUM.

De uma forma geral e de acordo com os conceitos apresentados acima, pode-se então justificar com a seguinte equação:

$$Si = \sum_{j=1}^i (x_j - \mu_0)$$

Onde μ_0 é a media do processo.

A aplicação do CUSUM as séries temporais poderá permitir visualizar a taxa de erros relativamente a média e obter as tendências positivas e negativas se comparados a um valor produzido pela variação dos dados que poderá ser nulo.

Para compreendermos o processo de mudança para o acto de detecção de anomalias, podemos verificar que o CUSUM é o somatório de uma função adicionando seus erros pelo número dos dados de entrada desta função. A expressão abaixo vem demonstrar a equivalência que poderá permitir isolar o erro de modo a obtê-lo como uma adição especial.

$$Si = \sum_{j=1}^i (x_j - \mu_0) \leftrightarrow f(y) = \frac{\sum f(x) + \sum e}{n}$$

A função $f(y)$ representa a mesma função, mas com o facto de esta admitir $\sum e$ como uma sequência de erros.

Detectar anomalias com o CUSUM

Sendo uma metodologia não paramétrica, um detector CUSUM detecta pontos de mudança, ou seja, as bordas de subida e descida de um período anómalo, o que indica

a duração em que o anomalia dura. Quando a média de um subconjunto de amostras torna-se mais elevado do que um valor limite, um ponto de mudança é registrado, e o nível de mudança passa à um acumulador. O acumulador altera o seu estado quando o ponto de outra mudança é detectada.

Em estatística a detecção pode ser explicada quando um processo está sob controlo, x_i que tem uma distribuição normal, ou seja, uma média e um desvio padrão que indicam o quanto houve dispersão no sistema. Se o processo tende a se afastar do valor pretendido, o gráfico CUSUM indica a presença de uma causa assinalável que deve ser investigada.

Assim sendo, este método poderá ser importante no intuito de se perceber as reais variações do processo, indicando até que ponto os dados se desviam das suas médias normais e apontar quais os meses que contribuem para estes desvios..

2.3.2. ALGORITMO MOVING AVERAGE

A aplicação deste algoritmo, possibilitará tratar domínios de dados sazonais que apresentam as mesmas características médias [Wonnacott e Wonnacott, 1990]. Em estatística, a média móvel, é um tipo de filtro utilizado para analisar um conjunto de pontos de dados, criando uma série de médias de diferentes subconjuntos de um conjunto de dados.

Uma vantagem é o facto de poder-se manipular a estrutura de dados como um todo ou apenas domínios específicos. Entretanto, este método será utilizado para gerar-se médias específicas tendo em conta um reordenamento de diversos grupos considerando características semelhantes.

Numa análise efectuada pelo *Moving Average*, a cada instante temporal, a previsão é apenas a média das últimas N observações, a média k . Um dos problemas com este método é a escolha de N , o tamanho da janela a ser utilizada. Se N for muito pequeno a previsão tenderá a oscilar muito, ou seja, “suave” que poderá implicar uma previsão como poucos ruídos e uma compreensão menos complexa. Uma característica importante do método de médias móveis é: todas as observações usadas para o cálculo em um intervalo deverão ter o mesmo peso ($1/N$). Mas na prática é razoável supor que as observações mais recentes sejam mais relevantes para previsão dos próximos

valores da serie, e portanto deveriam receber um peso maior que as observações mais antigas.

O objectivo será a compreensão teórica de como aplicar o MM a uma função CUSUM, assim sendo teremos:

Para toda uma sequência temporal, poderá ser aplicado a função seguinte:

$$S_i = 1/n \sum_{j=1}^i (x_j - \mu_0)$$

Sendo que S_i , representa o CUSUM achado para cada intervalo n .

Esta proposta de estudo surge da necessidade de poder analisar-se gráficos que suas médias apresentam longas distâncias entre os cumprimentos de onda, ou seja, a ideia será encurtar as distâncias entre os pontos dados de maneiras a termos uma maior precisão na detecção de mudanças anulando a sazonalidade tendo em conta a característica da serie que iremos aplicar, estaremos então diante de um processo de discretização, em que os dados estarão distribuídos para diferentes estados.

Para um exemplo real, para efectuar-se o cálculo de uma média móvel de 3 meses, no final de cada mês, somam-se as vendas dos últimos três meses e divide-se por 3. Se o pretendido for uma média móvel de quatro meses, no final de cada mês, somam-se as vendas para os últimos quatro meses e divide-se por 4. No final de um período, a média móvel simples de período, pode ser usada como uma previsão para um período se a série de dados é estacionária, ou seja, convergente em que os valores pertencentes a série em causa tendem a um número diferente de ∞ .

Analiticamente, dada uma sequência de valores n, x_1, x_2, \dots, x_n e um tamanho de janela de $k > 0$, a média de k -ésimo movimento da sequência dada é definida como se segue:

$$\begin{aligned}
 Y_1 &= \frac{1}{k}(x_1 + x_2 + \dots x_k) \\
 Y_2 &= \frac{1}{k}(x_2 + x_3 + \dots x_{k+1}) \\
 Y_3 &= \frac{1}{k}(x_3 + x_4 + \dots x_{k+2}) \\
 &\dots \\
 Y_{n-k+1} &= \frac{1}{k}(x_{n-k+1} + x_{n-k+2} + \dots x_n)
 \end{aligned}$$

A sequência de média móvel tem n-k + 1 elementos como mostrado acima. Após aplicarmos a média móvel relativa ao tamanho de uma janela, poder-se-á achar as mudanças de maneira mais precisa, sendo que o erro é relativo a média em que temos.

As médias móveis com k = 4 de uma sequência de 10 valores (n = 10) é mostrada a seguir:

i	1	2	3	4	5	6	7	8	9	10
Input	10	20	30	40	50	60	70	80	90	100
y1	25 = (10+20+30+40) / 4									
y2	35 = (20+30+40+50) / 4									
y3	45 = (30+40+50+60) / 4									
y4	55 = (40+50+60+70) / 4									
y5	65 = (50+60+70+80) / 4									
y6	75 = (60+70+80+90) / 4									
y7	85 = (70+80+90+100) / 4									

Assim, a sequência de média móvel tem n-k + 1 = 10-4 + 1 = 7 valores.

Uma vez aplicado o *Moving Average*, aplicamos o CUSUM de modo a obtermos o controlo das mudanças, mas de forma mais precisa.

2.3.3. INTERPRETAÇÃO DE UM RESULTADO BASEADO EM AGRUPAMENTO DOS DADOS

Uma solução de problemas baseados no estudo e análise de dados, poderá ser tão melhor executada quanto mais regular e uniforme for o comportamento dos dados em questão. Quando se tem uma grande quantidade de dados dificilmente se encontra este tipo de comportamento, principalmente nos casos em que se trata de dados multidimensionais, ou seja dados compostos de diversos atributos, que são objectos da chamada análise multivariada, em que as funções poderão estar dependentes de outras funções externas. A estratégia utilizada normalmente nestes casos é procurar identificar na população de dados, grupos cujos dados apresentam um comportamento semelhante segundo um determinado critério ou factor. O caminho mais comumente utilizado para uma visualização clara é a análise de agrupamento de dados que engloba uma série de técnicas para uma devida segmentação dos dados dependentemente de suas características e propriedades.

Para [Puntar, 2003], dentre os métodos utilizados na Análise Multivariada (Multivariate Analysis), a Análise de Agrupamento (Cluster Analysis) é sem sombra de dúvidas um dos mais importantes, eficazes e comumente utilizados. Estes métodos surgiram da necessidade de classificar, reunir e separar os objectos em estudo de acordo com suas características mensuráveis [Acarini & Amaral, 1992]. Trata-se então de uma ferramenta bastante poderosa para o estudo e compreensão do comportamento de dados nas mais diferentes situações. Um exemplo disto é o caso de dados recolhidos através de pesquisas, onde podemos obter uma quantidade extremamente grande de informações que, observados sob um contexto geral, podem não apresentar nenhum sentido, porem quando classificados e separados em grupos passam a fornecer informações com respeito ao comportamento de cada um destes grupos, pois a necessidade de ser específico no tratamento e compreensão de dados possibilita chegar a conclusões óptimas e com tolerância a erros.

[Acarini & Amaral, 1992] afirmam que, o objectivo é pesquisar a melhor partição de dados, em que cada grupo formado abrangerá os casos mais próximos em relação a sua média. De acordo com esta definição [Reis, 2001] apresenta a abordagem de que, o agrupamento de dados é uma tarefa que procura segmentar populações heterogéneas em subgrupos ou segmentos homogéneos. Os registros são agrupados conforme alguma similaridade em si.

“A análise de Clusters designa uma série de procedimentos estatísticos sofisticados que podem ser usados para classificar objectos e pessoas sem preconceito, isto é,

observando apenas as semelhanças ou dissemelhanças entre elas, sem definir previamente critérios de inclusão em qualquer agrupamento.” [Reis, 2001].

Dependendo dos interesses e objectivos que se deseja atingir, a visualização de agrupamento de dados é de grande realce, pois possibilita obter uma preciosa ajuda para um rápido entendimento e assimilação de informações através das quais possamos avaliar factores relativos a características individuais tais como, tamanhos, a pertinência total ou parcial de uma amostra.

Como pudemos observar, para a adopção da média móvel, existe a necessidade de inferir sobre um modelo de dados classificados dependentemente de suas características com o objectivo de evitar-se interferências, e retirarem-se conclusões para a realidade de cada grupo classificado.

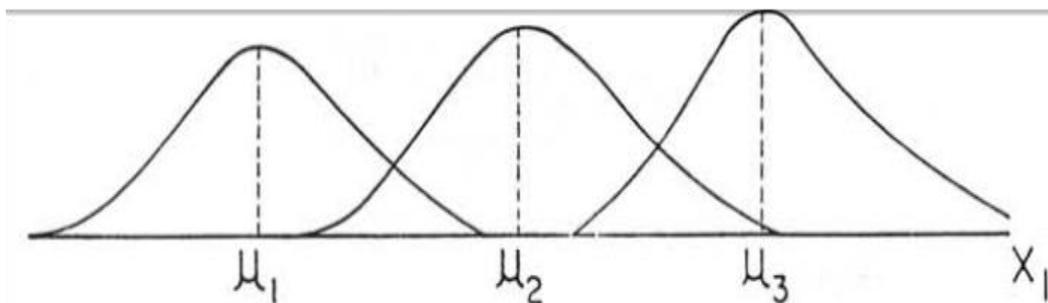


Ilustração VIII – População diferente pela análise de variância [Acarini & Amaral, 1992]

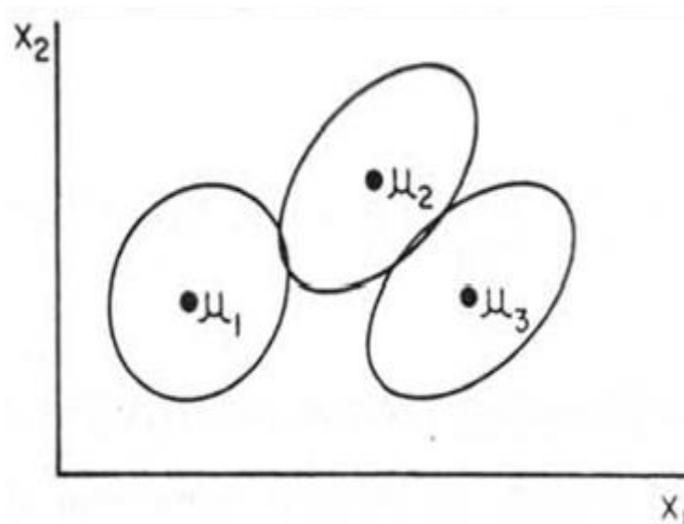


Ilustração IX – Populações distinguidas pela análise multivariada (apenas duas variáveis) [Acarini & Amaral, 1992]

A técnica baseada em agrupar valores, pode resolver eficientemente o problema das tendências e sazonalidades, eis esta a solução encontrada para dar resposta a realidade em causa a ser analisada por tratar-se de grupos a parte. Os grupos surgem a justificar a aplicação das médias por janelas (k-means) para a eliminação dos chamados ruídos. Para uma sequência de valores aleatórios pode-se então agrupar-se estes dependentemente de suas distâncias, sendo valores próximos, pertencentes ao mesmo grupo e valores distantes pertencentes a grupos distintos. Desta forma, pode-se eficientemente identificar quando é que um determinado valor não pertence a nenhuma das classes previstas, sendo este facilmente denotado como uma anomalia. Tais valores, graficamente de maneira simples poderão ser identificados.

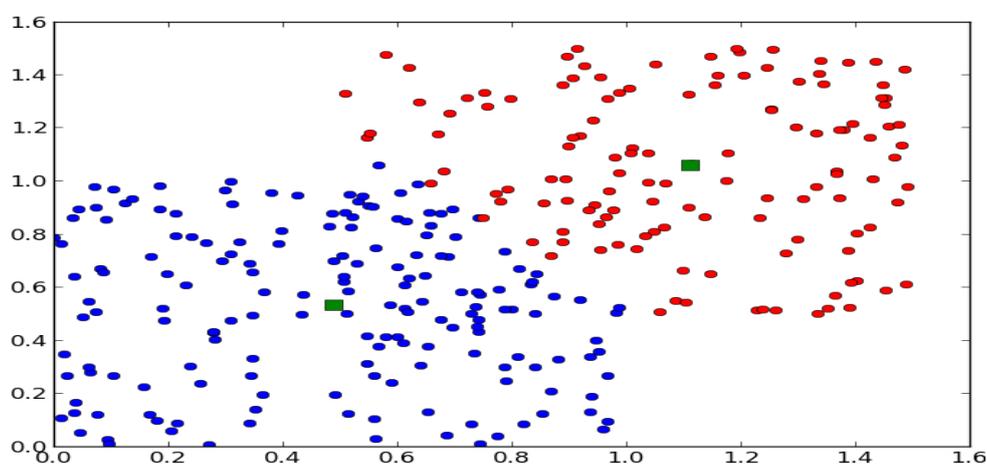


Ilustração X – Representação do agrupamento de dados com a identificação de pontos fora de seus domínios de actuação [Scikits, 2010]

Como pudemos ver, uma tarefa especial frequentemente associada com o agrupamento de dados, é a identificação de pontos fora de determinados padrões. Os grupos representam registros ou objectos similares, entretanto existem muitos objectos que não apresentam uma forte pertinência a nenhum dos grupos em questão.

Pontos fora dos padrões são, portanto, as observações cujas características as identificam distintamente das demais, ou seja, o ponto resultante da combinação de suas características, apresenta-se comparativamente diferente dos demais. Pontos fora dos padrões não podem ser categoricamente caracterizados como benéficos e nem tão pouco como problemáticos, mas sim observados dentro do contexto da análise e avaliados em função do tipo de informação que poderão fornecer.

Um ponto fora do padrão poderá ser indicativo de alguma característica da população que não tenha sido revelada durante o curso normal da análise, neste caso, ainda que distante da maioria das demais observações, apresenta um importante benefício na análise. Por outro lado, poderá também ser contrário aos objectivos da análise, distorcendo seriamente os testes estatísticos e, neste caso, trata-se de um ponto problemático, não sendo portanto representativo da população.

Pontos fora dos padrões poderão ser classificados segundo quatro diferentes classes conforme a seguir:

- Numa **primeira classe** enquadram-se aqueles originados de um erro de procedimento como por exemplo um erro de entrada de dados ou um engano na codificação. Estes pontos seriam identificados na fase de preparação dos dados porem, se não notados, seriam eliminados da análise ou registados como valores erróneos. Vale a pena considerar que, recolher, verificar e processar os dados são actividades que podem e devem ser realizadas conjunta e progressivamente;
- A **segunda classe** são aquelas observações que resultaram de algum evento extraordinário, reflectindo assim o resultado de uma única observação. Cabe então procurar identificar o evento extraordinário que ocasionou tal resultado, devendo o analista decidir pela sua permanência ou não, em função dos objectivos que pretende alcançar através da análise;
- A **terceira classe** abrange observações extraordinárias para as quais não se tem uma explicação plausível. Embora, via de regra, estes pontos tenham que ser provavelmente omitidos, eles poderão ser mantidos caso se note que representam um segmento válido da população;
- Finalmente na **quarta e última classe** enquadram-se as observações cujos valores em cada variável apresentam uma variedade comum porém, quando combinados, através destas, resultam pontos fora dos padrões. Nesta situação as observações deveriam ser mantidas a não ser que exista uma evidência específica que permita desconsiderar o ponto como um membro válido da população.

Para esta abordagem, podemos verificar que um comportamento anormal descreve um determinado indivíduo em que suas características de consumo não vão de encontro a uma suposta classe, entretanto, a metodologia de agrupamento vem dar suporte à resolução desta problemática.

Vantagens que podem ser extraídas no agrupamento de dados:

- Eliminação do efeito sazonal ao aplicarmos em estações climáticas diferentes;
- Possibilidade de segmentar melhor os consumidores com vista a evitar melhor uma anomalia;
- Interpretação gráfica eficiente dos dados;
- Distribuição espacial eficiente dos dados;
- Filtragem para o tratamento semântico dos dados;
- Possibilidades de aperfeiçoamento de métodos de previsão de séries temporais, classificando dados por subgrupos.

3. RESULTADOS DOS MÉTODOS PARA A DETECÇÃO DE MUDANÇAS E ANOMALIAS NO CONSUMO DE ENERGIA ELÉCTRICA DA CIDADE DE LUANDA

3.1. ANÁLISE DOS GRÁFICOS DE CONSUMO

Os dados a serem estudados são relativos ao consumo de energia eléctrica da cidade de Luanda. Foram disponibilizados pelo Gabinete de Sistemas da EDEL no dia 16 de Janeiro de 2013 e representam a facturação dos anos de 2009 até 2012. Estes dados estão expressos em milhões de *kwanzas* (kz), moeda nacional angolana.

Os dados de consumo de Luanda estão classificados em três tipos:

- Consumo de energia de baixa tensão;
- Consumo de energia de baixa tensão especial;
- E consumo de energia de média tensão.

Nesta sessão, o objectivo será a interpretação e a diferenciação entre os três tipos de consumo com o intuito de anotar a forma como estes oscilam. A partir da Ilustração XII pode-se observar a representação das três categorias de consumo como base inicial de comparação. Assim sendo, do ano de 2009 até ao ano de 2012, este foi o comportamento registado no consumo de energia eléctrica da cidade de Luanda.

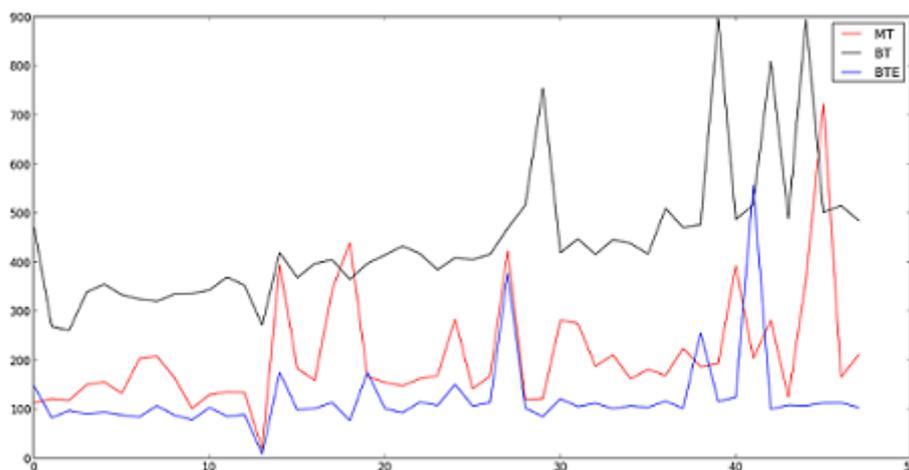


Ilustração XI – Gráfico comparativo entre o consumo BT, MT e BTE

De acordo com este comportamento obtido, alguns dos grandes objectivos a alcançar poderão ser: Quais as principais características a detectar-se de 2009 a 2012. Será que registou algumas diferenças consideráveis? O que poderá ter causado tais diferenças caso tenham ocorrido? Será que é possível apresentar tais diferenças graficamente? Quais poderão ser os métodos aplicados?

3.2. DESCRIÇÃO DA VARIAÇÃO DOS DADOS

Mediante a projecção gráfica dos dados de consumo, nota-se que o consumo de baixa tensão regista os maiores valores, sendo que a maior parte de seus pontos está acima dos pontos de média tensão e de baixa tensão especial. Este comportamento descreve-se porque ao apurar-se os dados, enquanto para a baixa tensão os valores mensais oscilam entre 261 e 897 milhões de *kwanzas*, para a média tensão os valores oscilam entre 18 e 723 milhões de *kwanzas* e para a baixa tensão especial os valores oscilam entre 9 e 559 milhões de *kwanzas*.

Para a baixa tensão e a média tensão verifica-se que existe um crescimento com o passar dos meses. A baixa tensão especial apresenta uma variação estável sem apresentar algum crescimento ou decréscimo. De acordo com os objectivos deste trabalho, os crescimentos ou decréscimos registados passarão a ser denominados como mudanças, pelo facto de tratar-se do deslocamento da média em relação a variação dos valores. Estas mudanças poderão tornar-se mais sensíveis de serem detectadas ao aplicar-se o CUSUM para descrever a soma acumulada dos desvios em relação a média uma vez que esta será a metodologia que atribuirá maior base para este estudo.

O crescimento verificado na baixa e na média tensão poderá descrever que o aumento de consumo de energia eléctrica tem aumentado para os últimos anos. Este facto poderá ser normal, uma vez que a cidade tem crescido em diversos sectores implicando que haja um aumento para o consumo ao longo dos anos. Mais concretamente, o aumento para o consumo de baixa tensão poderá ser uma consequência resultante de um aumento no número de residências enquanto para a baixa tensão especial poderá ser resultante do aumento de número de pequenas e médias empresas e para a média

tensão poderá ser resultante de um crescimento em grandes infra-estruturas registadas nesta categoria.

Como pudemos observar, o resultado obtido na Ilustração XII, surge como uma primeira solução de visualização do comportamento dos dados, uma vez que é base sobre a qual poderá aplicar-se outros métodos de modos a aprimorar os momentos em que registaram-se mudanças. Assim sendo, embora verificar-se a variação dos dados e ter-se obtido o comportamento da variação dos dados, surge a necessidade de aplicar-se então algum método capaz de melhor detectar períodos em que valores passam a crescer ou decrescer para melhor justificar-se os níveis de consumo se comparados a sua média. Uma vez que o CUSUM regista sua soma acumulada, esta apresenta como resultado as mudanças, momentos em que valores crescem e decrescem, entretanto o CUSUM acaba em ser um método capaz de transparecer a realidade da descrição das mudanças para os três tipos de consumo, mas se bem aplicado e filtrado.

Uma vez que os métodos para a detecção de mudanças são aplicados sobre o comportamento de cada tipo de consumo, será normal extrair-los primeiramente características para o modo como os dados variam. As principais características extraídas da variação dos dados para os 3 tipos de consumo poderão resumir-se da seguinte forma:

- Registo dos maiores valores no consumo de baixa tensão;
- A curva do consumo de baixa tensão possibilita a interpretação de um crescimento de ano para ano;
- Detecção de uma tendência de crescimento que poderá continuar por mais anos para o consumo de baixa tensão;
- Possibilidade de interpretação da diferença entre os valores para as três categorias;
- Possibilidade da interpretação da variação para cada tipo de consumo;
- Comparação entre a inclinação dos tipos de consumo descrita pelos valores.

- Assim sendo, o gráfico apresentado na Ilustração I apresentou simples vantagens para a detecção de mudanças no seguinte:
- Interpretação inicial dos dados de consumo identificando tendências de crescimento ou de decrescimentos;
- Diferenciação entre os tipos de consumo;
- Interpretação da real situação sobre a oscilação do consumo de energia eléctrica de Luanda;
- Identificação da necessidade de aplicar algum método específico para a identificação de mudanças no consumo.

Pela análise dos três tipos de consumo, conclui-se que estamos diante de series não estacionárias, em que se verifica a componente sazonalidade e tendência. Como objectivo pretende-se eliminar tais características.

Sendo assim, o método CUSUM possibilitará a projecção das tendências de subidas e descidas relativamente a médias e a posterior aplicação do *Moving Average* permitirá eliminar os ruídos caracterizados como tendência e posteriormente o reajuste das janelas em que o CUSUM deverá actuar poderá implicar também a eliminação da sazonalidade.

3.3. DETECÇÃO DE MUDANÇAS E ANOMALIAS COM O CUSUM

De acordo aos resultados apresentados através da aplicação do CUSUM, a variação do consumo da energia, gera um comportamento a parte que acaba por ser um processo de monitoração baseado nos desvios relativamente a média. Através dos valores são garantidas as suas diferenças em relação a um comportamento supostamente normal que pode ser válido a um grupo que apresenta as mesmas características. A maneira como são apresentadas essas diferenças é que possibilitará fazer a leitura do CUSUM reconhecendo os momentos em que há desvios no consumo e saber até que ponto estes desvios podem vir a influenciar o processo no seu todo.

A leitura do consumo da energia com o CUSUM vem contribuir de forma bastante positiva, uma vez que este método está a ser aplicado numa área bastante crítica e que poderá possibilitar uma melhor segmentação e controlo dos consumidores. O controlo do consumo da energia de Luanda, ainda passa por diversas dificuldades devido ao crescimento desordenado da cidade que influencia na procura e seu mau uso. A aplicação do CUSUM poderá reverter este quadro, uma vez que o processo poderá ser acompanhado ano após ano.

3.3.1. DESCRIÇÃO DO CUSUM RELATIVAMENTE ÀS MUDANÇAS NO CONSUMO

De acordo a sustentação fornecida o CUSUM poderá representar os momentos de subida e de descida do consumo da energia, tratados como mudanças, pois que a mudança não será de acordo a variação dos dados de consumo mensal, mas sim nas deslocações desses valores relativamente a média que ao serem somados poderão descrever a trajectória dos erros que se foram admitindo ao longo dos 4 anos.

Para a interpretação e descrição do CUSUM, precisa-se perceber que, a soma acumulada de desvios positivos permitirão a subida do gráfico e a soma acumulada de desvios negativos irão permitir a descida do gráfico. Uma vez que para cada ponto superior a média irá obter-se valores positivos e para cada ponto inferior irá obter-se valores negativos, assim sendo, a soma de um valor abaixo da média implica um deslocamento do CUSUM para baixo, e a soma de um valor acima da média implica um deslocamento do CUSUM para cima. Para os períodos em que o CUSUM poderá registar sucessivamente alguma descida poderá implicar que somaram-se sucessivamente valores negativos que geram desvios derivados de valores inferiores a média e para os períodos em que poderá registar subidas poderá implicar que somaram-se sucessivamente os valores positivos que geram desvios derivados de valores superiores a média. Neste sentido, os resultados apresentados nos gráficos são apresentados como resultados de algoritmos que funcionam obedecendo a estas regras, embora que houve a necessidade de aplicar-se o *Moving Average* como método que possibilitou uma melhor classificação dos dados de modos a não distanciarem muito da média e melhor acompanharem o processo.

3.3.2. APLICAÇÃO DO CUSUM NO CONSUMO DE BAIXA TENSÃO

A Ilustração XIII, representa a aplicação do CUSUM relativamente aos dados do consumo de baixa tensão. Após ter-se aplicado o CUSUM, verifica-se através do efeito causado, a diferença em relação aos dados mensais que os desvios causam. Em comparação com a Ilustração XII manteve-se a representação dos dados de consumo e obteve-se o resultado da soma acumulativa em relação a estes dados através da curva verificada a baixo. Assim sendo, temos a descrição do registo da diferença dos dados mensais em relação a média, em que esta diferença acaba por descrever a situação real de como as mudanças poderão estar a variar tendo em conta a alguma base comportamental entre todos os utilizadores. Assim sendo, como objectivo, este registo apresenta como comportamento as fases em que os dados distanciam-se pouco ou muito dos padrões de consumo esforçando alguma mudança para cima ou para baixo dos níveis de consumo.

De forma geral, para o consumo de energia de baixa tensão, primeiramente o CUSUM apresenta uma descida, uma vez que para a soma acumulativa os valores são inferiores a média e apresentam uma grande diferença em relação a esta, assim sendo, os resultados da soma esforçam o gráfico a deslocar-se para baixo. A seguir, para a subida verificada, implica que foi-se obtendo a soma acumulativa através de valores cada vez mais acima da média, que resultam em desvios positivos sucessivos.

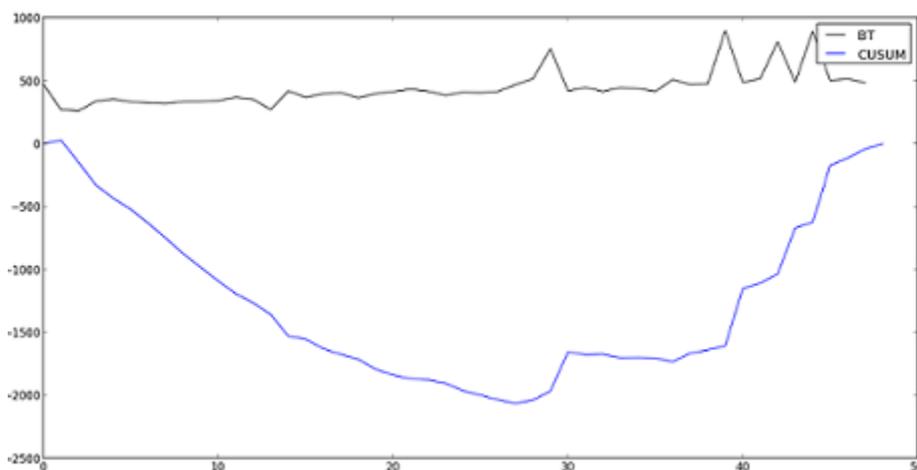


Ilustração XII – CUSUM da BT relativamente aos seus valores de consumo

Tendo em conta a esta metodologia, o comportamento do CUSUM para a baixa tensão poderá indicar os meses que contribuem para consumos baixos e exagerados através do percurso tomado pela soma mas ainda assim pode-se admitir muitos erros pelo facto do intervalo dos meses ser muito grande. Embora o CUSUM distanciar-se

demasiadamente da média dos dados, já indica-nos a presença de uma tendência que poderá ser resultado de uma realidade da cidade de Luanda mas não ainda com tanta legibilidade. Os resultados da soma acumulativa que obrigam a subida do gráfico poderão justificar o crescimento da população podendo implicar um maior consumo para a categoria de baixa tensão, embora exista ainda também os consumos irregulares. Assim sendo, pelo facto do CUSUM distanciar-se muito da média não se pode admitir com tanta legibilidade a realidade de Luanda que é o comportamento de subida no decorrer dos anos. Como objectivo, para dar volta a esta situação poderá ser buscado como auxílio algum método para aproximar os valores da média e ter-se um melhor controlo dos momentos de mudança.

Para a categoria de baixa tensão, poderão ser extraídas algumas vantagens:

- Detecção de subida e descida da taxa de desvios relativamente a média do consumo;
- Detecção e acompanhamento de uma tendência de crescimento populacional através do consumo;
- Interpretação da variação da taxa de mudança no consumo mediante aos valores que deslocam-se da média;
- Contribuição no alerta do consumo anormal entre a população e controlo dos consumos ilícitos.

Caso Luanda verifique um aumento na sua população nos próximos anos, poderá continuar a haver um aumento no consumo de energia eléctrica para a baixa tensão, mas importante também levar-se em consideração a um controlo de consumos irregulares. Através do CUSUM, este crescimento poderá ser acompanhado se aplicar-se o *Moving Average* como método de melhoria para o CUSUM e acompanhar-se as falhas através da monitoração, deste modo, para a categoria de baixa tensão o CUSUM poderá aproximar-se melhor a média e o gráfico melhor poderá descrever a tendência de crescimento. Assim sendo, o consumo poderá ser um indicador determinante para contribuir no processo de monitoração do crescimento populacional.

3.3.3. APLICAÇÃO DO CUSUM NO CONSUMO DE MÉDIA TENSÃO

O CUSUM para a média tensão, embora que também regista uma descida e logo a seguir uma subida, a descida deslocou-se bem menos para baixo, isso implica que a soma acumulada dos desvios em relação a média não desloca-se muito em relação a sequencia dos dados, assim sendo, para a baixa tensão os valores distanciam-se mais da sua média se comparado a média tensão, neste sentido o CUSUM acaba em adaptar-se melhor para esta categoria pelo facto dos valores não apresentarem grandes diferenças em relação a média favorecendo a soma acumulada embora tal facto não implica que as subidas e descidas do consumo já estejam a ser bem acompanhadas por estar-se ainda a trabalhar com todos meses dos 4 anos sem que seja aplicado uma forma de melhor classificar os dados e as médias serem relativas a cada classe.

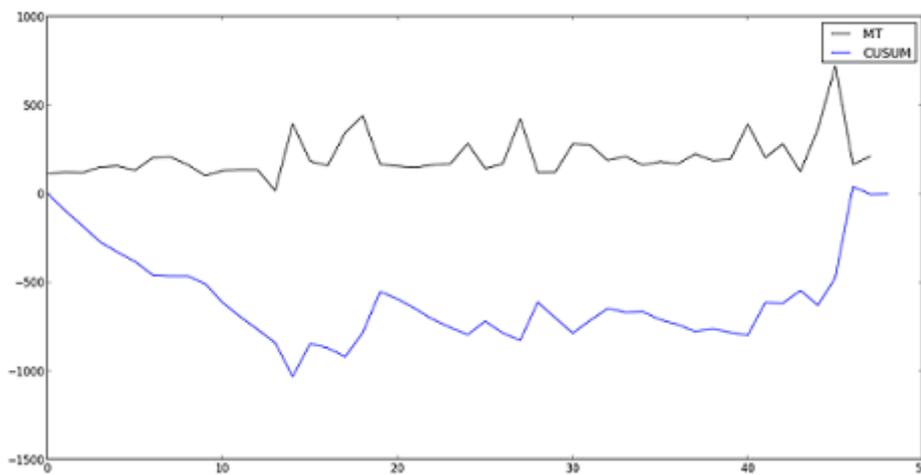


Ilustração XIII – CUSUM da MT relativamente aos seus valores de consumo

Para esta categoria, poderá admitir-se que os dados do consumo têm obedecido melhor a sua média, ou seja, que os registos para cada mês não diferenciaram-se tanto quanto para a baixa tensão. Este resultado demonstra nos últimos anos a média tensão tem estado mais estável em relação a baixa tensão.

Uma vez que os dados do consumo de média tensão são relativos a grandes infra-estruturas, também se poderá considerar normal o comportamento do seu CUSUM, pois que para Luanda existe um crescimento no número dessas infra-estruturas implicando um maior consumo, logo, para o consumo de média tensão, existe uma relação entre o consumo e o crescimento no número destas infra-estruturas. Este facto poderá fazer entender que dependentemente do crescimento económico de Luanda, haverá um maior crescimento no consumo de energia eléctrica para grandes infra-estruturas durante estes anos, mas tal facto poderá ser melhor acompanhado graficamente caso

aplicar-se o MA, uma vez que o CUSUM poderá aproximar-se melhor a média e o gráfico melhor poderá descrever a tendência de crescimento. Assim sendo, este indicador passa a ser determinante para contribuir-se na monitoração relativa ao crescimento das infra-estruturas.

Algumas vantagens extraídas a partir da detecção de mudanças do consumo para a média tensão poderão ser classificadas no seguinte:

- Detecção de subida e descida da taxa de desvios relativamente a média do consumo de média tensão;
- Indicador de acompanhamento de desenvolvimento do sector empresarial;
- Interpretação da variação da taxa de consumo anormal para a média tensão;
- Método que poderá contribuir de forma global para o acompanhamento do crescimento e desenvolvimento da cidade de Luanda.

As vantagens extraídas podem ser apenas consideradas como ponto de partida para o acompanhamento do processo uma vez que o CUSUM ainda se distancia muito da média dos valores mensais.

3.3.4. APLICAÇÃO DO CUSUM NO CONSUMO DE BAIXA TENSÃO ESPECIAL

Para a baixa tensão especial, inicialmente o CUSUM também apresenta uma descida, a tendência de subida também é verificada mas de forma bem mais suave. Este comportamento implica que os valores não apresentam grandes diferenças em relação à média.

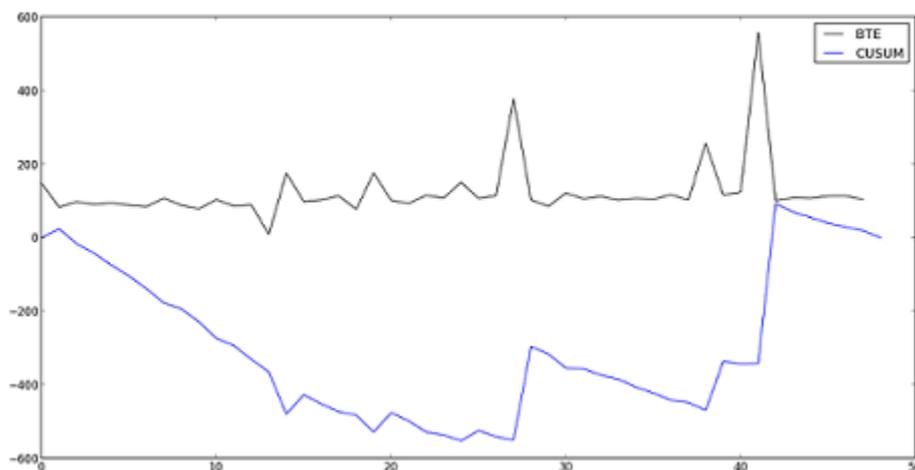


Ilustração XIV– CUSUM da BTE relativamente aos seus valores de consumo

Uma vez que os dados do consumo de baixa tensão especial são relativos a pequenas empresas, também poderá considera-se normal o comportamento do CUSUM uma vez que para Luanda existe um crescimento no número de empresas, logo, para o consumo de baixa tensão especial, verifica-se uma relação entre o consumo e o crescimento no número de empresas. Este facto poderá fazer entender que dependentemente do crescimento económico de Luanda, uma vez que existem indicações de que continuará a crescer para os próximos anos, então o sector empresarial poderá continuar a implicar um maior crescimento no consumo de energia eléctrica para as empresas durante estes anos, mas tal facto também poderá ser acompanhado ao aplicar-se o MA, uma vez que o CUSUM poderá aproximar-se a média e o gráfico irá descrever melhor a tendência de crescimento. Assim sendo, este indicador passa a ser determinante para contribuir-se na monitoração do desenvolvimento neste sector.

Algumas vantagens extraídas a partir da detecção de mudanças do consumo para a baixa tensão especial poderão ser classificadas no seguinte:

- Detecção de subida e descida da taxa de desvios relativamente a média do consumo de baixa tensão especial;
- Detecção e acompanhamento da tendência de crescimento ou decréscimo do consumo de baixa tensão especial para os próximos anos, sendo determinante de acompanhamento do crescimento de grandes Infra-estruturas.

De forma geral, para o CUSUM dos três tipos de consumo, embora possa existir uma diferença considerável, ainda assim, a média distancia-se muito do CUSUM, logo não é certo confiar nos resultados obtidos. Deste modo, poderá ser demonstrado um método para filtragem da média de modos a reajustar-se de acordo a obtenção de submédias no decorrer dos meses, pois assim as tendências poderão ser melhor descritas pelo facto do CUSUM poder adaptar-se ser capaz de acompanhar as submédia sem a necessidade de desviar muito.

3.4. APLICAÇÃO DO MOVING AVERAGE PARA FILTRAR O CUSUM

Relativamente aos dados, o CUSUM apresenta um problema que é o facto de admitir um intervalo muito grande para a média, logo a diferença dos valores em relação a média poderá ser enorme, podendo possibilitar a adição de enormes desvios positivos ou negativos à soma acumulada, assim verificado, a soma dos desvios varia somente entre a distância da média em relação ao valor, então, quanto menor for a distância da média em relação aos valores, melhor poderá ser o CUSUM.

A distância entre a média e os valores pode implicar a perda de muita informação, ou seja, a possibilidade de uma leitura rigorosa do CUSUM na obtenção de suas reais tendências. Podemos verificar que pelo facto do consumo de energia de Luanda registar uma tendência de crescimento, e as médias dos dados distanciarem-se muito, o CUSUM somente não é um método eficaz, este precisa ser adaptado para a esta realidade.

Uma vez que Luanda pode registar um aumento na sua população e em Infra-estruturas nos próximos anos, tal facto poderá implicar um aumento no consumo de energia eléctrica, tal facto poderá ser melhor confirmado ao aplicar-se o MA, de modos a permitir que o CUSUM apresente um melhor comportamento em relação a média e que o gráfico descreva melhor a tendência de crescimento ao longo dos anos. Assim sendo, para este verifica-se um comportamento considerando um ajustamento de Janelas para média móvel abrangendo 1 ano, 6 meses ou 3 meses.

Ou seja,

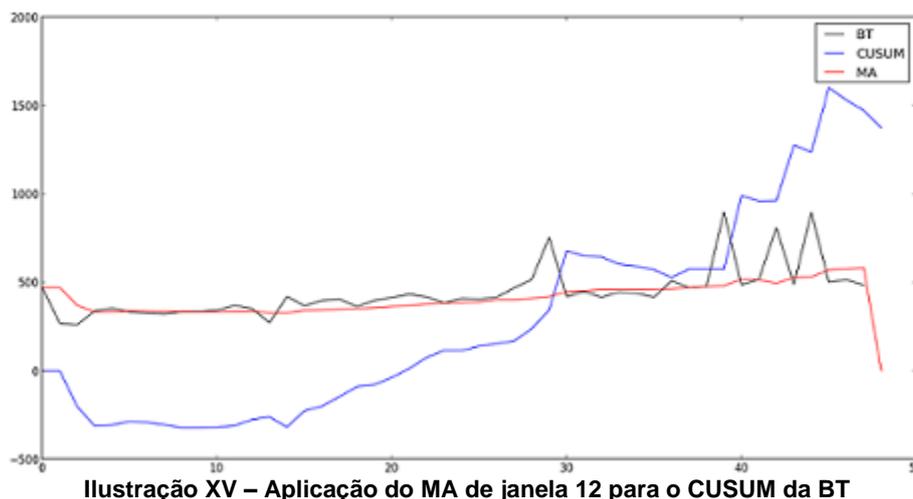
Para toda serie temporal foi aplicado o seguinte:

$$S_i = 1/n \sum_{j=1}^i (x_j - \mu_0)$$

Em que n representa a janela, ou o número o número total dos meses tendo em conta o intervalo que se pretende.

3.4.1. MOVING AVERAGE PARA FILTRAR O CUSUM NA BAIXA TENSÃO

Ao ser aplicado o MA com uma Janela igual a 1 ano para os três tipos de consumo, a média é gerada de 12 em 12 meses, possibilitando que os valores somados ao CUSUM não causam grandes deslocamentos graficamente, mas sim em apenas intervalos considerados de 12 em 12 meses. Como resultado deste método tem-se a Ilustração 5, que representa a curva do CUSUM de acordo com a janela de 12 meses.



O modelo aplicado para a série no caso da Ilustração XVII, está abaixo representado.

Para toda série temporal com o objectivo de ajustamento com uma janela igual a 12 foi aplicado o seguinte:

$$S_i = 1/12 \sum_{j=1}^i (x_j - \mu_0)$$

Neste caso, o CUSUM será aplicado para cada período de 12 meses.

Com este método, o CUSUM não tem a possibilidade de variar muito como anteriormente devido ao intervalo permitido, este acompanha melhor a variação dos dados para um conjunto de médias mais justas e os resultados obtidos as somas acumuladas também não variam. Quanto menos for o intervalo, menos o CUSUM poderá afastar-se da média dos valores.

Sendo assim, pelo facto do CUSUM estar próximo a média, então o comportamento que a média apresentar o CUSUM tenderá a acompanhar apresentando melhor os momentos em que há mudança, ou seja, a medida que a média deslocar-se para cima, o CUSUM também deslocará, e a medida que deslocar-se para baixo, então este também se deslocará, mas com o objectivo de descrever os reais instantes de mudança. Neste caso, a tendência descrita pelo CUSUM é devido a um registo em que houve subida na média, ou seja, que foram-se registando com o passar dos meses maiores consumos relativamente a média que varia de tempo em tempo. Neste caso, é evidente que um dos objectivos é do CUSUM descrever a diferença que houve nos momentos de subida dos meses em relação a média.

Caso optar-se em comparar a baixa tensão já com a aplicação do MA, percebe-se que anteriormente pelo facto do CUSUM basear-se numa média que abrangia todos os meses, os resultados da soma acumulada traduzidas como as mudanças, poderiam subir ou descer muito devido a quantidade de valores somados.

Embora registando-se uma melhoria, ainda poderá perder-se alguma informação importante a ver com o objectivo do trabalho, pois ainda admite-se mudanças enormes derivadas do número de valores que a soma acumulativa ainda regista. O algoritmo gerador dos gráficos, através da aplicação do MA, tem a capaz de prever este problema, dependentemente do número de meses pode-se seleccionar o intervalo máximo que a média móvel pode determinar o CUSUM devido ao facto de acompanhar-se as mudanças o mais próximo possível do consumo dos meses. Esta base funciona partindo do princípio que a forma geral do CUSUM será determinada através da soma da diferença entre os meses e as médias móveis de cada mês. Como exemplo, caso a média mensal de um determinado mês estiver com uma chave igual a 6, então serão geradas médias de consumo de 6 em 6 meses e o CUSUM adaptara-se de acordo a estas médias podendo descrever a trajectória da mudança de acordo a este ajustamento.

De acordo a base apresentada e pela necessidade do CUSUM estar sempre próximo da média, o resultado da Ilustração 6 é demonstrado através do reajuste do MA com uma janela igual a 6 (chave), médias de 6 em 6 meses, de modos a tentar-se comprovar que o resultado poderá ser melhor tendo em conta aos objectivos a cumprir-se.

Verifica-se que ao aplicar o MA com uma Janela Móvel de 6, o CUSUM aproxima-se bem mais da média se comparado ao caso anterior de janela igual a 12 e verifica-se como resultado uma realidade em que o CUSUM descreve o processo de mudança como uma curva bem mais ascendente. As médias são achadas para cada 6 meses e o CUSUM desloca-se admitindo que os valores poderão apenas permitir sequência de desvios apenas para intervalos de 6 meses em 6 meses, logo o CUSUM passa acompanhar as tendências da média com maior precisão se comparado ao caso anterior.

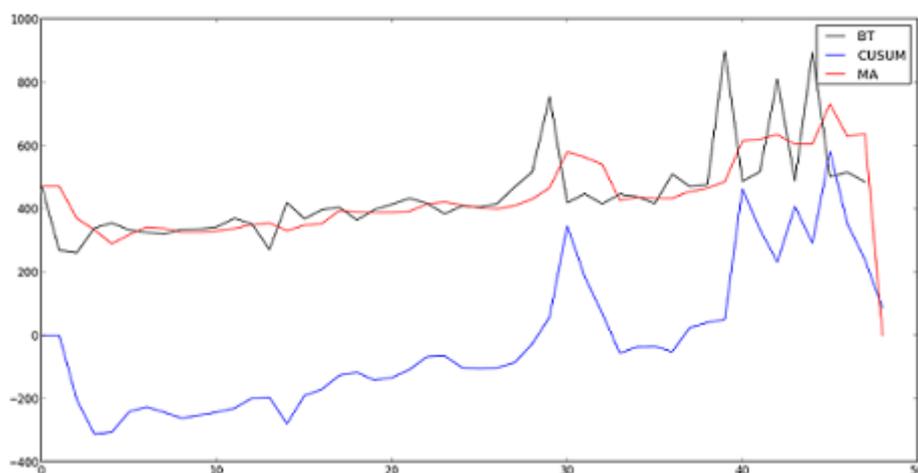


Ilustração XVI – Aplicação do MA de janela 12 para o CUSUM da BT

Com o reajuste de janela igual a 6, já é possível observar realmente o que foi acontecendo no processo de mudança do consumo de energia de 2009 à 2012, mas de modo a cumprir-se o objectivo na íntegra, a Ilustração XVII é resultado de uma janela igual a 3. Assim sendo, irá obter-se uma média de três em 3 meses e uma vez que o CUSUM estará em função para cada uma dessas médias e este irá estar bem próximo a média se comparado ao caso em que a janela é igual a 6.

O modelo com o *Moving Average* com janela de 6 está a baixo representado.

Para toda serie temporal com o objectivo de ajustamento com uma janela igual a 6 foi aplicado o seguinte:

$$S_i = 1/6 \sum_{j=1}^i (x_j - \mu_0)$$

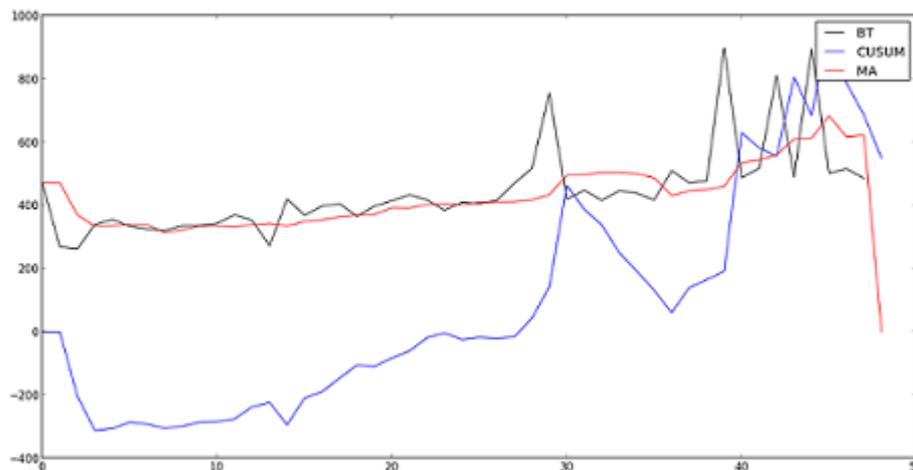


Ilustração XVII – Aplicação do MA de janela 3 para o CUSUM da BT

Uma vez que os valores tendem a aumentar, então a média readapta-se a cada instante e o CUSUM também apresenta uma tendência de subida sendo que estará muito próximo a média. Neste caso, mesmo para os maiores consumos registados, o CUSUM já não se desloca tanto em relação a média e verifica-se que a tendência de crescimento que anteriormente se distanciava bastante da média, já apresenta uma boa proximidade.

Para este caso, o CUSUM comprova a particularidade prevista em que o reajuste considerando uma janela igual a 3 possibilitou que o CUSUM obedecesse melhor o comportamento da média se comparado aos casos anteriores em que as janelas são superiores, sendo que mesmo para os pontos que registam as maiores tendências de subida, o CUSUM aproxima-se bem aos dados.

Para toda série temporal com o objectivo de ajustamento com uma janela igual a 3 foi aplicado o seguinte:

$$S_i = 1/3 \sum_{j=1}^i (x_j - \mu_0)$$

Com a adaptação aplicada ao algoritmo, facilmente interpreta-se o comportamento das mudanças uma vez que o CUSUM descreve melhor o processo. Verifica-se também que “quanto menos o CUSUM deslocar-se da média, melhor será a capacidade de resposta ao problema que é de acompanhar-se as tendências de subida ou de descida”. Nesta base tem-se que o grande objectivo da aplicação do MA foi de possibilitar que o gráfico CUSUM descrevesse melhor a sua trajectória relativamente aos valores mensais não distanciando-se pouco da média. Com a metodologia anterior sem o MA, existia uma maior dificuldade de obtenção dos resultados de acordo ao objectivo deste trabalho. Logo, para este caso, já descreve-se melhor os momentos de mudança do consumo de energia.

O grande contributo que esta parte do trabalho vem dar é de possibilitar uma leitura digna das fases em que houve mudança de 2009 para 2012. “Neste caso poderá dar-se resposta de como ocorreu o processo de mudança e acordo aos dados de 3 em 3 meses, de 6 em 6 meses ou de 1 em 1 ano”, mas tendo em conta que quanto menos forem os meses melhor comportamento o gráfico terá.

3.4.2. MOVING AVERAGE PARA FILTRAR O CUSUM NA MÉDIA TENSÃO

Dependentemente do crescimento económico de Luanda, uma vez que existem indicações de que continuará a crescer para os próximos anos, para o consumo de média tensão poderá continuar haver um crescimento, uma vez que regista-se um crescimento em grandes infra-estruturas, assim, será aproveitado o MA como o método já proposto para melhor se poder confirmar estas indicações. A aproximação do CUSUM a média mediante a aplicação deste método permitirá o gráfico descrever realmente se o consumo tem ou não crescido.

Assim sendo, para efectuar-se a confirmação do consumo de energia nesta categoria também demonstrara-se o resultado pela atribuição ao CUSUM as janelas de 1 ano, 6 meses e 3 meses através da sua média móvel.

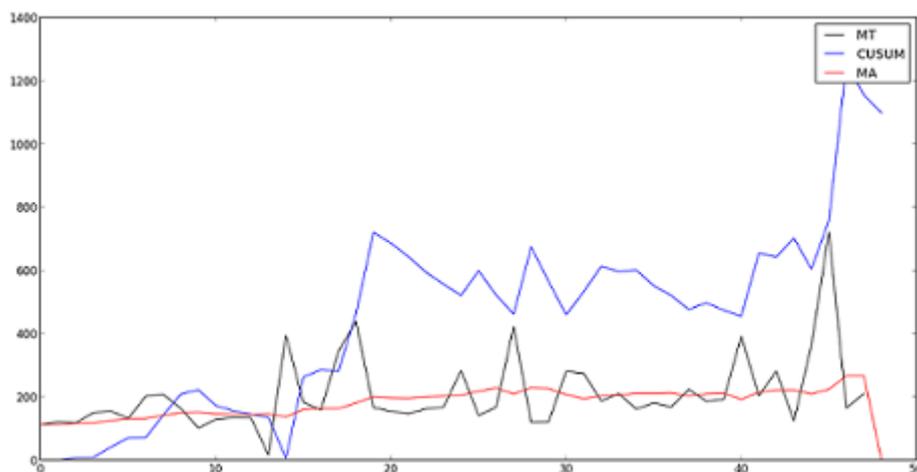
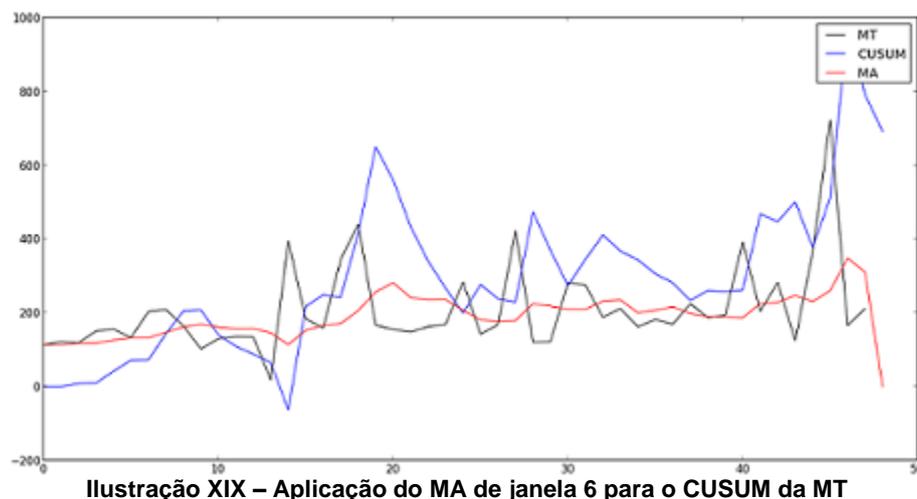


Ilustração XVIII – Aplicação do MA de janela 12 para o CUSUM da MT

Poderá constatar-se como particularidade o facto que boa parte dos pontos do CUSUM ultrapassam com algum significado os valores e para os últimos pontos desenvolve-se uma tendência de subida que representa uma grande diferença com os pontos iniciais, a quantidade de somas de muitos valores positivos, ou seja, superiores a média possibilitou o CUSUM a deslocar-se bem mais para cima. Assim, verifica-se que ao longo dos 4 anos, o consumo registou um crescimento repentino. Este dado acaba importante no sentido de despertar interesses na busca de reais motivos que levaram a isto acontecer para a baixa tensão, mas acaba em ser um aspecto positivo por tratar-se de um tipo de consumo que tem registado clientes que são bons impulsionadores no desenvolvimento de Luanda.

Neste caso, verifica-se uma diferença considerável em relação a baixa tensão e a baixa tensão especial, ao tratar-se da maior parte do CUSUM situar-se acima da média a uma distância superior e a soma acumulada ter ultrapassado os máximos valores de consumo. Ainda não existe muita precisão para uma boa descrição do processo, pois nos casos anteriores foi demonstrado que os reais resultados passam a existir após efectuar-se um ajustamento com números cada vez menores.

De modos a melhor apurar-se ou contrariar-se o comportamento verificado inicialmente será então verificar-se o resultado com o ajustamento na janela igualado a 6.



Para a janela igual a 6, uma vez que o MA está bem mais próximo dos valores, este acompanha melhor o comportamento dos dados e escreve exactamente as tendências de subida e descida já com a proximidade à média, justificando que a inclinação não na realidade não é tão acentuada quanto parecia anteriormente, mas acaba-se por verificar que o crescimento foi maior para este caso. A questão anterior, em que o consumo registou um crescimento repentino já não se poderá justificar com o ajustamento igual a 6, pois também que se acaba mais uma vez por comprovar que não poderá ser certo tirar conclusões quando o ajustamento ainda é igual a 12, pois perde-se informação preciosa que possibilita tirar conclusões mais precisas.

Com os acontecimentos observados, poderá pensar-se que enquanto para o consumo de baixa tensão considerando a janela de tamanho 6, o CUSUM já encontrava-se próximo a média, neste caso os valores ainda encontram-se a uma distância que ainda dificulta na real descrição das mudanças. Neste caso acaba em ser mais interessante a necessidade de aplicar-se a seguir o ajustamento com a janela de tamanho 3 a ver o que poderá obter-se como resultado.

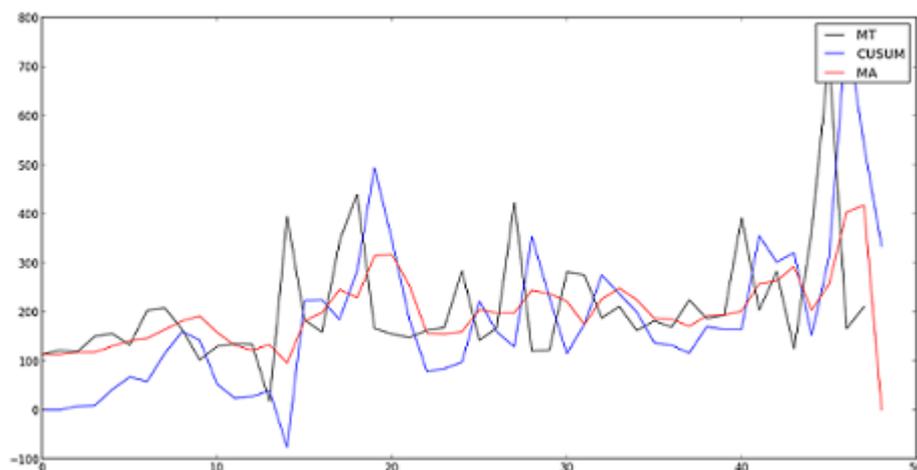


Ilustração XX – Aplicação do MA de janela 3 para o CUSUM da MT

Observa-se na Ilustração XX, que para a média móvel com janela igual a 3, a curva acaba em se situar bem mais próxima da média e a trajectória de crescimento mais uma vez acaba menos ascendente, pois implica que as conclusões que estavam a ser tiradas inicialmente, acabam aqui descartadas.

Importante também tomar como nota conclusiva que embora a baixa tensão possa ter registado valores de consumo maior, tal facto não implicou que o seu CUSUM iria logo deslocar-se mais relativamente a média tensão.

3.4.3. MOVING AVERAGE PARA FILTRAR O CUSUM NA BAIXA TENSÃO ESPECIAL

Dependentemente do crescimento económico de Luanda, uma vez que existem indicações de que continuará a crescer para os próximos anos, o sector empresarial poderá continuar a implicar um maior crescimento no consumo de energia eléctrica para as empresas durante estes anos, mas tal facto poderá ser melhor confirmado ao aplicar-se o MA, uma vez que o CUSUM poderá aproximar-se melhor a média e o gráfico melhor poderá descrever a tendência de crescimento. Assim sendo e como anteriormente, para este tipo de consumo verificou-se o comportamento do CUSUM considerando um ajustamento de janelas para média móvel abrangendo 1 ano, 6 meses e 3 meses.

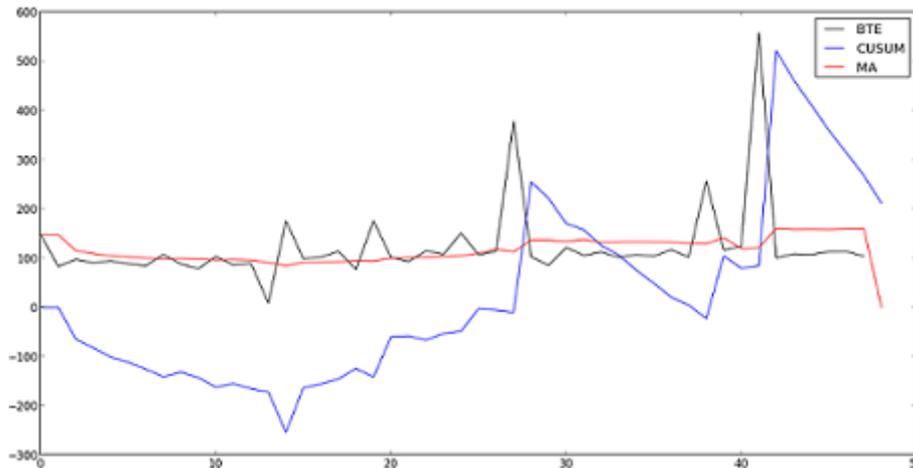


Ilustração XXI – Aplicação do MA de janela 12 para o CUSUM da BTE

Considerando a janela de 12 para o consumo de baixa tensão especial, o comportamento do CUSUM é bem melhor relativamente ao consumo de baixa tensão com o MA de janela 12, este obedece melhor o deslocamento da média, embora que o caso anterior apresenta uma subida bem mais ascendente. Logo a partida, os dados não apresentam grandes diferenças em relação aos desvios, ou seja, andam bem mais próximos da média, possibilitando a melhor descrição das subidas e descidas. Verifica-se mais uma vez a tendência de subida também confirmando que se trata de uma realidade, mas não tanto quanto na baixa tensão, uma vez que admitindo logo a Janela de um ano o CUSUM desloca-se mais próximo a média, logo a partida pode-se pensar que para esta categoria, embora verificar-se o crescimento, mas existe uma boa diferença em relação a baixa tensão. Já é possível afirmar que o consumo na baixa tensão especial cresceu menos em relação a baixa tensão.

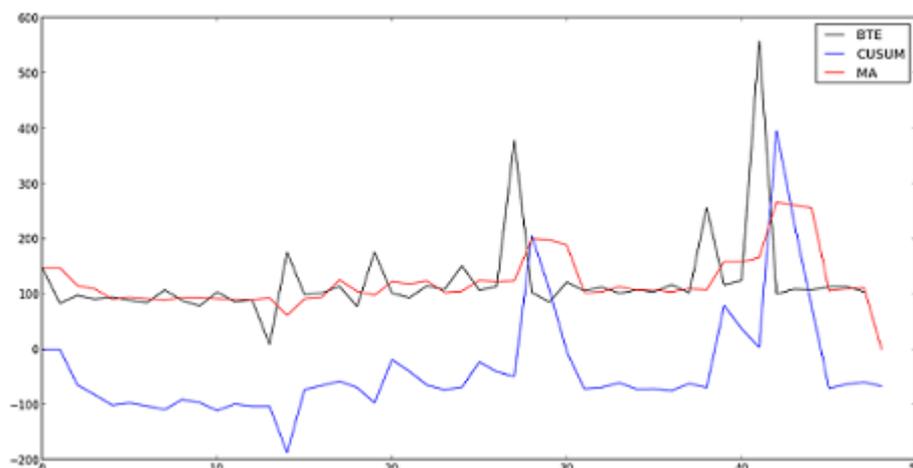


Ilustração XXII – Aplicação do MA de janela 6 para o CUSUM da BTE

De modo a melhor se comparar ao caso do consumo de baixa tensão e de média tensão, também gerou-se o gráfico com a Janela igual a 6 e igual a 3. Para o caso da Ilustração 12, em que a Janela é igual a 6, nota-se também que o CUSUM apresenta o melhor resultado em relação a Janela igual a 12, mas neste caso, ao diminuir-se o tamanho da Janela, melhor comprova-se que houve um equilíbrio no decorrer dos 4 anos, ao invés de registar-se subida. Assim sendo, caso possa existir algum crescimento para este tipo de consumo, talvez poderá ter ocorrido em anos anteriores.

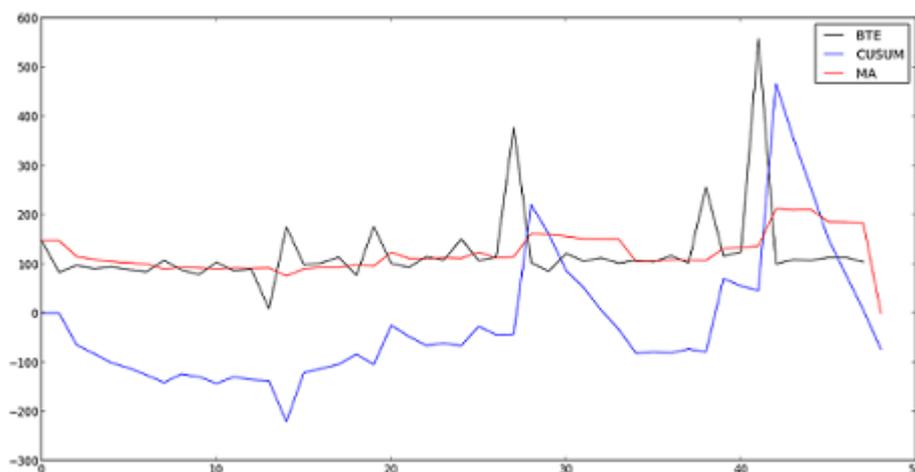


Ilustração XXIII – Aplicação do MA de janela 3 para o CUSUM da BTE

Para uma Janela de 3, verifica-se melhor que o CUSUM, e nesta fase já verifica-se que o consumo de baixa tensão especial embora não tanto mas regista também uma subida de 2009 para cá. Com esta observação, percebe-se que é preciso não tirar conclusões precipitadas quando os valores atribuídos para a janela são ainda grandes, é preciso testar-se valores cada vez mais inferiores de modo a obter-se as características do estudo em causa. Assim poderá ficar provado da necessidade que houve de diminuir-se a dimensão da janela não só para as outras categorias, mas também para esta.

Assim, acaba-se então por verificar que o consumo de baixa tensão especial não registou grandes mudanças, mas certo que tal experimentação acaba em comprovar mais uma vez o processo evolutivo que Luanda tem registado.

Esta categoria é a que melhor pode comprovar que para determinar as mudanças no consumo com estes métodos, deve-se ter em conta o comportamento dos desvios em

relação a média e não somente por tratar-se de valores altos, pois viu-se que a média tensão mesmo não tendo os tais valores tão altos quanto a baixa tensão mas foi a categoria em que houve uma maior subida na mudança verificada.

Mediante a aplicação do método MA a esta categoria poderá destacar-se os seguintes pontos essenciais:

- Detecção de subida e descida da taxa de desvios relativamente a média do consumo;
- A aplicação do MA permitiu que o gráfico CUSUM descrevesse melhor a sua trajectória relativamente aos valores mensais não se distanciando muito da média;
- O CUSUM passa a acompanhar as tendências de mudança com maior precisão quando comparado ao caso anterior;
- Com o reajuste de janela igual a 6, já é possível observar realmente o que foi acontecendo no processo de mudança do consumo de energia de 2009 à 2012;
- Com a metodologia anterior sem o MA, existia uma maior dificuldade de obtenção dos resultados de acordo ao objectivo deste trabalho;
- Detecção de níveis de descida ou subida que podem contribuir no acompanhamento do desenvolvimento do sector empresarial através do processo de mudança.

3.5. APLICAÇÃO DA PERIODICIDADE PARA O CUSUM

Embora o MA resolve o problema do CUSUM, ainda assim o facto de os dados pertencerem a diferentes períodos, estes transportam a característica da sazonalidade.

Para resolver-se o problema da sazonalidade, foram considerados períodos de modos a evitar-se características extras a cada dado. O algoritmo considera as médias de cada período e acha o CUSUM para os mesmos meses de cada ano, assim sendo, um mês representa todos os valores deste mesmo para cada ano. Desse modo, poderá evitar-

se a componente sazonal. Os períodos estão agrupados por uma estrutura de meses de cada ano.

Neste caso, ao aplicarmos o CUSUM para cada mês de todos os anos, poderá existir uma estrutura que além de evitar a sazonalidade poderá ser melhor para acompanhar-se a tendência. Este facto poderá implicar que para cada mês, o consumidor pode apresentar características somente naquele mês sem interferir-se a outros meses.

Ou seja,

Para cada mês em todos os anos, foi aplicado o seguinte modelo:

$$S_i = 1/4 \sum_{j=1}^i (x_j - \mu_0)$$

Assim sendo, o objectivo é a aplicação do algoritmo CUSUM ordenando a serie temporal do consumo por mês, enquanto que anteriormente a aplicação do MM, baseou-se no mesmo modelo mas aplicado para todos os meses seguidos.

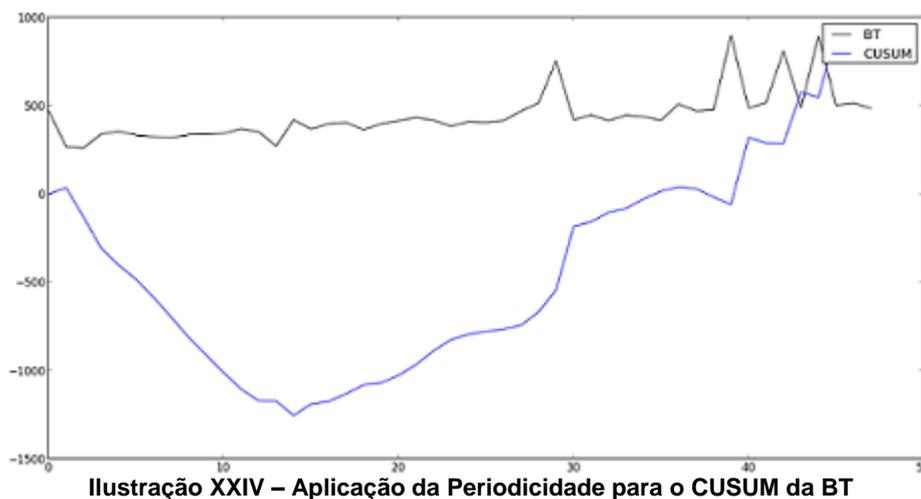
Com a aplicação da periodicidade, poderá obter-se a vantagem de resumir-se o processo todo num único gráfico e globalmente analisar-se as tendências considerando a eliminação da sazonalidade e o acompanhamento dos períodos analisando mês a mês de todos os anos.

3.5.1. PERIODICIDADE PARA O CUSUM DE BAIXA TENSÃO

Este método é aplicado pelo facto das mudanças ocorrerem por períodos e como um capaz de representar uma boa metodologia de previsão, poderá resolver-se a questão através da eliminação do factor sazonal.

Este método surge com o objectivo de resolver-se o problema tendo em conta que os dados transportam como componente a sazonalidade. Assim sendo, sobre o CUSUM, ao aplicar-se a periodicidade como método, a média será considerada de acordo ao mês, isto é, a média dos meses de Janeiro, dos meses de Fevereiro e assim sucessivamente. Após resolver-se as médias para cada mês de 2009 à 2012, a seguir será então aplicado o CUSUM. Isto implica que o CUSUM irá funcionar tendo em função

a diferença dos valores com a média de cada mês para todos os anos. É nesta base que o Algoritmo prevê a resolução do problema da sazonalidade e este método poderá ser adequado pelo facto dos dados serem temporais.



Para este caso, inicialmente nota-se uma descida do CUSUM e isso implica que foram adicionados sucessivamente desvios negativos. Mas em seguida, verifica-se logo a tendência de subida pela adição sucessiva de desvios positivos, acontecendo uma subida ultrapassando a média no último período, o que não havia acontecido antes de ser aplicado a periodicidade. Isto acontece pelo facto do CUSUM acompanhar o crescimento da média a medida que avança-se nos meses. Este resultado vem também demonstrar, que ao obter-se os momentos de mudanças do consumo de energia para a baixa tensão de 2009 para 2012 considerando que as alterações climáticas transportam alguma influência ao longo dos anos, tem-se também como resultado um aumento no consumo no decorrer destes anos que poderá ser resultado do crescimento de Luanda e que este resultado poderá melhor representar a realidade a ser estudada.

Ao comparar este caso com a Ilustração XIII, a maneira de lidar com os dados tornou o algoritmo melhor no cumprimento dos objectivos traçados. Neste caso, poderá perceber-se a evolução do processo de mudança do primeiro ao último ano para todos os meses, ou seja, em uma só questão afirmativa chave pode-se ter a noção da dimensão do problema. “Para todos os meses dos anos de 2009, 2010, 2011 e 2012, a partir de Janeiro a Dezembro, ache o processo evolutivo do consumo de energia”.

Para este caso, acaba-se por considerar que as médias periódicas atribuem um novo efeito as mudanças do CUSUM, por uma melhor justificação da tendência de subida. Neste sentido para a energia de baixa tensão mais uma vez confirma-se haver um crescimento no consumo de energia eléctrica da cidade de Luanda e o CUSUM por si poderá ser um método eficaz, mas desde que elimina-se a sazonalidade.

3.5.2. PERIODICIDADE DE MÉDIA TENSÃO

Para a análise na média tensão, verifica-se também a tendência de crescimento do CUSUM. E também se verifica o facto de ser verdade a tendência de subida. As médias periódicas para a média tensão revelam uma subida do CUSUM muito acima da média ao comparar-se com a Ilustração XIII. Assim sendo, a tendência de crescimento acontece pelo facto do CUSUM acompanhar o desenvolvimento da média, logo o consumo médio foi registando subidas, ou seja, o seu crescimento na passagem de mês em mês, podendo assim facilmente perceber os meses em que houve maior contributo para o aumento no consumo para esta categoria.

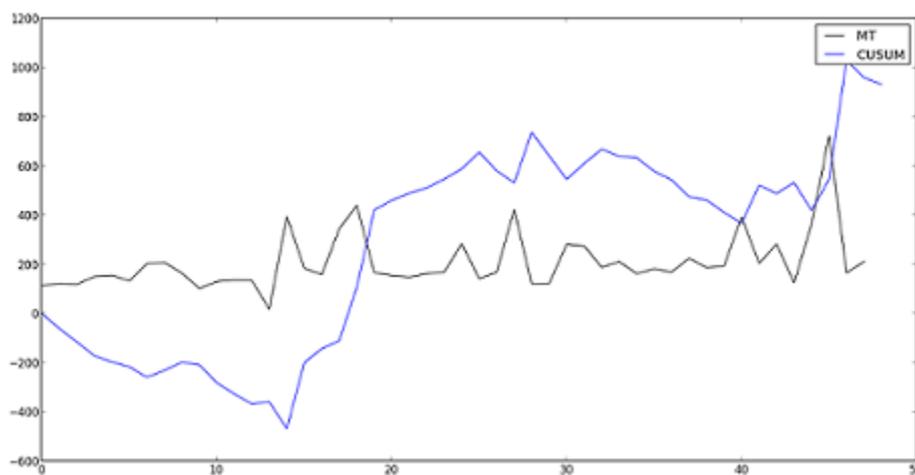


Ilustração XXV – Aplicação da Periodicidade para o CUSUM da MT

Neste caso foram somados mais desvios negativos que positivos. O facto de existirem mais somas positivas implica uma maior tendência a subida de valores e verifica-se que o CUSUM passou a subir mais cedo em relação a Ilustração XIII, o que implica um maior deslocamento para cima.

Em seguida, verifica-se logo a tendência de subida pela adição sucessiva de desvios positivos, ultrapassando a média, o que não havia acontecido antes de ser aplicado a

periodicidade. Isto acontece pelo facto do CUSUM acompanhar o crescimento da média a medida que se avança nos meses, definitivamente entende-se que para a baixa tensão existe uma tendência de subida ao considerarmos os intervalos periódicos por mês.

O facto de verificar-se a partir do gráfico que aproximadamente um terço do consumo está acima da média, poderá justificar o facto de Luanda registar um crescimento bastante elevado para grandes infra-estruturas, a medida que os anos vão passando, e assim sendo um resultado bastante útil que poderá servir de base para outras contribuições.

Quanto ao resultado obtido, acaba-se em descrever um novo efeito do CUSUM, ao confirmar a subida da energia de média tensão e justificar o facto de ser o tipo de consumo que regista uma maior subida ao considerarmos sequencialmente os meses. Este processo evolutivo do CUSUM, poderá ser tida como normal pelo facto de existir muito interesse no sector empresarial, uma vez que para estas categorias boa parte dos contractos são de empresas.

3.5.3. PERIODICIDADE PARA A BAIXA TENSÃO ESPECIAL

Na análise dos dados de baixa tensão especial, verifica-se que é a categoria que não apresenta uma tendência de crescimento, pelo contrário, ainda se nota uma ligeira inclinação para baixo. Neste caso, os valores de consumo relativamente a média baixaram ligeiramente, se considerarmos os meses de Janeiro a Dezembro de 2009 a 2012.

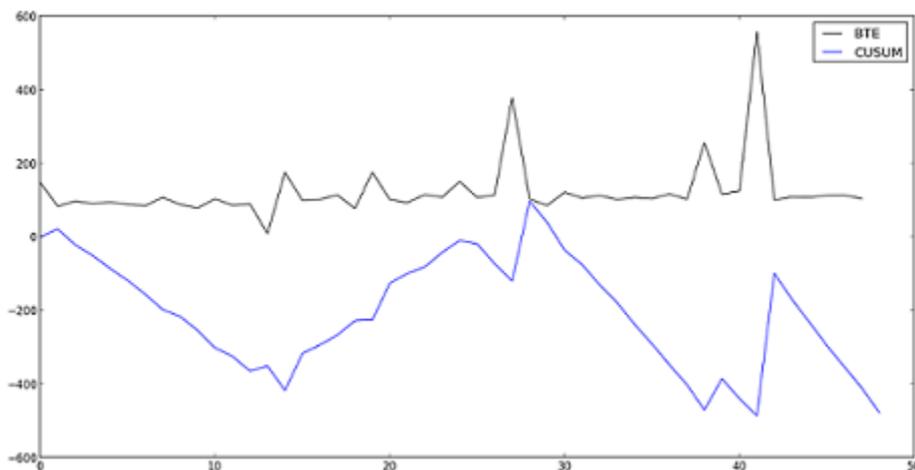


Ilustração XXVI – Aplicação da Periodicidade para o CUSUM da BTE

Ao comparar o CUSUM desta categoria com o CUSUM da Ilustração XIV nota-se que sem a aplicação da periodicidade tem-se uma informação ao contrário, enquanto com a periodicidade o CUSUM regista um ligeiro deslocamento para baixo, sem a aplicação da periodicidade, isso na Ilustração XIV, o CUSUM apresenta um ligeiro deslocamento para cima embora de maneira bem suave. Este facto verifica-se porque a soma cumulativa obteve-se do resultado da adição de mais erros negativos do que positivos, embora não haver tanta diferença relativamente as outras categorias, uma vez que o CUSUM encontra-se mais ou menos equilibrado. Assim sendo, embora registar uma ligeira descida, trata-se da categoria em que os valores rondam mais próximos em volta da média.

As médias periódicas, atribuem um novo efeito as mudanças do CUSUM, pela confirmação de uma ligeira descida. Neste sentido para a energia de baixa tensão especial, passa a existir uma ligeira descida no consumo de energia da cidade de Luanda, demonstrando que no final de 2012 houve uma pequena queda no consumo de energia por pequenas empresas.

Vantagens na aplicação deste método:

- Método eficaz na resolução do problema da sazonalidade;
- Detecção de subida e descida da taxa de desvios relativamente à média do consumo organizada por meses para a baixa tensão;
- Detecção e acompanhamento da tendência de crescimento e contribuição de cada mês ao longo de todos os anos para a baixa tensão;
- Para este caso, acaba-se por considerar que as médias periódicas atribuem um novo efeito as mudanças do CUSUM, por uma melhor justificação da tendência de subida;
- Interpretação da variação da taxa de mudança no consumo populacional organizado por meses para todos os anos;

- Detecção de níveis de descida para o consumo populacional organizado por meses para todos os anos;
- Contribuição e acompanhamento do crescimento populacional organizado por meses para todos os anos.

4. CONCLUSÕES

4.1. CONCLUSÕES GERAIS

A característica das séries temporais aplicadas a este trabalho, permitiu optar-se por um algoritmo que melhor se adequa a realidade dos dados de consumo, uma vez que a mudança de hábito dos consumidores está relacionado com os intervalos de tempo. Importou considerar o tempo por fases distintas. Este estudo permitiu verificar o impacto que estas metodologias estão a contribuir para a evolução e inovação de sistemas cada vez mais personalizados, pois a ideia é ir de encontro as mudanças de comportamento entre os utilizadores/consumidores.

O consumo de energia pode ser considerado um dos principais indicadores do desenvolvimento económico e do nível de qualidade de vida de um país. Uma leitura bem aprimorada sobre como os dados de consumo apresentam mudanças para os diferentes tipos, acaba por um método poderoso de filtragem que poderá possibilitar a compreensão do poder do consumidor, assim sendo, de forma geral o poder de consumo do consumidor da capital de Luanda tem aumentado, numa fase que se regista um aumento da economia Angolana. Para os resultados obtidos, a aplicação da periodicidade ao CUSUM melhor possibilitou descrever a realidade na qual se baseou o estudo, uma vez que apresenta o crescimento classificando os quatro anos mediante cada mês resolvendo a questão da sazonalidade. Para este facto poderemos saber quais os meses que apresentam maiores indicadores do consumo de energia eléctrica e quais as contribuições dos resultados desses para o sistema de consumo em geral.

De uma forma geral as grandes oscilações nas variações observadas no consumo de energia decorrem principalmente do aumento no número de consumidores e de alterações no consumo específico de energia nos diferentes sectores. Um dos principais factores também poderá ser o crescimento económico.

Há evidências claras de que o crescimento do consumo e o crescimento da renda são muito mais ligados se comparado a previsão de determinadas teorias [Squalli e Kenneth, 2006]. Em [Squalli e Kenneth, 2006], parece que a suavização do consumo ocorre durante períodos de vários anos, não de várias décadas. Isso poderá implicar que a proposta de controlo anual do consumo de energia acaba em ser uma boa prática, pois que poderá possibilitar uma visão daquilo que é o crescimento da renda populacional,

num país em crescimento, pois é neste conceito que poderá dizer-se que as boas práticas no consumo de energia eléctrica acabam em ser um requisito fundamental para que o processo esteja controlado.

O Algoritmo aplicado aos dados foi capaz de apresentar três possíveis formas distintas de detectar-se as mudanças, embora que duas das quais são consequências da necessidade de melhor obter-se os resultados. Com esta mobilidade, fica evidente que o estudo poderá continuar através da inserção de novos métodos e que facilmente poderá ser adaptável em outras áreas. Estes métodos são de fácil aplicação, uma vez que a computação actualmente apresenta ferramentas poderosas com pacotes completos capazes de fornecer resultados cada vez mais desejados e específicos.

A aplicação do MA foi bastante importante, pois demonstrou-se que o CUSUM pode funcionar sempre em concordância com o MA desde que este apresente sempre um ajustamento de acordo a realidade. Pois que o ajustamento do MA em relação ao CUSUM, possibilita uma classificação dos dados para médias justas de modos a evitar que uma média pode aproximar-se para alguns dados mas distanciar-se de outros. No caso do consumo de energia eléctrica, o modo de consumo para o mês de Setembro pode não ser considerado igual ao modo de consumo para o mês de Janeiro, então achar uma média ideal para situações diferentes possibilita solucionar o problema de acordo a variação comportamental dos consumidores para os diferentes meses.

Na análise actual dos dados surge uma controvérsia caso pretenda-se determinar os pontos críticos (anomalias), visto que para a determinação de anomalias há que se levar em questão a sensibilidade e origem dos dados para que estes sejam válidos nos seus devidos segmentos, ou seja, uma vez detectada um anomalia, esta poderá ser considerada simplesmente anomalia para um determinado grupo, assim como pode não ser considerada para outro grupo.

Uma das abordagens que se pretende levar em consideração como sendo de grande importância para a compreensão das anomalias é a classificação dos dados, denominadas também como agrupamento. Esta prática acaba em tornar algoritmos bem mais adaptáveis para determinadas realidades, pois que a título de exemplo pode-se verificar que a questão social de um consumidor pode ser muito importante caso se pretenda analisar médias e variâncias de consumo de energia eléctrica. Ao aplicar-se

uma metodologia de agrupamento de acordo ao nível social dos consumidores conclusões mais específicas e surpreendentes poderão ser retiradas, uma vez que pode-se admitir muitos erros ao afirmar-se que as razões que levam um indivíduo de baixa renda a atingir um nível alto de consumo serem as mesmas a de um indivíduo de alta renda. Para a resolução desta questão, o *moving average* baseado em valores k , poderá resolver.

Notou-se em princípio que valores distantes das médias não são certas para descrever-se tendências de crescimento. O alisamento pelo MA, possibilitou uma análise mais ao pormenor, mas não foi suficiente basear-se apenas a este método uma vez que ainda houve a necessidade de eliminar-se a componente sazonal dos dados, assim sendo a periodicidade como método, forneceu uma estrutura diferente de classificação dos dados de modos a solucionar-se este problema.

Entre consumidores também é importante realçar que a falta de uma linha de orientação de boas práticas pode contribuir para um aumento no consumo, embora não tão significativo, mas o crescimento desordenado de uma parte da cidade poderá implicar um aumento ilegal do consumo, neste caso não se pode afirmar a 100% que o crescimento verificado no gráfico de consumo não é simplesmente resultado do desenvolvimento que a cidade de Luanda regista. Questões como esta poderão surgir como resultado do contributo deste trabalho.

Para a falta de controlo de consumo de energia que pode ser causado pela falta de Informação e por consumidores derivados do crescimento desordenado de uma parte da cidade de Luanda poderá surgir algumas orientações a tomar-se como sugestões de melhoria graças a aplicação dos métodos propostos. Linha de orientações que poderão estar resumidas a seguir:

Acompanhar o crescimento da cidade através do aperfeiçoamento de métodos para a detecção das mudanças e anomalias;

Acção de comunicação para a sensibilização e orientações para a economia no consumo de energia eléctrica a medida que registar-se subidas no consumo;

Reforçar-se a importância de diminuição no consumo de energia eléctrica tendo em conta a sua implicação negativa ao meio ambiente contemplando um conjunto de boas práticas entre os consumidores;

Aperfeiçoamento nos métodos para a uniformização do consumo para as diferentes categorias (um cumprimento de igualdade de deveres entre os consumidores nas três categorias).

As leituras que desenvolvemos, não pretendem ser uma resposta fechada e definitiva, é importante no intuito de poder perceber-se as mudanças que têm havido no consumo de energia de modos a atrair outros estudos sendo nesta ou em outras áreas. Será apenas uma possível compreensão da realidade em estudo. Uma questão que também poderá alertar-se é a observação do comportamento dos utilizadores e quais são os principais factores causadores para a mudança de suas atitudes no consumo.

“As mudanças existem, acompanhar através de boas metodologias é que pode originar um grande desafio, pois que não se deve apenas basear-se nos dados, mas é preciso investigar-se métodos que melhor adequam-se a esta especialidade.”

4.2. CONTRIBUIÇÃO PARA O CONHECIMENTO CIENTÍFICO

Este trabalho pode ser muito importante pelo facto de providenciar aos académicos um estudo em uma área ainda pouco explorada. Concretamente para o consumo de energia em Luanda, poderá servir de um guia de orientação para o surgimento de novos estudos nesta área. A apresentação dos resultados sobre a mudança poderá servir de base de estudo para outras áreas, como a Economia, desenvolvimento do sector empresarial, crescimento populacional, etc.

A sequência dos métodos aplicados, facilmente poderá ser importado para a análise de dados em outras áreas.

Uma mais valia é o facto de este trabalho servir de uma base que posteriormente poderá atrair interesses com objectivos de analisar-se outros processos semelhantes.

A aplicação destes métodos, tornou-se mais simples como desenvolvimento de novas ferramentas, assim como novas linguagens da computação. Lidar com grandes quantidades de dados anteriormente envolvia muito mais paciência e tempo.

Actualmente acaba em ser da responsabilidade de especialistas da computação apresentar algumas técnicas clássicas que estavam já a merecer pouca atenção e suas vantagens através das novas ferramentas e novas técnicas.

Uma grande vantagem, é também o facto de estas metodologias estarem a ser cada vez mais aplicadas a medida que linguagens de programação vão apresentando pacotes completos que podem com soluções completas.

4.3. LIMITAÇÕES DO ESTUDO

Após ter-se decidido o tema uma grande limitação foi verificada no processo para a obtenção dos dados, pois por tratar-se de dados que normalmente são protegidos, é necessário aguardar algum tempo até que o pedido seja avaliado e aprovado, logo este facto implicou algum atraso no avanço de determinadas fases da Dissertação.

Para o consumo de energia, eléctrica também não existia muitas publicações que estudam causas que influenciam na mudança de consumo de energia eléctrica, teve que ser necessário efectuar-se uma recolha de informação dos mesmos métodos estudados com tipos de dados diferentes, tais como, industriais, climáticos, ambientais, vigilância em hospitais, etc. Especificamente em Angola existem não existem estudos sobre consumo numa visão da computação, e de forma geral, existem poucas publicações que poderiam fornecer alguns dados importantes de modos a compreender certas implicações no comportamento dos consumidores.

4.4. RECOMENDAÇÕES FUTURAS

Este trabalho é recomendado organismos que lidam com análise de dados. Os algoritmos aplicados para a obtenção dos resultados facilmente adaptam-se em outros estudos, desde que os dados obedeçam uma estrutura em série temporal.

Academicamente este trabalho poderá ser importante no fornecimento de bases para quem pretende saber sobre relações importantes entre as ciências económicas, estatísticas e da computação. Assim sendo, poderá criar estímulos para o conhecimento de linguagens de programação viradas a computação científica.

Algumas áreas em que tais métodos são frequentemente aplicados, são: Finanças, Marketing, Ciências Económicas Sociais, Seguros, Demografia, Meteorologia, etc.

REFERÊNCIAS

Acarini, E., and G. Amaral. "Métodos estatísticos para agrupamento não hierárquicos de dados." *Boletim IG-USP. Publicação Especial* 13 (1992): 45-63.

Alpuim, Teresa. "Detecção de periodicidades em séries temporais: o Periodograma." (1995).

Balestra, Pietro, and Marc Nerlove. "Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas." *Econometrica: Journal of the Econometric Society* (1966): 585-612.

Berberidis, Christos, et al. "Multiple and partial periodicity mining in time series databases." *ECAI*. Vol. 2. 2002.

Bingham, Brian. "On to Python: Engineering computing using Python." (2012).

Box, George EP, Gwilym M. Jenkins, and C. Reinsel. "Forecasting and control." *Times Series Analysis* (1970).

Cardoso, Maria Regina Alves. "Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos." *Rev. bras. Epidemiol* 4.3 (2001).

Chatfield, Christopher. "The analysis of time series: an introduction Chapman and Hall." (1982).

Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." *ACM Computing Surveys (CSUR)* 41.3 (2009): 15.

Das, Kaustav, and Jeff Schneider. "Detecting anomalous records in categorical datasets." *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007.

EDEL - Empresa de Distribuição de Electricidade, EP. "Relatório de Actividade." (2011).

Eftimie, Nicolae. "COMPUTER PROGRAM FOR THE CUSUM CHART." *Bulletin of the Transilvania University of Braşov*• Vol 2 (2009): 51.

Elfeky, Mohamed G., Walid G. Aref, and Ahmed K. Elmagarmid. "Periodicity detection in time series databases." *Knowledge and Data Engineering, IEEE Transactions on* 17.7 (2005): 875-887.

Ergun, Funda, S. Muthukrishnan, and S. Cenk Sahinalp. "Sublinear methods for detecting periodic trends in data streams." *LATIN 2004: Theoretical Informatics*. Springer Berlin Heidelberg, 2004. 16-28.

ESLING, PHILIPPE, and CARLOS AGON. 2012. "Time-Series Data Mining." *ACM Computing Surveys* 45, no. 1: 12:1-12:34. Business Source Elite, EBSCOhost (accessed February 17, 2013).

Guedes, Manuel Vaz. "Grandezas Periódicas Não Sinusoidais." (1992).

Gujarati, Damodar, and Porter, Dawn. (2008). *Basic econometrics*. 5th edition, New York, USA, McGraw-Hill companies.

Guralnik, Valery, and Jaideep Srivastava. "Event detection from time series data." *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1999.

Hunter, John D. "Matplotlib: A 2D graphics environment." *Computing in Science & Engineering* (2007): 90-95.

Hunter, John, and Darren Dale. "The Matplotlib User's Guide." *Matplotlib 0.90. 0 user's guide* (2007).

Hwang, Seokyon. 2011. "Time Series Models for Forecasting Construction Costs Using Time Series Indexes." *Journal Of Construction Engineering & Management* 137, no. 9: 656-662. Academic Search Premier, EBSCOhost (accessed February 17, 2013).

Jones, Eric, Travis Oliphant, and P. Peterson. "Open source scientific tools for Python." (2001).

Kawahara, et. al. "Change-point detection in time-series data based on subspace identification." *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on.* IEEE, 2007.

Kawahara, Yoshinobu, and Masashi Sugiyama. "Change-point detection in time-series data by direct density-ratio estimation." *Proceedings of 2009 SIAM International Conference on Data Mining (SDM2009).* 2009.

Kou, Yufeng, Chang-Tien Lu, and Dechang Chen. "Spatial Weighted Outlier Detection." *SDM.* 2006.

Lin, Weiqiang, Mehmet A. Orgun, and Graham J. Williams. "An overview of temporal data mining." *Proceedings of the 1st Australian data mining workshop (ADM02). Canberra, Australia.* 2002.

Lundh, Fredrik. *Python Standard Library.* O'Reilly Media, Inc., 2001.

Makridakis, Spyros G., and Steven C. Wheelwright. "Forecasting methods for management." (1989).

Martins, Lourdes Conceição, et al. "Relação entre poluição atmosférica e atendimentos por infecção de vias aéreas superiores no município de São Paulo: avaliação do rodízio de veículos." *Rev Bras Epidemiol* 4.3 (2001): 220-29.

Montgomery, Douglas (2009) – *Statistical Quality Control.* Com a colaboração de Luís Carlos Barroso. 6.ªed. Lisboa : Arizona State University.

Morettin, Pedro Alberto, and Clélia Maria de Castro Toloí. *Modelos para previsão de séries temporais.* Vol. 1. Instituto de matemática pura e aplicada, 1981.

Morgenstern, V. M., B. R. Upadhyaya, and M. Benedetti. "Signal anomaly detection using modified CUSUM method." *Decision and Control, 1988., Proceedings of the 27th IEEE Conference on.* IEEE, 1988.

Nenes, George, and George Tagaras. 2008. "An economic comparison of CUSUM and Shewhart charts." *IIE Transactions* 40, no. 2: 133-146. Academic Search Premier, EBSCOhost (accessed February 5, 2013).

Perry, Marcus B., and Joseph J. Pignatiello. 2011. "Estimating the time of step change with Poisson CUSUM and EWMA control charts." *International Journal Of Production Research* 49, no. 10: 2857-2871. Business Source Elite, EBSCOhost (accessed February 7, 2013).

Puntar, Sergio Gonçalves. *Métodos e Visualização de Agrupamentos de Dados*. Diss. UNIVERSIDADE FEDERAL DO RIO DE JANEIRO, 2003.

Salvador, Stan, Philip Chan, and John Brodie. "Learning States and Rules for Time Series Anomaly Detection." *FLAIRS Conference*. 2004.

SAS Institute Inc. 2008. *SAS/ETS® 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

SciKits Learn. 2010. "A demo of the mean-shift clustering algorithm." [Em linha]. Acedido em Janeiro 4, 2013, em http://scikit-learn.org/0.6/auto_examples/cluster/plot_mean_shift.html

Shekhar, Shashi, Chang-Tien Lu, and Pusheng Zhang. "Detecting graph-based spatial outliers." *Intelligent Data Analysis* 6.5 (2002): 451-468.

Song, Xiuyao, et al. "Conditional anomaly detection." *Knowledge and Data Engineering, IEEE Transactions on* 19.5 (2007): 631-645.

Squalli, Jay, and Kenneth Wilson. "A bounds analysis of electricity consumption and economic growth in the GCC." *EPRU, Zayed University, WP-06-09* (2006).

Sujatha, B. "A SURVEY ON PERIODICITY DETECTION IN TIME SERIES DATABASE." *Journal of Global Research in Computer Science* 4.5 (2013): 30-32.

Vargas, Vera do Carmo C., Luis Felipe Dias Lopes, and Adriano Mendonça Souza. 2004. "Comparative study of the performance of the CuSum and EWMA

control charts." *Computers & Industrial Engineering* 46, no. 4: 707-724. Business Source Elite, EBSCOhost (accessed February 17, 2013).

Vieira, Sonia. *Estatística para a qualidade: como avaliar com precisão a qualidade em produtos e serviços*. Campus, 1999.

Weigend, Andreas S., Morgan Mangeas, and Ashok N. Srivastava. "Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting." *International Journal of Neural Systems* 6.04 (1995): 373-399.

Wong, Weng-Keen, and Daniel B. Neill. "Tutorial on Event Detection KDD 2009." *Age* 9 (1909): 30.

Zarpeão, B., et al. "Detecção de anomalias em redes de computadores." *SIMPÓSIO BRASILEIRO DE TELECOMUNICAÇÕES, XXVII. Anais* (2009).

BIBLIOGRAFIA

Jones, Eric, Travis Oliphant, and Pearu Peterson. "SciPy: Open source scientific tools for Python." <http://www.scipy.org/> (2001).

Langtangen, Hans Petter, ed. *Python scripting for computational science*. Vol. 3. Springer, 2008.

Larget, Bret, and Donald L. Simon. "Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees." *Molecular Biology and Evolution* 16 (1999): 750-759.

Lopes, Celi Aparecida Espasandin, and R. C. C. P. Moran. "A estatística e a probabilidade através das atividades propostas em alguns livros didáticos brasileiros recomendados para o ensino fundamental." *Artigo publicado nos anais de artigos selecionados para a Conferência Internacional: Experiências e Perspectivas do Ensino da Estatística – Desafios para o século*. Vol. 21. 1999.

Lundh, Fredrik. *Python Standard Library*. O'Reilly Media, Inc., 2001.

Neal, Radford M. "Probabilistic inference using Markov chain Monte Carlo methods." (1993).

Oliphant, Travis E. "Python for scientific computing." *Computing in Science & Engineering* 9.3 (2007): 10-20.

Pérez, Fernando, Brian E. Granger, and John D. Hunter. "Python: an ecosystem for scientific computing." *Computing in Science & Engineering* 13.2 (2011): 13-21.

Sanner, Michel F. "Python: a programming language for software integration and development." *J Mol Graph Model* 17.1 (1999): 57-61.

Van Rossum, Guido, and Fred L. Drake Jr. "The Python Language Reference." (2013).

ANEXOS

LISTA DE ANEXOS

- Anexo A** - Tabela do consumo de energia eléctrica de Baixa Tensão de Luanda 2009 - 2012
- Anexo B** - Tabela do consumo de energia eléctrica de Média Tensão de Luanda 2009 - 2012
- Anexo C** - Tabela do consumo de energia eléctrica de Baixa Tensão Especial de Luanda 2009 - 2012

ANEXO A

Tabela do consumo de energia eléctrica de Baixa Tensão de Luanda 2009
- 2012

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2009	471.4	268.8	261.3	339.1	355.2	333.4	325.0	320.2	335.1	336.2	343.0	370.0
2010	352.6	271.6	419.9	368.5	397.4	404.9	364.8	397.6	414.6	433.1	417.4	384.4
2011	409.3	405.6	416.0	469.4	516.1	754.9	419.2	447.5	415.9	446.1	438.5	416.7
2012	509.8	470.8	476.6	897.8	487.1	517.0	809.9	488.3	895.3	501.9	515.5	485.2

ANEXO B

Tabela do consumo de energia eléctrica de Média Tensão de Luanda 2009 -
2012

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2009	113.5	121.7	118.7	150.4	155.8	132.8	203.3	208.5	164.2	102.2	129.9	135.3
2010	134.5	17.6	394.8	183.7	158.9	344.6	439.9	166.4	155.0	148.0	162.5	168.4
2011	284.0	142.2	167.8	423.2	120.5	121.8	281.7	275.2	188.0	211.2	162.3	181.7
2012	169.2	224.6	187.0	193.4	392.6	204.0	282.2	124.4	363.9	722.6	165.8	211.1

ANEXO C

Tabela do consumo de energia eléctrica de Baixa Tensão Especial de Luanda
2009 - 2012

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2009	147.5	83.2	97.3	90.6	94.3	88.5	84.6	107.4	88.1	78.7	103.4	86.0
2010	89.4	8.8	175.6	98.7	102.1	114.0	77.3	175.8	101.3	93.2	115.0	107.9
2011	151.0	106.9	113.9	378.0	102.6	85.4	121.6	105.8	112.4	101.9	107.2	104.2
2012	116.9	102.3	256.3	116.5	124.4	557.9	100.5	108.8	107.5	113.1	113.5	104.1