

**TÉCNICAS ESTATÍSTICAS APLICADAS EM  
IMAGENS DO *SPECKLE* DINÂMICO**

**RICARDO MARQUES DA COSTA**

**2009**

**RICARDO MARQUES DA COSTA**

**TÉCNICAS ESTATÍSTICAS APLICADAS EM IMAGENS DO  
*SPECKLE* DINÂMICO**

Dissertação apresentada à Universidade Federal  
de Lavras como parte das exigências do  
Programa de Mestrado em Engenharia de  
Sistemas, para a obtenção do título de "Mestre".

Orientador

Profa. Dra. Thelma Sáfadi

LAVRAS  
MINAS GERAIS - BRASIL  
2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Costa, Ricardo Marques.

Técnicas estatísticas aplicadas em imagens do *Speckle*  
dinâmico / Ricardo Marques Costa. – Lavras : UFLA, 2009.  
83 p. : il.

Dissertação (Mestrado) – Universidade Federal de Lavras, 2009.

Orientador: Thelma Sáfadi.

Bibliografia.

1. *Bio-speckle*. 2. Matriz STS. 3. Análise multivariada. 4. Séries  
temporais. 5. Agrupamentos. I. Universidade Federal de Lavras. II. Título.

CDD – 519.535

**RICARDO MARQUES DA COSTA**

**TÉCNICAS ESTATÍSTICAS APLICADAS EM IMAGENS DO *SPECKLE*  
DINÂMICO**

Dissertação apresentada à Universidade Federal  
de Lavras como parte das exigências do  
Programa de Mestrado em Engenharia de  
Sistemas, para a obtenção do título de "Mestre".

Aprovada em 20 de fevereiro de 2009

Prof. Dr. Roberto Alves Braga Jr.      UFLA

Prof. Dr. João Domingos Scalon      UFLA

Profa. Dra. Thelma Sáfadi  
UFLA  
(Orientadora)

LAVRAS  
MINAS GERAIS – BRASIL

A meu pai, Waldemar Costa (*in memoriam*),  
pelo amor, apoio, compreensão e exemplo de  
vida.

***DEDICO***

## AGRADECIMENTOS

A Deus, por ter sido paciente e ter me ajudado tanto nas horas mais difíceis.

Aos professores Thelma Sáfy e Giovanni Francisco Rabelo, pelo apoio, estímulo, amizade e orientação no desenvolvimento deste trabalho.

Ao Professor Roberto Alves Braga Jr., pelas valiosas sugestões que ajudaram na condução dos ensaios.

A todos os professores da Pós-Graduação em Engenharia de Sistemas da UFLA, que deram apoio para que o trabalho prosseguisse.

Aos professores membros do Colegiado de Engenharia de Sistemas e aos coordenadores Giovanni Francisco Rabelo e Tadayuki Yanagi Jr. Que, como coordenadores, buscaram manter o estudo de qualidade e o bem-estar dos alunos para desenvolver suas dissertações.

A minha amiga Luciene, por todo o companheirismo, pelas sugestões e ajuda durante essa jornada, além da sincera amizade demonstrada.

A todos os meus amigos da graduação, João, Cleiton, Helon, Carlos, Eder, Pedro, Márcio e Murilo, que me incentivaram e me apoiaram.

Aos meus amigos Paulo Eduardo e Denis, que me incentivaram, pelos momentos de alegria e descontração.

Aos meus amigos da cidade de Areado, que sempre confiaram em meu trabalho.

A todos os colegas da pós-graduação em Engenharia de Sistemas e do Departamento de Ciências Exatas da UFLA, em especial Fabrício, Litle, Popó, Ricardo, Edcarlos, Marlon, Crysttian, Gabriel, Leandro e Alison, pela amizade e companheirismo.

Aos colegas Claudinei, Anderson, Patrícia Siqueira, Nádia, Renata e Denise, pela amizade e troca de conhecimentos nos grupos de estudo.

A minha amiga Iza, pelas orações e apoio.

Ao Jaime e as minhas tias Terezinha e Neusa, pelos conselhos e incentivo.

Aos meus cunhados, Wagner e Glaucia, pela amizade e carinho.

Aos meus irmãos, Silmara e Júlio, pelo apoio e incentivo.

A minha mãe, minha luz, que não poupou esforços para que eu chegasse onde estou.

Ao meu pai (*in memoriam*), meu grande e maior amigo, que, ainda com vida, me ajudou e me incentivou para que eu não desistisse.

Meus sinceros agradecimentos.

## SUMÁRIO

LISTA DE ABREVIATURAS .....	i
RESUMO .....	ii
ABSTRACT .....	iii
1 INTRODUÇÃO .....	1
2 REFERENCIAL TEÓRICO .....	4
2.1 Interferometria digital de <i>speckle</i> (DSI) .....	4
2.2 Aspectos relevantes do padrão de <i>speckle</i> .....	8
2.2.1 Propriedades do padrão de <i>speckle</i> .....	8
2.2.2 Estatísticas do padrão de <i>speckle</i> .....	10
2.2.2.1 Momento de inércia .....	12
2.2.3 Padrões de <i>speckle</i> objetivo e subjetivo .....	14
2.3 Técnicas estatísticas .....	16
2.3.1 Estatística multivariada .....	16
2.3.1.1 Análise de agrupamento .....	18
2.3.1.1.1 Distância euclidiana .....	23
2.3.1.2 Técnicas para a construção de conglomerados .....	25
2.3.1.2.1 Método das k-médias .....	26
2.3.2 Séries temporais .....	27
2.3.2.1 Teste do sinal .....	30
2.3.2.2 Teste de Fisher .....	31
2.3.2.3 Modelos para séries temporais .....	32
3 MATERIAL E MÉTODOS .....	36
3.1 Material .....	36
3.1.1 Softwares utilizados nas análises .....	37



3.2 Métodos .....	37
3.2.1 Análise de agrupamento.....	38
3.2.2 Séries temporais.....	39
4 RESULTADOS E DISCUSSÃO.....	41
4.1 Estatística multivariada: análise de agrupamento .....	41
4.2 Séries temporais: ajuste de modelos .....	47
4.2.1 Análise das séries dos grupos de média e de alta atividade .....	59
5 CONCLUSÕES .....	80
6 REFERÊNCIAS BIBLIOGRÁFICAS.....	81

## LISTA DE ABREVIATURAS

AR	Modelo autorregressivo
ARIMA	Modelo autorregressivo integrado de médias móveis
CCD	<i>Charge coupled device</i>
DSI	Interferometria digital de <i>speckle</i>
FAC	Função de autocorrelação
FACP	Função de autocorrelação parcial
LAG	Número de defasagens
MA	Modelo médias móveis
MI	Momento de inércia
MOC	Matriz de ocorrências
STS	<i>Spatial temporal speckle</i>
THSP	História temporal e espacial do padrão <i>speckle</i>

## RESUMO

COSTA, Ricardo Marques da. **Técnicas estatísticas aplicadas em imagens do *speckle* dinâmico**. 2009. 83 p. Dissertação (Mestrado em Engenharia de Sistemas) - Universidade Federal de Lavras, Lavras, MG.<sup>1</sup>

A análise de materiais por meio da iluminação laser e a aplicação de técnicas estatísticas de análise são de fundamental importância para diversos campos da ciência. Apesar da gama de métodos que possibilitam a captura e a análise dos dados de materiais, a técnica do *speckle* se destaca. Esta técnica consiste em estudar os efeitos da interação da luz, emitida por um laser, em uma superfície rugosa chamada de padrão de *speckle*, que pode ser de material biológico ou não. Para a análise do padrão de *speckle*, uma das abordagens é a utilização de uma imagem secundária, formada a partir de imagens sucessivas do *speckle* de um material, conhecida por matriz STS (*spatial temporal speckle*). Na análise destes STSs, várias podem ser as técnicas e os procedimentos para a obtenção de informações. Todavia, observam-se muitos esforços no sentido de incrementar a interpretação dos dados. Pensando nisso, este trabalho teve por finalidade aplicar a metodologia de séries temporais e as técnicas de análise de agrupamento da estatística multivariada, em um conjunto de células vivas, para facilitar a interpretação dos resultados gerados pela técnica do *bio-speckle*. Além disso, observar relações entre os dados pela realização de uma análise exploratória em algumas linhas que compõem cada matriz, para identificar se comportamentos, como tendência e perda de atividade, ocorrem em cada conglomerado formado. Os resultados mostraram que as técnicas de séries temporais e de análise multivariada foram aplicadas com sucesso ao conjunto de células vivas, possibilitando a redução da dimensão das estruturas e facilitando a interpretação das análises pela construção de clusters compostos de um número menor de informações.

Palavras-chave: *Bio-speckle*. Matriz STS. Análise multivariada. Séries temporais. Agrupamentos.

---

<sup>1</sup> Comitê orientador: Thelma Sáfiadi – UFLA (orientador) e Giovanni Francisco Rabelo – UFLA (coorientador).

## ABSTRACT

COSTA, Ricardo Marques da. **Statistic techniques applied in dynamic *speckle* images**. 2009. 83 p. Dissertation (Mastering in Systems Engineering) - Universidade Federal de Lavras, Lavras, MG.<sup>1</sup>

Analysis of materials through laser beams and application of statistical analysis techniques are now fundamentally important in several science-related fields. In spite of the wide range of methods which enable for the capture and analyses of material data, the *speckle* technique does stand out. Such technique consists of studying the effects of laser-emitted light beam interaction over a creased surface, called a *speckle pattern*, which may be biological material or otherwise. For the purpose of *speckle pattern* analyses, one of the possible approaches consists of the use of a secondary image, formed from successive *speckle* images of a certain piece of material, known as STS matrix (*spatial temporal speckle*). A large number of techniques and procedures may be used for obtaining information during such STS analyses. However, a great deal of effort in the sense of improving data interpretation has been observed. Taking this into consideration, this piece of work aimed at applying the temporal series methodology and the techniques of multi-varied statistical grouping analysis on a set of living cells, so as to enable for an easier interpretation of results generated by the *bio-speckle* technique, besides watching relations among the data through an exploratory analysis of some of the lines which compose each matrix, so as to identify whether behaviors, such as tendency and loss of activity, will or not occur in each formed conglomerate. Results allowed to verifying that the application of the temporal series and the multi-varied analysis techniques were successful as for the living-cells set, as they allowed for the reduction in structures dimension and enabled for an easier interpretation of the analyses by building clusters which comprise a smaller amount of information.

**Key-words:** *Bio-speckle*. STS matrix. Multi-varied analysis. Temporal series. Groupings

---

<sup>1</sup> Guidance Comité: Thelma Sáfadi – UFLA (Guide) and Giovanni Francisco Rabelo – UFLA (Co-guide).

## 1 INTRODUÇÃO

A análise de materiais biológicos por meio de sistemas de visão de máquina é de fundamental importância para diversos campos da ciência. Estes sistemas permitem analisar e determinar comportamentos, observar a relação entre variáveis e simplificar estruturas.

Apesar da existência dos vários sistemas capazes de capturar, analisar e processar informações, algumas limitações ainda precisam ser eliminadas e deficiências corrigidas para facilitar a interpretação dos resultados. Entre estas, destaca-se a necessidade da eliminação do contato entre o experimentador e a amostra, da diminuição da subjetividade na análise dos resultados e da redução da quantidade de dados trabalhados para diminuição do esforço computacional.

Assim, dos métodos baseados no fenômeno da interferência da luz, destacam-se algumas técnicas interferométricas ópticas, como a técnica do *speckle*. Estas, geralmente, são empregadas para projeção e análise de imagens tridimensionais (3D) e análise de superfícies e contornos de materiais estáticos e não-estáticos, bem como para a determinação da atividade celular, viabilidade de sementes e sêmen e qualidade de frutos.

As técnicas de análise de imagens pelo *speckle*, sobretudo a técnica do *bio-speckle* ou *speckle* dinâmico (aplicada em materiais biológicos), caracterizam-se pela versatilidade, facilidade no tratamento e na coleta dos dados e economia de recursos financeiros.

Além disso, a técnica do *bio-speckle*, segundo Braga Jr. (2000), destaca-se pela possibilidade de um caráter mais automatizado, independente do julgamento humano e, portanto, subjetivo, além de permitir maior velocidade ao mesmo e, conseqüentemente, economia de tempo no processo.

Em materiais biológicos, como a análise de sêmen bovino pela técnica do *bio-speckle*, esta técnica é empregada para determinar níveis de intensidade

de atividade celular. Para a realização da análise exploratória destes níveis de atividade, ferramentas e procedimentos estatísticos são testados, visando alcançar a diminuição do custo de interpretação dos resultados.

Assim, a união da técnica do *bio-speckle* com técnicas de análise, como a estatística multivariada e as técnicas de séries temporais, visa facilitar o entendimento das análises e informar sobre novas relações que podem estar ocorrendo.

Portanto, a utilização de ferramentas que apresentam uma visão mais global do fenômeno que aquela possível numa abordagem univariada torna possível, por meio de técnicas estatísticas, depurar, caracterizar e obter conclusões a respeito dos dados representativos das imagens de *bio-speckle*, sendo uma parceria potencial a ser desenvolvida.

Partindo do exposto, este trabalho teve a finalidade de aplicar a metodologia de séries temporais e as técnicas de análise multivariada no estudo de um conjunto de células vivas de sêmen bovino iluminadas por um raio laser, visando facilitar a interpretação dos resultados gerados pela técnica do *bio-speckle* e observar relações entre os dados.

De forma específica objetivou-se:

- a) reduzir a dimensão original das estruturas e facilitar a interpretação das análises realizadas por meio da construção de agrupamentos. Esta redução foi feita em uma matriz STS pela separação das suas 512 linhas em três grupos pré-definidos. Após a geração dos grupos, realizou-se uma análise exploratória sobre os mesmos para a verificação do comportamento de cada grupo gerado dentro deste STS em relação ao comportamento geral apresentado pelas 512 linhas;
- b) determinar comportamentos, como tendência e variações sazonais e realizar ajustes de modelos para linhas, tomadas dentro de cada um dos três grupos formados do STS tomado como amostra para os ensaios;

- c) avaliar o comportamento de três grupos em três STS consecutivos. Analisou-se como se comporta cada grupo do primeiro STS, em relação ao mesmo grupo nos outros dois STS consecutivos. Além disso, pretendeu-se verificar se o comportamento de queda dos índices de atividade de uma matriz STS, em relação aos STS consecutivos, também se observa nos grupos formados.

## 2 REFERENCIAL TEÓRICO

### 2.1 Interferometria digital de *speckle* (DSI)

Pesquisas voltadas para os fenômenos de interferência de luz e processamento digital de imagens baseados na visão de máquina têm aumentado ao redor do mundo. Estas pesquisas buscam a otimização de processos, a redução de custos e o aumento da produtividade.

Briers (2007) mostra que este desenvolvimento não é recente e enfatiza que, desde a década de 1980, pesquisadores têm buscado encontrar métodos de diagnóstico não-invasivos.

Silva & Muramatsu (2007) chamam a atenção para a construção do primeiro laser, em 1960 e afirmam que foi este o fato que deu impulso às pesquisas da época, possibilitando a aplicação da fonte de luz laser em diversas áreas. Entre estas, destacam-se as áreas de engenharia, ciências agrárias, biológicas e médicas, sendo possível realizar a prescrição da superfície de objetos, análise de tecidos e células e a manipulação de drogas para tratamento de patologias.

Dessa forma, várias técnicas surgiram ao longo do tempo, visando reduzir o grau de contato entre o experimentador e a amostra. Entre as mais utilizadas destaca-se a técnica do *speckle* laser que, segundo Briers (2007), caracteriza-se por apresentar um caminho simples para a captura de informações.

Briers (2007) destaca que as flutuações dos materiais causam um “borramento” do *speckle*, levando a uma redução do contraste do *speckle* local. Esta variação do contraste, segundo Briers (2007), é codificada como as distribuições de velocidade do fenômeno em questão e, conforme Rabelo (2000), pode auxiliar nos processos de identificação de alterações fisiológicas em frutos



e sementes e permitir o cálculo da intensidade da atividade em materiais, não se prendendo a uma análise superficial.

Além destas, Gale et al. (2004) destacam outras aplicações, como controle de envelhecimento de componentes mecânicos, estudos botânicos e acerca da secagem de pinturas, enfatizando o estudo do espalhamento da luz a partir de mudanças das superfícies, seja pela oxidação, erosão ou amostras biológicas.

Quanto à secagem de pinturas, Arizaga et al. (2006) relatam que alguns benefícios são gerados nas situações em que o tempo de secagem é pequeno, uma vez que muitos processos industriais envolvem a cobertura de substratos com finas camadas de tinta, visando à proteção, à decoração e à transmissão de propriedades como brilho, adesão e outras características magnéticas.

Arizaga et al. (2006) também reforçam que técnicas como a do contraste do *bio-speckle* são alternativas que permitem caracterizar quão rápido é este processo de secagem e que a aplicação de procedimentos usando esta técnica permite um estudo mais detalhado do sistema em questão.

O *bio-speckle*, ou *speckle* dinâmico, corresponde a um fenômeno óptico conhecido e acontece quando uma luz laser se dispersa sobre uma superfície na qual se desenvolve um processo vivo (Dainty, 1984).

Em razão disso, há uma classificação da interferometria digital de *Speckle*, em consequência da ligação com o material em análise, fazendo com que o fenômeno seja identificado por *speckle* e *bio-speckle*. *Speckle*, quando o processo fizer tratamento de amostras de objetos inertes, como um cone ou outro objeto geométrico construído de material estático, e por *bio-speckle*, ou *speckle* dinâmico, quando fizer tratamento de amostras de material vivo, como sêmen ou um conjunto de células vivas.

Entretanto, independente das denominações e do fato de as amostras corresponderem a materiais que possuem ou não atividade, Pires et al. (2007)

destacam que este é um fenômeno de interferência, tipicamente ondulatório, observável não apenas no visível, mas também em outras partes do espectro eletromagnético e na acústica.

Braga Jr. (2000) explana que, para análise da viabilidade de sementes, por exemplo, uma das etapas do controle de qualidade, a técnica do *bio-speckle* apresenta-se como potencial para a eliminação da subjetividade, a redução de tempo de análise e a automatização do processo de avaliação, situação ideal de julgamento buscada pelos programas de controle de qualidade.

Pomarico et al. (2004) complementam e ressaltam que, na biologia, é possível avaliar a motilidade de parasitas nematoides expostos a diferentes drogas anti-helmínticas.

Assim, é importante destacar que, independente da superfície ser plana ou não, esta possui microirregularidades em seu plano, oferecendo condições distintas de reflexão que, segundo Pomarico et al. (2004), possibilitam a formação dos diagramas de *speckle*.

Estes diagramas podem ser avaliados, segundo Oulamara et al. (1989), pela utilização de uma imagem que se refere à história temporal e espacial do padrão de *speckle*, chamada de matriz STS ou de THSP. Esta matriz STS é uma imagem formada a partir da retirada sucessiva de colunas das imagens originais do material sob avaliação.

Assim, quando imagens sucessivas do *speckle* são obtidas ao longo do tempo, obtém-se a variação temporal e espacial de cada pixel, naquele determinado instante. No caso da análise de um material biológico, esta variação corresponderá à variação dos níveis de atividade.

De posse dessas imagens (que foram convertidas para o formato de arquivo txt), se for retirada uma mesma coluna de cada imagem capturada ao longo do tempo, haverá a formação de uma nova imagem, formada a partir das colunas das imagens originais. Esta nova imagem NxM é chamada de matriz

STS ou THSP e representa as N colunas ao longo do tempo. Normalmente, o STS tem uma composição  $N=M$  e, de acordo com Xu et al. (1995), é possível adotar estatísticas de segunda ordem, por meio de funções de autocorrelação. Conforme Nascimento et al. (2007), estas vão desde funções em que os coeficientes caem lentamente até aquelas funções nas quais o coeficiente cai rapidamente, sendo capazes de simular diferentes níveis de atividade.

A imagem de um STS formado a partir de imagens de sêmen bovino pode ser observada na Figura 1.

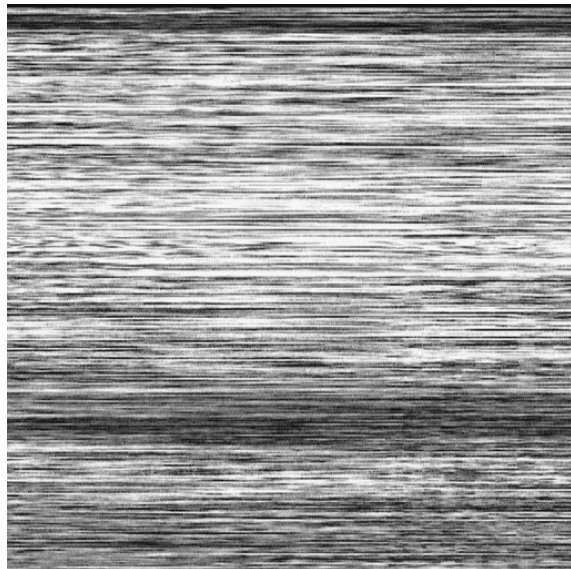


FIGURA 1 Matriz STS gerada a partir de imagens de sêmen bovino

Pela Figura 1 é possível observar a imagem de uma matriz STS formada pela captura de imagens de sêmen bovino.

Braga et al. (2008) explicam que, para os diferentes níveis de atividade no STS, as variações de intensidade dos pixels são mostradas na direção horizontal das diferentes colunas, enquanto a variação espacial, na direção vertical, que são as linhas.

A análise do STS, mostrado na Figura 1, permite diagnosticar comportamentos e propor modelos para as linhas que compõem a estrutura, possibilitando facilitar a interpretação dos resultados e o entendimento do fenômeno ocasionado pela interação da luz laser e o material sob amostra.

Em suma, o laser geralmente utilizado nas pesquisas realizadas por físicos e considerado, muitas vezes, uma ferramenta complexa e pouco popular, tem se adaptado satisfatoriamente aos desenvolvimentos tecnológicos e permitido a geração de aplicações e produtos científicos e comerciais (Nobre 2008).

## **2.2 Aspectos relevantes do padrão de *speckle***

### **2.2.1 Propriedades do padrão de *speckle***

Segundo Bergkvist (1997), o *bio-speckle* é criado em razão da luz (laser) ser espalhada por partículas em movimento e ser modulado pelo estado de movimentação dos *scatterers* ou elementos difusores. Este fato caracteriza o *speckle* variante no tempo, referenciando-o por temporal.

Bergkvist (1997) afirma, ainda, que o *bio-speckle* apresenta a aparência de uma espécie de líquido em ebulição, podendo ser referenciado também como “*speckle* borbulhante”. Segundo ele, quando a luz penetra num determinado objeto, esta é espalhada em todas as direções pelos elementos difusores antes de deixar o objeto, fazendo com que não seja possível reconhecer e estudar cada partícula que espalha a luz. Além disso, em cada ponto da imagem do *bio-speckle*, a luz origina-se de muitos elementos difusores que estão superpostos, o que faz com que a complexidade seja aumentada.

Pela Figura 2 pode-se constatar que o fenômeno é muito complexo e pode ser comparado com o gráfico da soma vetorial.

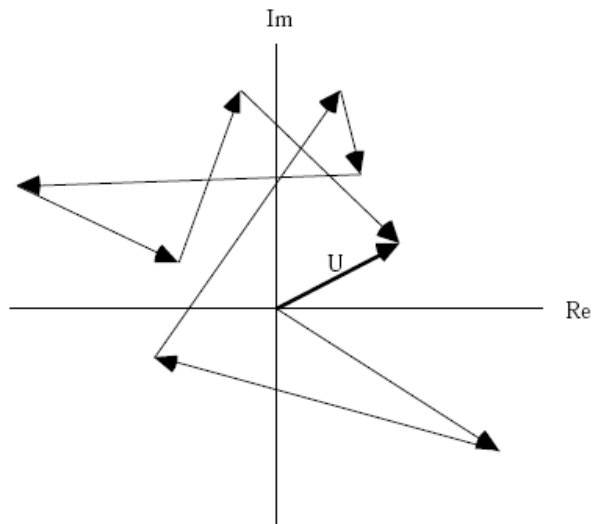


FIGURA 2 Exemplo de soma vetorial para criar uma área em um padrão *speckle*  
 Fonte: Bergkvist (1997, p. 17)

Pela Figura 2 é possível observar que a frequência das flutuações do *speckle* é diretamente relacionada à taxa de variação do passeio aleatório e, conseqüentemente, a velocidade de cada elemento difusor.

Outros fatores, como o tamanho do grão de *speckle*, contribuem para a complexidade do fenômeno gerado. Esta medida é de suma relevância e, segundo Rabelo (2000), relata que as dimensões de interesse do *speckle* são o tamanho de cada grão e sua luminosidade.

Xu et al. (1995) demonstraram que o *speckle* produzido pelo espalhamento dentro de um objeto tem um tamanho médio menor do que aqueles produzidos na superfície. Isto ocorre devido ao fato de a luz laser sofrer um espalhamento ao penetrar o objeto. Como a luz é refletida de volta, ela agora sairá de uma região do objeto de uma área maior do que quando ela tinha penetrado.

Bergkvist (1997) mostra, pela Equação 1, que o tamanho do grão de *speckle* corresponde ao diâmetro do local iluminado, conhecido por disco de Airy.

$$d = 2 * 1,22 \left( \frac{\lambda * z}{D} \right) \quad (1)$$

em que:

$\lambda$ : comprimento de onda;

z: distância da observação;

D: diâmetro da área circular observada.

Assim, aumentando o valor de D, é possível confirmar a relação destacada por Xu et al. (1995) ao inferirem sobre relacionamento entre área de uma superfície e o tamanho dos grãos de *speckle*.

Além disso, Drain (1980), citado por Bergkvist (1997), destaca que, devido à complexidade da luz espalhada, não há métodos matemáticos precisos para descrever este fenômeno. Mas, os diferentes vetores de ondas, os vetores de velocidade e os múltiplos espalhamentos, todos juntos, podem ser interpretados estatisticamente como uma expansão do efeito Doppler para luz espalhada.

### **2.2.2 Estatísticas do padrão de *speckle***

Segundo Rabal et al. (2008), o uso do *bio-speckle* como uma técnica de metrologia foi consolidado por testes sistemáticos de todas as abordagens sugeridas em muitas aplicações.

Rabal et al. (2008) explicam que é possível, por meio da média de variáveis biológicas ou físicas, obter informações úteis que permitam o

observador avaliar, por meio de números ou imagens processadas, as variáveis sob controle.

Bergkvist (1997) complementa e afirma que o *speckle* é um fenômeno estocástico que pode ser descrito estatisticamente. Além disso, que há várias aproximações estatísticas e que algumas delas são avaliações dos conceitos de estatística de primeira ordem e de segunda ordem.

A estatística de primeira ordem esclarece as propriedades de um padrão de *speckle* ponto a ponto, sem considerar as relações entre os diversos pontos distintos. E a estatística de segunda ordem descreve quão rápido varia o fenômeno, permitindo calcular o tamanho do grânulo de *speckle* e sua distribuição no padrão (Henão et al., 1997).

Briers (1993), citado por Rabelo (2000), afirma que o contraste de *speckle* demonstra que o desvio padrão  $\sigma$  da intensidade espacial é a média do padrão de *speckle*  $\langle I \rangle$  e que o contraste é expresso pela relação entre a variância e a intensidade média quadrática, que podem ser indicados, respectivamente, pelas Equações 2 e 3:

$$\sigma = \langle I \rangle \quad (2)$$

$$C = \frac{\sigma^2}{\langle I \rangle^2} \quad (3)$$

Segundo Rabal et al. (2008), o contraste *speckle*, em relação ao tempo, como estatística de primeira ordem, é seguido por métodos estatísticos de segunda ordem que são avaliados utilizando-se as imagens STS ou os THSPs. Entre eles, citam as funções de autocorrelação, cumulantes estatísticos e o método para cálculo do momento de inércia, conhecido por MI.

Rabal et al. (2008) afirmam que a estatística temporal de segunda ordem é o tratamento mais popular que pode ser utilizado para medir a velocidade ou a motilidade dos *scatterers* em objetos biológicos ou não-biológicos e é implementado utilizando-se dispositivos, como câmeras CCDs<sup>1</sup> ou fotodetectores, que obtêm informações em uma ou duas dimensões.

### 2.2.2.1 Momento de inércia

A utilização de estatísticas de segunda ordem para a caracterização de material biológico iluminado por uma luz laser é uma prática muito utilizada e que permite o exame das variáveis presentes no conjunto analisado.

O momento de inércia (MI), segundo Arizaga et al. (1999), mostrado por Rabal et al. (2003), é a medida que indica com que frequência ocorrem mudanças bruscas de intensidade na matriz STS e, geralmente, é utilizada para caracterizar os padrões de *speckle*. Esta medida exibe altos valores quando a amostra apresentar alta atividade e baixos valores em situação contrária.

Rabelo (2000) informa que as informações de uma imagem podem ser caracterizadas por seu nível de intensidade, conhecido como níveis de cinza, sua distribuição espacial e sua distribuição de probabilidades.

Quando um determinado material é iluminado, uma matriz STS pode ser criada para representar as informações acerca da sua atividade. Assim, diferentes materiais apresentarão diferentes STSs de acordo com os níveis de atividade.

A iluminação de uma fruta em condições normais e de outra em início de processo de apodrecimento mostrará duas matrizes STS, respectivamente, com menores e maiores níveis de atividade.

A medição destes níveis de atividade é importante como forma de se aplicar um sistema de medição em que o julgamento deixa de ser subjetivo e

---

<sup>1</sup> CCD ou dispositivo de carga acoplado é um sensor para a captação de imagens, formado por um circuito integrado, contendo uma matriz de capacitores ligados (acoplados).



visual e passa a ser representado por valores numéricos que podem ser utilizados como ferramentas de metrologia e análise.

A matriz STS é representada por pixels em níveis de cinza codificados em 8 bits, que configuram, no total, 256 tons. Assim, O MI representa o processo que se baseia nas ocorrências de valores de intensidade sucessivos dos pixels que compõem a imagem STS. O STS é transformado em uma matriz denominada matriz de ocorrências (MOC).

A MOC é definida por entradas, que são o número de ocorrências de um valor de intensidade  $i$ , seguido por um valor de intensidade  $j$  e pode ser definida de acordo com a Equação 4:

$$MOC = [N_{ij}] \quad (4)$$

em que:

$N_{ij}$  é o número de ocorrências de intensidades;

$i$  e  $j$  são as intensidades sucessivas.

Na MOC,  $N_{ij}$  representa o número de ocorrências, na imagem STS, de certa intensidade de cinza  $i$  seguida, imediatamente no próximo instante de tempo, por um nível de cinza de valor  $j$ .

Se um material apresentar baixa atividade, sua matriz de ocorrências é caracterizada por pixels que apresentam tons de cinza com pouca variação, ou seja, a mudança de intensidade do pixel  $i$  para o pixel  $j$  será pequena.

Entretanto, se o material a ser analisado apresentar alta atividade será possível observar um efeito contrário e um espalhamento, indicando uma mudança mais intensa de intensidade do pixel  $i$  para o pixel  $j$ .

### 2.2.3 Padrões de *speckle* objetivo e subjetivo

Jones et al. (1989), citados por Pires et al. (2007), afirmam que estruturas de *speckle* podem ser obtidas de diferentes maneiras, mas destacam que as duas principais formas de obtenção são classificadas como objetiva e subjetiva.

Na Figura 3 são mostrados, esquematicamente, os arranjos ópticos para a gravação de padrões de *speckle* objetivo e subjetivo, respectivamente.

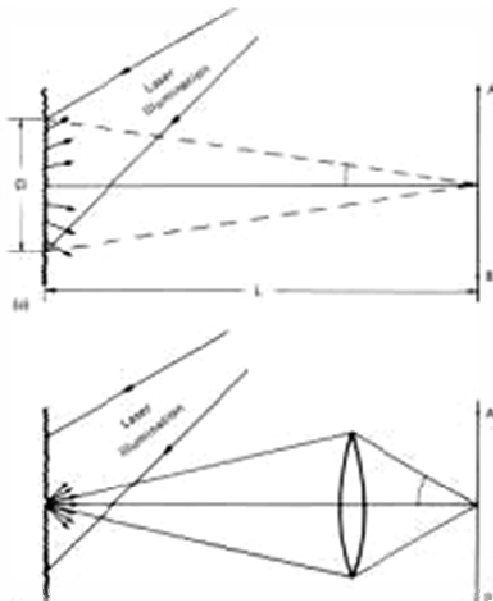


FIGURA 3 Arranjo óptico para a formação de *speckle* objetivo e subjetivo  
Fonte: Pires et al. (2007, p.5)

Pela Figura 3 observa-se que quando um ponto de uma superfície rugosa é iluminado e espalha luz em todas as direções é possível afirmar que, na gravação de padrões de *speckle* objetivo, o ponto de observação é iluminado por ondas espalhadas de cada porção de toda superfície. Porém, na gravação de padrões de *speckle* subjetivo, o ponto de observação é iluminado apenas pelos

raios emergentes da lente que convergem para o ponto de observação, os quais têm sua origem em uma pequena região da superfície iluminada.

Em Rabal et al. (2008), o padrão de *speckle* objetivo é denominado “Free propagation”, referindo-se ao caso em que o detector da câmera registra a imagem do *speckle* pela propagação livre da onda. Esta propagação livre é capturada sem lentes, ao contrário do padrão subjetivo, em que, para a captura da imagem da amostra, há a necessidade de uma lente de ajuste no sistema óptico.

Jones et al. (1989), citados por Pires et al. (2007), destacam que a principal diferença entre os padrões de *speckle* mostrados na Figura 3 é que um pequeno fragmento do padrão de *speckle* objetivo contém informação de toda superfície iluminada do objeto, enquanto uma pequena seção do padrão de *speckle* subjetivo corresponde a uma parte definida da superfície do objeto.

Pires et al. (2007) explicam que o *speckle* é chamado de objetivo porque sua escala depende somente do plano no espaço onde é observado e que o termo *speckle* subjetivo é utilizado porque o tamanho dos *speckles* no plano da imagem é dependente da abertura da lente de observação. Segundo ele, estes efeitos podem ser observados comparando-se o tamanho dos *speckles* quando uma imagem é observada diretamente pelo olho e quando uma abertura menor do que a pupila do olho é colocada em frente ao mesmo. Neste último caso, o tamanho do *speckle* será visto maior, assumindo que o tamanho do *speckle* é tido como a separação entre os dois mínimos da função de Bessel, que pode ser mostrada pela Equação 1.

Com base na retina do olho humano, é possível perceber que a mesma trabalha de forma semelhante aos sistemas de visão de máquina. Além disso, que esta age como se fosse um plano receptor de luz, semelhante ao plano de reflexão da Figura 3. Dessa forma, a partir da interação entre muitas ondas de diferentes partes do local iluminado, cada uma com uma fase diferente, a

intensidade da onda resultante consistirá de uma soma complexa, como mostrado na Figura 2 e pela Equação 1 (Bergkvist, 1997).

Assim, Rabal et al. (2008) afirmam que é de suma importância considerar qual a influência exercida pelo modo de captura da imagem sobre o padrão de *speckle* adquirido.

Além disso, Rabal et al. (2008) chamam a atenção para a situação em que o sistema óptico é desfocado em ambos os casos. Se isto ocorre, cada ponto da imagem recebe informação de toda área da mostra, que depende do grau de desfocagem.

Dessa forma, é possível notar que, apesar da existência das diversas técnicas e procedimentos de análise e processamento digital de imagens, diversos fatores e variáveis podem estar envolvidos, sendo algumas não observáveis e de difícil mensuração.

## **2.3 Técnicas estatísticas**

### **2.3.1 Estatística multivariada**

Em relação aos procedimentos de análise, verificação e certificação, é necessário considerar que, no geral, estes procedimentos trabalham com um volume muito grande de informações. Assim, tanto no mercado quanto nas pesquisas desenvolvidas nas universidades, os processos de experimentação e julgamento têm a necessidade de serem tratados em tempo polinomial sem perder, contudo, a eficiência.

Assim, devido este alto volume de informações, os métodos precisam mostrar a capacidade de obter informações de acontecimentos e dos fenômenos que estão sendo analisados para converter uma grande massa de dados em conhecimento.

Além disso, é de suma importância a utilização de ferramentas que apresentem uma visão mais global do fenômeno do que aquela obtida em uma

abordagem univariada. Manly (1995) afirma que o ponto a ser considerado na análise multivariada é avaliar as diversas variáveis simultaneamente, cada uma sendo considerada igualmente importante, pelo menos no início.

Por meio de técnicas estatísticas, como a análise estatística multivariada, pode-se depurar, caracterizar e obter conclusões a respeito de valores representativos de amostras diversas, de materiais biológicos ou não.

Segundo Ferreira (2008), o problema de se inferir em trabalhos científicos, a partir de dados mensurados pelo pesquisador, sobre processos, fenômenos físicos, biológicos ou sociais, que não se pode diretamente observar, é uma realidade constante.

Anderson (1971) reforça que as medidas e análises de dependência entre variáveis, entre conjunto de variáveis e entre variáveis e conjunto de variáveis são fundamentais para exames multivariados.

Portanto, ao contrário de outras áreas, que não consideram todos os fatores envolvidos, a estatística multivariada corresponde a uma metodologia que estuda os fenômenos, analisando suas diversas variáveis simultaneamente.

Além disso, cabe ressaltar que as respostas mensuradas nos fenômenos estudados são obtidas em diversas variáveis aleatórias, fazendo com que as técnicas de análise sejam capazes de lidar com muitos dados de forma simples e objetiva.

Atualmente, existem diversos métodos que empregam a estatística multivariada. Dentre eles, destacam-se a técnica de componentes principais, a análise de fatores e a análise de agrupamento (*cluster analysis*).

Hair (1995) complementa que acadêmicos e pesquisadores frequentemente enfrentam situações que podem ser resolvidas pela definição de grupos de objetos homogêneos. Estes objetos podem ser, segundo Hair (1995), indivíduos, empresas ou, até mesmo, comportamentos, que podem ser agrupados e segmentados por meio de uma metodologia objetiva.

Em suma, Hair (1995) afirma que, em todas as instâncias, o analista procura por uma estrutura “natural” entre as observações, baseadas num perfil multivariado e que o método mais comumente utilizado para esta proposta é a análise de cluster.

### **2.3.1.1 Análise de agrupamento**

De acordo com Mingotti (2005), a análise de agrupamento (também é conhecida por análise de conglomerado, classificação ou cluster) tem o objetivo de dividir os elementos de uma amostra ou população em grupos. Mingotti (2005) destaca, ainda, que os elementos pertencentes a um mesmo grupo devem ser o mais parecido possível entre si e elementos pertencentes a grupos diferentes o mais heterogêneo possível, de acordo com as variáveis que foram medidas.

A análise de agrupamento é de suma importância em diversas áreas. Entre elas, destacam-se as áreas médicas, as ciências agrárias e humanas.

Nas áreas médicas é possível separar de um conjunto total, subgrupos de pacientes, de acordo com sintomas e patologias. Nas ciências agrárias, McGarigal et al. (2002) ressaltam a aglomeração de espécies em subgrupos para classificação e ordenação. Speece et al. (1985), citados por Mingotti (2005), afirmam que, na psicologia e na psiquiatria, é possível apresentar perfis de personalidades e mostrar grupos de pessoas que apresentam uma semelhança comportamental ou indivíduos com personalidades agressivas, de acordo com determinadas características.

Mingotti (2005) afirma que a análise de agrupamentos também é muito popular em *Data Mining* (mineração de dados) para analisar e, por meio de ferramentas computacionais, buscar padrões em conjunto de dados.

Dessa maneira, ao analisar um conjunto de dados multivariados, as variáveis envolvidas podem ser classificadas de acordo com o esquema mostrado na Figura 4.

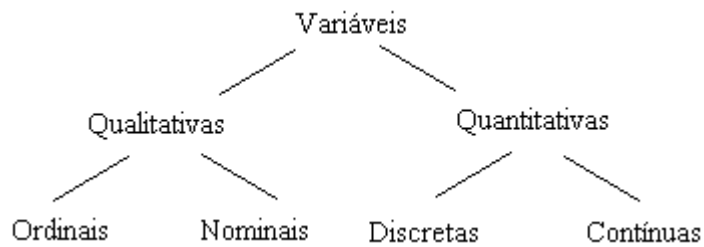


FIGURA 4 Divisão e classificação dos tipos de variáveis existentes

Na Figura 4 pode-se observar que o conjunto de variáveis pode ser dividido, classificado, em duas categorias: qualitativas e quantitativas, e estas, por sua vez, divididas cada uma em dois subgrupos. O conjunto das variáveis qualitativas pode ser separado em ordinais e nominais e o conjunto das variáveis quantitativas, em discretas e contínuas.

Segundo Ferreira (2008), as variáveis qualitativas ordinais, ao contrário das nominais, podem ser classificadas de acordo com uma ordem de grandeza. Conforme Ferreira (2008), é difícil classificar uma cor como sendo melhor ou pior que outra cor, mas é possível qualificar um carro classe A como sendo melhor que um carro classe D.

Ferreira (2008) diz, ainda, que as quantitativas discretas assumem, em geral, valores inteiros dentro de um intervalo finito ou infinito de valores, com valores predefinidos em intervalos de tempo  $t$ . As variáveis quantitativas contínuas são mensuradas em escala real como peso, altura e volume. Assim, a geração de um eletrocardiograma para um paciente em um intervalo de tempo  $\Delta t$  pode ser classificada como contínuo, já que se observa um comportamento dentro de um intervalo de tempo e não para cada instante  $t$ .

Desse modo, por meio do conhecimento dos tipos de variáveis envolvidas, é possível agrupar elementos de uma massa de dados ou de uma população com base em medidas de parença ou diferença.

A utilização do critério de parença consiste em agrupar elementos que são semelhantes entre si (medidas de similaridade) e as medidas de dissimilaridade caracterizam-se por agrupar indivíduos que são diferentes entre si.

De acordo com Johnson & Wichern (1992), no caso multivariado, o número  $p$  de variáveis pode ser mostrado por  $p > 1$  e deve-se, para isso, utilizar a noção de vetor aleatório para representar cada unidade amostral.

Mingotti (2005) afirma que, em um conjunto de dados constituído de  $n$  elementos amostrais e tendo-se medido  $p$ -variáveis em cada um deles, o objetivo será agrupar esses elementos em  $g$  grupos. Dessa forma, para cada elemento amostral  $j$ , tem-se, portanto, o vetor de medidas  $X_j$  definido por:

$$X_j = [x_{1j} x_{2j} \dots x_{nj}]$$

Ferreira (2008) afirma que num ensaio existem unidades amostrais ou experimentais que podem ser descritas em função de vetores de variáveis aleatórias:

$$\begin{array}{cccccc} U_1 & U_2 & \dots & U_j & \dots & U_n \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ y_{11} & y_{21} & \dots & y_{j1} & \dots & y_{n1} \\ y_{12} & y_{22} & \dots & y_{j2} & \dots & y_{n2} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ y_{1p} & y_{2p} & \dots & y_{jp} & \dots & y_{np} \end{array}$$



Na relação mostrada acima,  $y_{ij}$  representa a  $j$ -ésima variável para a unidade  $U_i$ . Assim, a observação  $y_{11}$  representa a observação para a variável 1 da unidade amostral 1.

Johnson & Wichern (1992) afirmam que uma forma de representar todo o conjunto multivariado contendo as  $p$ -variáveis mensuradas nas  $n$  unidades amostrais é por meio de uma matriz de dados ( $Y$ ) de dimensões ( $n \times p$ ):

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & \cdots & Y_{1k} & \cdots & Y_{1p} \\ Y_{21} & Y_{22} & \cdots & Y_{2k} & \cdots & Y_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_{j1} & Y_{j2} & \cdots & Y_{jk} & \cdots & Y_{jp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ Y_{n1} & Y_{n2} & \cdots & Y_{nk} & \cdots & Y_{np} \end{bmatrix}$$

Dessa forma, cada linha da matriz corresponde a um vetor  $p$ -dimensional de observações multivariadas, apresentado na sua forma transposta e cada coluna, um vetor  $n$ -dimensional, das observações de uma determinada variável, sendo  $Y_{jk}$  a representação da  $j$ -ésima unidade amostral na  $k$ -ésima variável. Assim, na matriz  $Y$ ,  $Y_{12}$  representa a unidade amostral 1 na segunda variável.

Em um sistema com plantas de uma determinada cultivar, é possível separar as variáveis altura, produtividade ou diâmetro do colmo. Estas variáveis estarão correlacionadas, mas as observações, em cada unidade amostral, serão independentes.

Consequentemente, Mingotti (2005) afirma que, para decidir até que ponto dois elementos de um conjunto de dados podem ser considerados como semelhantes ou não, é necessário utilizar medidas matemáticas (métricas), como o conceito de distâncias. Estas medidas permitem calcular as distâncias entre os

vetores de observações dos elementos amostrais e agrupar aqueles de menor distância.

Contudo, Mingotti (2005) destaca que, para que o agrupamento possa ser realizado, é necessária a decisão do critério de agrupamento a ser utilizado: medida de parença ou similaridade e diferença ou dissimilaridade.

Sprent (1993), citado por Mingotti (2005), afirma que existem algumas medidas de similaridade específicas para variáveis categóricas (qualitativas), como o coeficiente qui-quadrado, o de contingência de Person e o de concordância de Kappa.

Para as variáveis qualitativas, é possível encontrar também algumas técnicas estatísticas para o cálculo da medida de similaridade. Dentre estas se destacam a correlação de Person e a técnica do cosseno, em que é possível prever que quanto maior o valor absoluto de retorno encontrado, maior será a proximidade entre as variáveis e, conseqüentemente, mais linearmente relacionadas elas estarão.

Além destas, como exposto por Sprent (1993), citado por Mingotti (2005), coeficientes de associação não-paramétricos, como o de Spearman e o de Kendall, também podem ser utilizados para o agrupamento de variáveis quantitativas.

Entretanto, apesar da variedade de procedimentos para calcular a medida de similaridade entre grupos de objetos, é usual a utilização de medidas de dissimilaridade para variáveis quantitativas. Estas apresentam valores pequenos para elementos mais parecidos e valores maiores, em caso contrário.

Entre as medidas de dissimilaridades mais comuns, é necessário destacar a distância generalizada ou ponderada, a distância de Minkowsky e a distância euclidiana, sendo esta última uma das mais utilizadas.

### 2.3.1.1.1 Distância euclidiana

A distância euclidiana entre dois elementos  $X_l$  e  $X_k$ ,  $l \neq k$ , é definida, segundo Mingotti (2005), pela Equação 5:

$$d(X_l, X_k) = [(X_l - X_k)'(X_l - X_k)]^{\frac{1}{2}} = \left[ \sum_{i=1}^p (x_{il} - x_{ik})^2 \right]^{\frac{1}{2}} \quad (5)$$

Dessa forma, os dois elementos amostrais são comparados em cada variável pertencente ao vetor de observações.

Mingotti (2005) demonstra que as distâncias entre os elementos amostrais são armazenadas numa matriz de dimensão  $n \times m$ , chamada matriz de distâncias, como exemplificado na Figura 5:

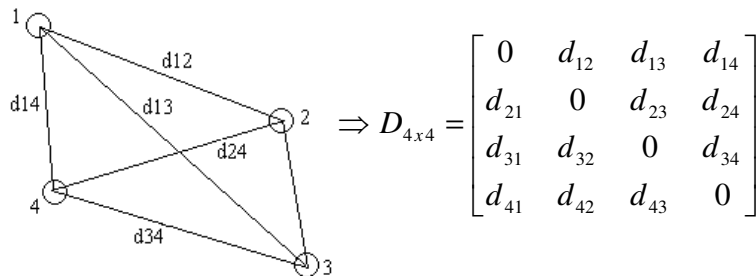


FIGURA 5 Esquema de armazenamento de distâncias numa matriz 4x4  
Fonte: Adaptado de Mingotti (2005)

Ainda de acordo com Mingotti (2005), é possível exemplificar o cálculo da distância euclidiana, tomando a renda mensal (em quantidade de salários mínimos) e a idade de seis indivíduos de uma localidade. Na Tabela 1 observam-se cada indivíduo, sua renda e sua respectiva idade.

TABELA 1 Renda e idade de seis indivíduos

Indivíduo	Renda	Idade
A	9,60	28
B	8,40	31
C	2,40	42
D	18,20	38
E	3,90	25
F	6,40	41
Média	8,15	34,17
Desvio padrão	5,61	7,14

Fonte: Adaptado de Ferreira (2008)

De acordo com a Tabela 1, a distância euclidiana, entre os indivíduos A e B, nas variáveis renda e idade é igual a:

$$d(X_A, X_B) = \left[ (9,60 - 8,40)^2 + (28 - 31)^2 \right]^{\frac{1}{2}} = 3,23$$

Pelo cálculo da distância,  $d_{ij}$  representa a distância do elemento amostral  $i$  ao elemento amostral  $j$ .

Assim, a distância euclidiana constitui uma das principais formas de se mensurar o grau de proximidade de um indivíduo ou objeto em relação a outro, permitindo a formulação de hipóteses e a confecção de inferências ou proposições acerca dos objetos componentes de um sistema.

### 2.3.1.2 Técnicas para a construção de conglomerados

É possível encontrar duas técnicas principais para a construção de clusters. Estas são classificadas, de acordo com Mingotti (2005), como técnicas hierárquicas e não-hierárquicas, de acordo com a Figura 6.

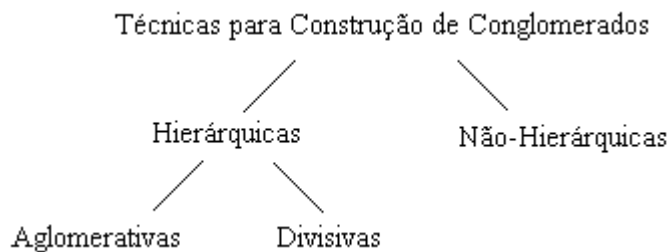


FIGURA 6 Principais técnicas para a construção de conglomerados  
Fonte: Adaptado de Mingotti (2005)

Na Figura 6, as técnicas hierárquicas ainda se subdividem em aglomerativas e divisivas.

As técnicas hierárquicas são utilizadas, na maioria das vezes, em análises exploratórias para se determinar agrupamentos. Estas visam agrupar uma grande quantidade de dados em  $k$  grupos pré-definidos, em que o número total de clusters é conhecido a priori.

As técnicas hierárquicas aglomerativas partem do princípio de que, no início, cada observação constitui um conglomerado isolado e, em cada passo do algoritmo, os elementos vão sendo agrupados, até o momento no qual todos os elementos considerados estão num único grupo (Mingotti, 2005).

Nas técnicas não-hierárquicas para que o agrupamento seja realizado, é necessário que o número de grupos seja especificado antes de se iniciar as análises.

Além da especificação prévia do número de grupos, outra característica que diferencia um agrupamento não-hierárquico de um hierárquico é o fato de

que dois elementos colocados inicialmente juntos num mesmo grupo não necessariamente estarão juntos na partição final. Isto se deve ao fato de que, em cada estágio do agrupamento, os novos grupos serão formados pela divisão ou junção de grupos já combinados em passos anteriores. Esta recombinação e permuta de elementos entre os grupos são para que os mesmos atendam a dois requisitos básicos: “coesão” interna (semelhança interna) e isolamento (ou separação) dos clusters formados (Mingotti, 2005).

Mingotti (2005) destaca, ainda, que os métodos não-hierárquicos utilizam algoritmos iterativos e, em comparação com métodos hierárquicos, apresentam maior capacidade de trabalhar com um volume grande de informações.

Os principais métodos de agrupamento baseado no modelo não-hierárquico são o método das k-médias, o Fuzzi c-Means e as redes neurais artificiais aplicadas à análise de agrupamento. Dentre estes, destaque para o algoritmo de k-médias, eficiente por trabalhar com uma grande quantidade de dados.

#### **2.3.1.2.1 Método das k-médias**

Segundo Ferreira (2008), o método das k-médias é o mais popular e é aplicado à matriz de dados  $Y$ . Conhecendo a priori o número de grupos  $k$ , as observações devem ser realocadas aos  $k$  grupos de acordo com uma função objetivo e esta realocação deve ser cessada quando uma regra de parada pré-especificada for contemplada.

O método das k-médias faz com que cada elemento amostral seja armazenado ao cluster cujo centroide é o mais próximo do vetor de valores observados para um respectivo elemento.

Mingotti (2005) e Ferreira (2008) destacam que há quatro passos básicos que o algoritmo das k-médias utiliza. São eles:

a) inicialmente, os  $n$  objetos são alocados arbitrariamente aos  $k$  grupos e os centroides ou “sementes” calculados para iniciar o processo de partição;

b) cada dado é, então, comparado com cada centroide, por meio de uma medida de distância que, em geral, é a euclidiana. O dado é alocado, então, ao grupo cuja distância é menor;

c) após aplicar o passo  $b$  para cada um dos dados amostrais, recalcula-se o centroide para cada grupo. Assim, o passo  $b$  deve ser repetido considerando os centroides destes novos grupos;

d) repetem-se os passos  $b$  e  $c$  até que não ocorram mais mudanças de objetos de um grupo para outro, ou seja, até que nenhuma realocação de elementos seja necessária.

Ferreira (2008) mostra e exemplifica o método das  $k$ -médias para o agrupamento de seis sistemas de uso da terra dos solos da Amazônia avaliados de acordo com as variáveis areia e argila.

### 2.3.2 Séries temporais

Outra ferramenta muito utilizada para análise de uma massa de dados são as ferramentas de séries temporais. Morettin & Tolo (2004) definem uma série temporal como sendo qualquer conjunto de observações ordenadas no tempo. São exemplos de séries temporais os índices diários medidos na bolsa de valores, precipitação atmosférica anual na cidade de Fortaleza, número médio anual de manchas solares e registro de marés no porto de Santos.

Assim, uma série temporal pode ser descrita de acordo com uma trajetória qualquer denotada por  $Z(t)$ , em que, para cada  $t$  fixo, têm-se valores de uma variável aleatória  $Z(t)$ , com certa distribuição de probabilidades.

Apesar de  $t$  ser referido, geralmente, como o tempo, Morettin & Tolo (2004) destacam que a série  $Z(t)$  pode ser construída em função de outro parâmetro físico, como espaço ou volume. Os mesmos autores também afirmam

que, se um índice for associado às observações, como a produção de café no mês, a série será discreta. Porém, se  $t$  pertencer a um intervalo e não a um índice, ou seja, se  $t$  variar de 0 a um determinado tempo, a série será contínua.

Entretanto, independente do parâmetro utilizado ou do fato de a série ser discreta ou contínua, cabe ressaltar que os dados de uma série são correlacionados e que uma observação  $Y_t$  pode ser influenciada por outra  $Y_{t-1}$  e  $Y_{t-2}$ .

Assim, inúmeras áreas podem utilizar os recursos das ferramentas de séries temporais, como em economia e finanças, oceanografia, meteorologia, engenharias e medicina, entre outras.

Em razão disso, a utilização destes métodos pode ter objetivos diversos. Morettin & Toloí (2004) enumeram os mais importantes:

- a) investigar o mecanismo gerador da série;
- b) fazer previsões para valores futuros da série;
- c) descrever comportamentos, como tendências, variações sazonais, presença de ciclos ou diagramas de dispersão;
- d) procurar periodicidades relevantes nos dados, como a análise espectral.

De tal modo, se forem tomadas as observações de  $Z(t)$  para  $t$  variando de 1 a  $n$ , um modelo com propósito determinado pode ser escrito para explicar o comportamento da mesma.

Estes modelos são escritos em função da soma de três componentes não-observáveis, descritos por Morettin & Toloí (2004), conforme a Equação 6:



$$Z_r = T_t + S_t + a_t \quad (6)$$

em que:

$T_t$  e  $S_t$  representam, respectivamente, tendência e sazonalidade;

$a_t$  representa a componente aleatória.

A componente tendência pode ser definida como uma força ou orientação que imprime determinado movimento ou orientação aos dados da série. E sazonalidade como sendo flutuações ocorridas ao longo de um determinado período de tempo, ou seja, observações que se repetem a cada  $k$  elementos, devido a fatores exógenos.

A estimação das componentes tendência e sazonalidade é de suma importância para ajuste de modelos, visto que, em geral, são bastante relacionadas e a influência da tendência sobre a componente sazonal pode ser muito forte.

A estimação da tendência e da sazonalidade permite obter a série ajustada para a tendência e a sazonalidade ou livre de tendência e sazonalidade (condição de estacionaridade).

Para testar a hipótese da tendência, destaca-se o método do sinal (Cox-stuart) e, para o teste da sazonalidade, o teste de Fisher, como os mais utilizados.

A retirada da tendência pode ser feita de várias formas, dentre as quais se destaca a o método das diferenças, que pode ser exemplificado pela Equação 7:

$$\Delta^b Z(t) = \Delta(\Delta^{b-1} Z(t)) \quad (7)$$

em que,  $\Delta Z(t) = Z(t) - Z(t-1)$

sendo:

$\Delta$ : operador diferença

$Z(t)$ : série analisada

Tendo-se uma tendência linear, por exemplo,  $T(t) = \beta_0 + \beta_1 t$ , é possível escrever que  $Z(t) = \beta_0 + \beta_1 t + a_t$ .

em que:

$T(t)$ : coordenada y (tendência)

t: coordenada x

$\beta_0$ : coeficiente linear y

$\beta_1$ : coeficiente angular y, não-nulo por hipótese

De acordo com a Equação 8, tem-se:

$$\Delta Z(t) = \beta_0 + \beta_1 t + a_t - (\beta_0 + \beta_1(t-1) + a_{t-1}) \quad (8)$$

$\Delta Z(t) = \beta_1 + a_t - a_{t-1}$ , que é uma série sem tendência

### 2.3.2.1 Teste do sinal

No teste de Cox-Stuart, as observações são agrupadas em pares  $(Z_1, Z_{1+c}), (Z_2, Z_{2+c}), \dots, (Z_{N-c}, Z_N)$ , em que  $c = N/2$ , se N for par e  $c = (N+1)/2$ , se N for ímpar.

A cada par  $(Z_i, Z_{i+c})$  associa-se o sinal de + (positivo), se  $Z_i < Z_{i+c}$  e o sinal de - (negativo), se  $Z_i > Z_{i+c}$ , eliminando-se os empates.

Assim, são testadas as hipóteses  $H_0$ , não existe tendência e  $H_1$ , existe tendência, como descrito a seguir.

$H_0$ :  $P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c})$ ,  $\forall i$ : não existe tendência;

$H_1$ :  $P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c})$ ,  $\forall i$ : existe tendência.

Segundo Morettin & Toloi (2004), para  $n \leq 20$ , a regra de decisão é baseada na distribuição binomial e, para  $n > 20$ , pode-se usar a distribuição normal.

Além disso, considerando  $T_2$  é igual ao número de pares com sinal positivo, para o teste unilateral, valores grandes de  $T_2$  indicarão que positivo (+) é mais provável que negativo (-).

Assim, rejeita-se  $H_0$ , se  $T_2 \geq n - t$ , em que  $t$  é encontrado numa tabela de distribuição binomial, com  $p = 0,5$  e  $n$ , para um dado nível de significância  $\alpha$ .

### 2.3.2.2 Teste de Fisher

O teste para verificar periodicidade na série foi proposto por Fisher (1929) e é baseado na estatística demonstrada pela Equação 9:

$$g = \frac{\max_{\substack{[N/2] \\ 2}} I_j^N}{\sum_{j=1}^{\substack{[N/2] \\ 2}} I_j^{(N)}} = \frac{I^{(1)}}{\sum_{j=1}^{\substack{[N/2] \\ 2}} I_j^{(N)}} \quad (9)$$

em que:

$I_j$ : é o valor do periodograma;

$N$  é o tamanho da amostra.

Este teste foi proposto para avaliar, para o maior período, as hipóteses  $H_0$  e  $H_1$ , sendo  $H_0$  definido como a ausência de periodicidade e  $H_1$  como presença de periodicidade.

Segundo Morettin & Tolo (2004), mesmo que a hipótese  $H_0$  seja confirmada, é possível que ocorram picos nas ordenadas do periodograma devido às flutuações aleatórias. Assim, mesmo que o periodograma apresente vários picos, não é possível considerar, a priori, que cada um destes corresponde a uma componente periódica em uma série  $Z(t)$ .

Segundo Fisher (1929), a distribuição exata de  $g$  é dada pela Equação 10:

$$P(g > Z) = \alpha = N(1 - Z)^{N-1} \quad (10)$$

em que:

$N = n/2$ ;

$\alpha$  é o nível de significância.

Assim, se  $g > Z$ , rejeita-se  $H_0$ .

Desse modo, a presença de periodicidade quer dizer que alguma situação se repete em períodos relativamente curtos de tempo. Pode-se citar o pico de produção da cultura do café em alguns estados, a cada intervalo de tempo de dois anos.

### **2.3.2.3 Modelos para séries temporais**

Segundo Morettin & Tolo (2004), os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas.

Assim, há um número muito grande de modelos diferentes para descrever o comportamento de uma série particular. Morettin & Tolo (2004)

destacam que a construção desses modelos depende de vários fatores, tais como o comportamento do fenômeno ou o conhecimento *a priori* do objetivo da análise.

Os modelos de ajustes em séries temporais podem ser classificados em duas classes: paramétricos e não-paramétricos:

- a) paramétricos: são aqueles com número de parâmetros finito;
- b) não-paramétricos: são aqueles com número de parâmetros infinito.

A análise dos modelos paramétricos é feita no domínio do tempo e os modelos mais usados são os de erros (de regressão) e os autorregressivos integrados e de médias móveis (ARIMA).

Diferentemente dos paramétricos, a análise dos modelos não-paramétricos é feita no domínio da frequência e os modelos mais usados são a função de autocovariância e a transformada de Fourier.

Uma metodologia bastante empregada na análise de modelos paramétricos é conhecida por abordagem de Box & Jenkins (1970). Segundo Morettin & Toloí (2004), tal metodologia consiste em ajustar modelos autorregressivos integrados de médias móveis, ARIMA (p, d, q), a um conjunto de dados.

Morettin & Toloí (2004) destacam, ainda, um ciclo iterativo para ajustes dos modelos, que é dividido nos seguintes estágios:

- a) uma classe geral de modelos é considerada para a análise (especificação);
- b) identificação do modelo com base em funções de autocorrelação, autocorrelação parcial e outros critérios;
- c) fase de estimação, na qual os parâmetros do modelo identificado são estimados;
- d) verificação ou diagnóstico do modelo ajustado, para verificar se os objetivos propostos foram atingidos;

Morettin & Tolo (2004) afirmam que caso o modelo ajustado não seja o adequado, o ciclo deve ser repetido e voltar à fase de identificação. Além disso, a fase crítica é a identificação e é possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal.

Para ajuste do modelo autorregressivo integrado médias móveis (ARIMA), proposto por Box & Jenkins (1976), é necessário que a série atenda à condição de estacionaridade, ou seja, que ela se desenvolva aleatoriamente no tempo ao redor de uma média constante. Esta condição de “equilíbrio” pode ser obtida pela tomada de diferenças, como é normalmente feito para séries econômicas.

O modelo ARIMA (p, d, q) pode ser expresso pela relação da Equação 11:

$$\phi(B)(1 - B)^d Z(t) = \theta(B) \varepsilon_t \quad (11)$$

em que:

B: é o operador de retardo tal que  $B^j Z_t = Z_{t-j}$

d: é o número de diferenças necessárias para tornar a série estacionária

$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ : é o polinômio autorregressivo de ordem p

$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ : é o polinômio de médias móveis de ordem q.

Estes polinômios constituem casos particulares do modelo ARIMA (p,d,q). Assim, se d = 0 e q = 0, tem-se o modelo autorregressivo de ordem p, AR (p), se d = 0 e p = 0, tem-se o modelo médias móveis de ordem q, MA (q) e se d = 0, tem-se o modelo autorregressivo médias móveis ARMA (p,q).

Segundo Morettin & Tolo (2004), após estimar um modelo para uma série, é necessário verificar se este representa, ou não, adequadamente, o conjunto de dados.

Esta verificação pode ser feita analisando-se os resíduos. Se o modelo for adequado, então, segundo Morettin & Tolo (2004), os erros constituirão ruído branco. Quando constatada a presença de ruído branco, os resíduos serão não-correlacionados e independente e identicamente distribuídos em relação à distribuição normal ( $N(0, \sigma^2)$ ).

Vários são os testes para diagnóstico do modelo ajustado. Dentre eles, destaca-se o teste de Box-Pierce, que considera as autocorrelações dos resíduos, e pode ser mostrado pela Equação 12:

$$Q(K) = n(n+2) \sum_{j=1}^k \frac{\hat{r}_j}{(n-j)} \quad (12)$$

em que:  $\hat{r}_j$  representa a função de autocorrelação e  $n$  o número de observações dividido por 2.

Assim, se o modelo for apropriado, a estatística da Equação 12 terá, segundo Morettin & Tolo (2004), distribuição  $\chi^2$  com  $k-p-q$  graus de liberdade, em que  $p$  e  $q$  referem-se às ordens do modelo ajustado e  $k$ , o número de *lags* utilizados.

A hipótese de ruído branco é rejeitada para valores grandes de  $Q(K)$ , ou seja, se  $Q(K) < \chi^2$ , então, o resíduo será ruído branco.

### 3 MATERIAL E MÉTODOS

Para a confecção do trabalho, este foi dividido em duas etapas.

A primeira etapa consistiu da escolha e da obtenção das imagens, que foram utilizadas como base de dados para processamento e geração dos resultados.

A segunda etapa caracterizou-se pelo emprego das técnicas estatísticas: adoção dos métodos de análise multivariada (análise de agrupamento) e dos procedimentos de julgamento utilizados em séries temporais.

#### 3.1 Material

A base de dados utilizada foram matrizes STS de 512x512 pixels de dados representativos de sêmen de bovinos, cedidas pelo Departamento de Engenharia e pelo Departamento de Medicina Veterinária da UFLA<sup>2</sup>. Foram utilizadas três matrizes STS das sete totais disponíveis.

As imagens, que deram origem aos sete STS, foram capturadas por um sistema de aquisição de imagens (com intervalo de captura de 80 milissegundos) instalado no Laboratório de Laser e Óptica da mesma universidade. O *setup* para captura, utilizado nesses procedimentos, foi composto por uma câmera CCD, um laser He-Ne, de 632 nm de comprimento de onda e um monitor para a visualização e os ajustes do arranjo experimental montado. Além disso, estas imagens foram capturadas de acordo com o padrão objetivo ou “free propagation”, em que um ponto da imagem recebe contribuição de todos os outros pontos.

As matrizes STS formadas a partir das imagens de sêmen bovino, obtidas desta iluminação e utilizadas neste trabalho, exibem o comportamento da atividade do material com o decorrer do tempo. Mais especificamente, uma

---

<sup>2</sup> UFLA – Universidade Federal de Lavras



tendência decrescente que se refere à queda dos índices de atividade celular em relação aos intervalos sucessivos de tempo.

### 3.1.1 Softwares utilizados nas análises

O *Software* R 2.8.0 (The R Foundation for Statistical Computing) foi utilizado como ferramenta para as análises de séries temporais. Também foi utilizado o *software* Minitab® 15.0<sup>3</sup> para a confecção da análise de agrupamento relativo aos procedimentos de estatística multivariada.

### 3.2 Métodos

Nesta fase, foram utilizados as estatísticas de séries temporais e os procedimentos para a realização da análise multivariada do sêmen.

A análise por séries temporais consistiu, basicamente, em ajustar modelos para algumas linhas de cada grupo de cada matriz e, assim, obter informações sobre o comportamento do material, por meio da avaliação da presença de tendência e da análise dos gráficos das funções de autocorrelação.

A análise multivariada foi feita com o tratamento dos dados do padrão de *speckle* pela utilização da técnica hierárquica divisiva, que aplicou a distância euclidiana como medida de dissimilaridade, e o algoritmo das k-médias, para possibilitar a construção dos conglomerados.

Além disso, para completar este tratamento, foi realizada uma análise exploratória nas matrizes STS, para encontrar uma estrutura natural de agrupamento e, assim, obter o entendimento do comportamento da atividade celular, exibido pelas imagens de sêmen bovino, com uma quantidade menor de informações disponíveis.

---

<sup>3</sup> Minitab® 15 Statical Software é um aplicativo utilizado para análises estatísticas. Foi utilizada a versão demo disponível no sítio: <http://www.minitabbrasil.com.br/>

### 3.2.1 Análise de agrupamento

A análise de cluster, formação dos grupos, foi feita seguindo os seguintes passos:

- a. calcularam-se os momentos de inércia, para todas as linhas de todas as matrizes STS. Cada índice de atividade foi considerado uma variável na separação dos clusters;
- b. com base na quantificação dos momentos de inércia de cada matriz STS, três “linhas” foram selecionadas de cada uma delas, para serem utilizadas como semente e servirem de base para a formação de três clusters (um cluster para representar as linhas com baixa atividade, ou baixos índices de atividade, outro para representar as linhas com atividade média e um terceiro para representar as linhas com alta atividade). A escolha do número de grupos foi feita casualmente sendo possível a formação de um número maior ou menor de grupos.

De acordo com Milligan (1980), o algoritmo k-médias não apresenta uma boa performance quando as sementes são escolhidas arbitrariamente. Entretanto, reforça que, se as sementes escolhidas como pontos iniciais de partida forem boas representantes de cada grupo que se deseja formar, o agrupamento por k-médias comporta-se de forma robusta.

Dessa forma, a escolha de cada linha, semente, foi tomada com base na observação de seu módulo (intensidade) da seguinte forma:

- para representar o grupo com baixos índices de atividade, foi tomada uma semente com pequeno valor de MI;
- para representar o grupo de valores médios, foi tomada uma semente com módulo de MI intermediário aos valores totais presentes na matriz STS;
- para representar o grupo de índices com alta atividade, foi tomada uma semente com alto valor de MI.

Posteriormente, foi empregado o algoritmo das k-médias e da distância euclidiana, para término do processo de separação das 512 linhas e geração dos 3 grupos em cada matriz STS.

Assim, foi possível, após a geração dos três grupos nos três STS, realizar uma análise intergrupos. Neste trabalho, foram utilizados o primeiro STS e seus respectivos grupos para esta avaliação interna.

Diferentemente do processo de verificação dos grupos de um mesmo STS, o processo de análise entre grupos de STS diferentes, foi possível, pelo cálculo do momento de inércia médio de cada grupo em cada STS. Assim, para cada STS, foram gerados três valores de MI médio, sendo um representativo para cada grupo.

A avaliação entre os grupos dos três STS foi realizada neste trabalho para verificar se a tendência de queda da atividade celular com o decorrer do tempo também se mantém nos grupos.

### **3.2.2 Séries temporais**

Para ajuste dos modelos de séries temporais, seguiu-se a metodologia descrita seguir.

- a. Para as três matrizes STS, retiraram-se, aleatoriamente, três linhas de cada grupo, formado na análise multivariada realizada anteriormente.
- b. Cada linha foi considerada uma série temporal a ser estudada.
- c. As linhas tomadas eram compostas de um total de 512 informações variantes no tempo, em intervalos iguais a 80ms.

Para cada linha, seguiram-se os seguintes passos:

- a. construção dos gráficos das séries para análise de comportamentos, como variações atípicas, dados discrepantes, ciclos, etc.;

- b. construção dos gráficos da função de autocorrelação (*fac*) para observação analítica da tendência e/ou sazonalidade;
- c. aplicação do teste do sinal (Cox-Stuart) para a verificação analítica da existência da tendência e do teste de Fisher para exame da presença de sazonalidade;
- d. realização de diferenças para eliminar a componente tendência, caso a presença desta fosse confirmada;
- e. construção dos gráficos das funções de autocorrelação (*fac*) e autocorrelação parcial (*facp*) da série diferenciada para a estimação das ordens do modelo;
- f. Ajuste do modelo e verificação da validade das estimativas.
- g. Aplicação do teste de Box-Pierce para verificar se os resíduos são ruído branco.

## 4 RESULTADOS E DISCUSSÃO

Os resultados obtidos neste trabalho são apresentados em duas partes. Na primeira parte são mostrados os efeitos obtidos pela tabulação dos dados segundo os procedimentos estatísticos multivariados e, na segunda parte, aplicada a metodologia de séries temporais para ajuste de modelos e determinação de comportamentos.

### 4.1 Estatística multivariada: análise de agrupamento

Aplicando-se o algoritmo das k-médias foi possível dividir as matrizes STS em três grupos de informações com características semelhantes (clusters).

O conjunto total das 512 linhas foi dividido em três grupos, nomeados de baixa, média e alta atividade. Na Tabela 2 é exibido o resultado do processo de agrupamento para a primeira matriz STS e exemplificada a formação dos grupos de linhas, dentro desta matriz, de acordo com os seus momentos de inércia.

TABELA 2 Resultado do processo de agrupamento para uma matriz STS

<b>Clusters</b>	<b>Número de observações em cada cluster</b>	<b>Soma dos quadrados dentro dos clusters</b>	<b>Distância média ao centroide</b>	<b>Distância máxima ao centroide</b>
Cluster 1	195	1,26107E+08	789,861	1197,244
Cluster 2	120	8,16893E+07	795,850	1583,172
Cluster 3	197	1,62261E+08	887,439	1583,172

Na Tabela 2 foram classificadas as 512 linhas da primeira matriz STS como 195 linhas pertencentes ao grupo de baixa atividade (cluster 1), 120 como

pertencentes ao grupo de atividade média (cluster 2) e 197 como pertencentes ao grupo de índices com alta atividade (cluster 3).

Além disso, na Tabela 2 são mostradas a soma dos quadrados dentro dos grupos formados, a distância média das observações ao cluster centroide e a distância máxima da observação ao cluster centroide.

Em geral, um cluster com uma pequena soma dos quadrados é mais compacto que um com uma grande soma de quadrados. Assim, pelos dados da Tabela 2 é possível notar que os clusters 1 e 3, que representam, respectivamente, os grupos de baixa e alta atividade, apresentam um conjunto de informações que estão mais próximas uma das outras do que em relação ao cluster 2, que representa as informações com índices de atividade celular intermediários. O centroide é o vetor de variáveis médias para as observações no cluster e é utilizado como um ponto médio.

Pelo módulo de cada semente escolhida para representar cada grupo de baixa, média e alta atividade e pela configuração final apresentada pela Tabela 1, percebe-se que, independente do número de grupos que foi formado, o grupo que representou as linhas com alta e média atividade apresentou uma quantidade de informações maior. Este fato comprova, realmente, a alta atividade do material após a sua coleta, que é mostrada pela primeira matriz STS.

Como descrito por Rabal et al. (2008), sobre a forma de captura das imagens, é importante considerar que este fato diz respeito à quantidade de informação que se pretende obter, já que, ao se aumentar o foco, o tamanho do grão pode se tornar próximo à resolução do sistema óptico de gravação e influenciar o julgamento sobre o que está ocorrendo na matriz STS.

Na Figura 7 observa-se um gráfico de dispersão confeccionado a partir dos grupos formados na primeira matriz STS.

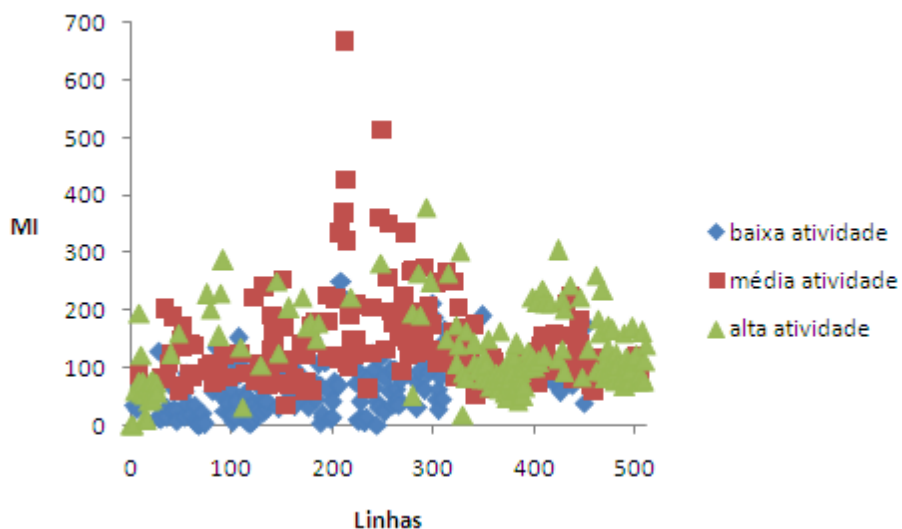


FIGURA 7 Gráfico de dispersão dos grupos de baixa, média e alta atividade

O gráfico da Figura 7 mostra a primeira matriz STS, dividida em três grupos de baixa, média e alta atividade. Cada grupo apresenta uma quantidade variável de linhas em razão da semente selecionada. Assim, se, em um novo processo, uma nova semente for escolhida, o algoritmo por k-médias selecionará um novo conjunto de observações (linhas) para cada grupo, de acordo com esta nova semente.

A análise da Figura 7 que mostra os grupos formados a partir da primeira matriz STS permite confirmar os resultados da Tabela 2. Pela Figura 7 é possível, por meio de uma inspeção visual, observar que, realmente, os grupos de baixa e alta atividade são mais compactos.

Além disso, o gráfico da Figura 7 mostra que, à medida que os índices de atividade vão decrescendo, os módulos das observações dos grupos de alta e média atividade tendem a ficar muito próximos e homogêneos, indicando a aparência de um único grupo e dificultando o processo de separação dessas informações em um destes dois conglomerados.

Esta dificuldade de separação destas informações em relação ao grupo de média ou alta atividade faz com que algumas informações sejam classificadas de forma incoerente e, assim, trocadas entre estes dois grupos.

Na Figura 7, a visualização clara de cada grupo formado fica prejudicada pelo volume total de informações (512 dados) e pela presença de valores de MI que diferem, algumas vezes, somente na terceira casa decimal.

Na Figura 8 observa-se o gráfico dos 7 STS utilizados neste trabalho.

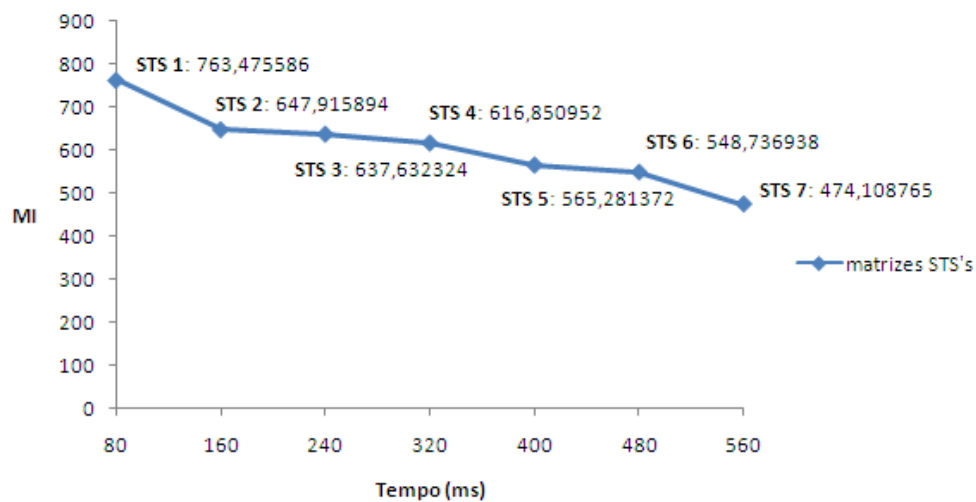


FIGURA 8 Comportamento da atividade do material com o decorrer do tempo

O gráfico da Figura 8 exibe o comportamento dos índices de atividade celular em relação ao tempo, das sete matrizes STS. Pela mesma Figura é possível perceber que o MI destas matrizes apresenta tendência decrescente, evidenciando a diminuição de atividade celular à medida que o tempo passa.

O gráfico da Figura 9 mostra o comportamento do momento de inércia médio de cada grupo de baixa, média e alta atividade nos três STS utilizados no trabalho.



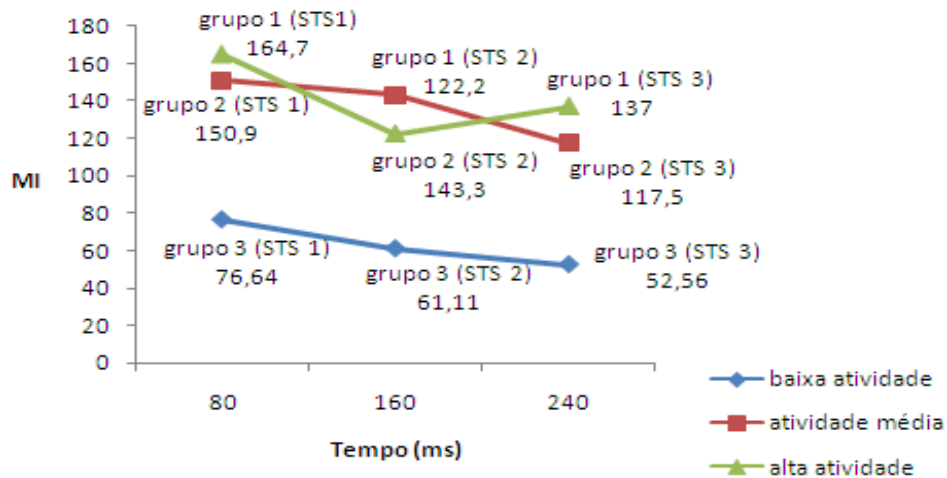


FIGURA 9 Momento de inércia nos grupos de baixa, média e alta atividade.

O gráfico da Figura 9 mostra que as três matrizes STS foram divididas, cada uma, em três grupos, sendo o grupo 1 referente ao grupo de alta atividade, o grupo 2 referente ao grupo de atividade média e o grupo 3, ao grupo de baixa atividade.

Visualmente, observa-se, no gráfico da Figura 9, que, mesmo após a divisão das matrizes STS em grupos de baixa, média e alta atividade, com um número menor de linhas, o comportamento geral do conjunto de células, evidenciando a diminuição gradativa do MI se mantém.

É possível notar também que o MI do grupo de alta atividade, do STS 3, aumentou e que esta elevação pode ser explicada devido ao aumento do número de observações deste grupo, que tendem a ter valores de MI próximos dos grupos de média atividade, como exposto na Figura 7 e, sobretudo, na Tabela 2.

As estatísticas geradas pelo programa *Minitab* ao realizar o processo de clusterização para os três STS estão mostradas na Tabela 3.

TABELA 3 Estatísticas do processo de agrupamento e geração dos três grupos para as três matrizes STS.

	<b>Clusters</b>	<b>Número de observações em cada cluster</b>	<b>Soma dos quadrados dentro dos clusters</b>	<b>Distância média ao centroide</b>	<b>Distância máxima ao centroide</b>
STS1	Baixa atividade	195	1,26107E+08	789,861	1197,244
	Média atividade	120	8,16893E+07	795,850	1583,172
	Alta atividade	197	1,62261E+08	887,439	1583,172
STS2	Baixa atividade	160	1,05620E+08	791,224	1375,608
	Média atividade	191	1,73312E+08	929,798	1656,802
	Alta atividade	161	1,31499E+08	865,597	1948,345
STS3	Baixa atividade	138	8,00505E+07	744,642	1391,360
	Média atividade	167	1,46895E+08	895,223	1958,939
	Alta atividade	207	1,92904E+08	944,617	1595,620

De acordo com os dados da Tabela 3, o agrupamento por k-médias separou as 512 linhas de cada STS (STS 1, STS 2 e STS 3) em três clusters (grupos) que foram chamados de baixa atividade, média atividade e alta atividade, referindo-se aos índices de atividade celular.

A primeira imagem ficou com 195, 120 e 197 observações para os grupos de baixa atividade, média e alta atividade, respectivamente. A segunda com 160, 191 e 161 observações, respectivamente, para os grupos de baixa, média e alta atividade. E a terceira imagem, com 138, 167 e 207 observações, também para os grupos de baixa, média e alta atividade.

É possível notar, pela análise dos dados da Tabela 3, que as estatísticas geradas confirmam o gráfico da Figura 9.

À medida que os grupos de cada STS foram sendo gerados, as linhas sofreram um processo de realocação. Observa-se que, no STS 1, os grupos de baixa e alta atividade perderam informações para o grupo de média atividade no STS 2.

Analisando-se os grupos formados no STS 3 em relação aos grupos formados no STS 1 e STS 2, é possível avaliar que a realocação das observações nos grupos continuou. Pelo aumento do número de observações do grupo de alta atividade do STS 3, como mostrado também na Figura 9, observa-se uma situação ilógica. Esta situação ilógica ocorre devido ao fato de o MI médio aumentar no grupo de alta atividade e não manter a tendência de queda como nos outros STS.

Na Tabela 3 também é possível visualizar a soma dos quadrados dentro dos clusters e a distância média e máxima das observações ao cluster centroide. É possível confirmar a proximidade das informações do grupo 1 e do grupo 3, no início da captura das imagens, como mostrado pela Figura 7 e, com o decorrer do tempo, no STS 3, a aproximação das intensidades do MI das linhas dos grupos de média e alta atividade.

#### **4.2 Séries temporais: ajuste de modelos**

Nesta seção, são analisadas séries retiradas de cada grupo de baixa, média e alta atividade da primeira matriz STS, formados na análise multivariada. Os gráficos exibem o comportamento de cada série e mostram os indícios relevantes que devem ser considerados na análise de séries temporais. Além disso, sugerem, por meio dos gráficos da função de autocorrelação e autocorrelação parcial, as ordens do modelo autorregressivo integrado e de médias móveis, ARIMA (p,d,q).

#### 4.2.1 Análise das séries do grupo de baixa atividade

Nas Figuras 10, 11 e 12 são mostrados os gráficos, de três séries (linhas) da primeira matriz STS, das séries originais e das suas respectivas funções de autocorrelação.

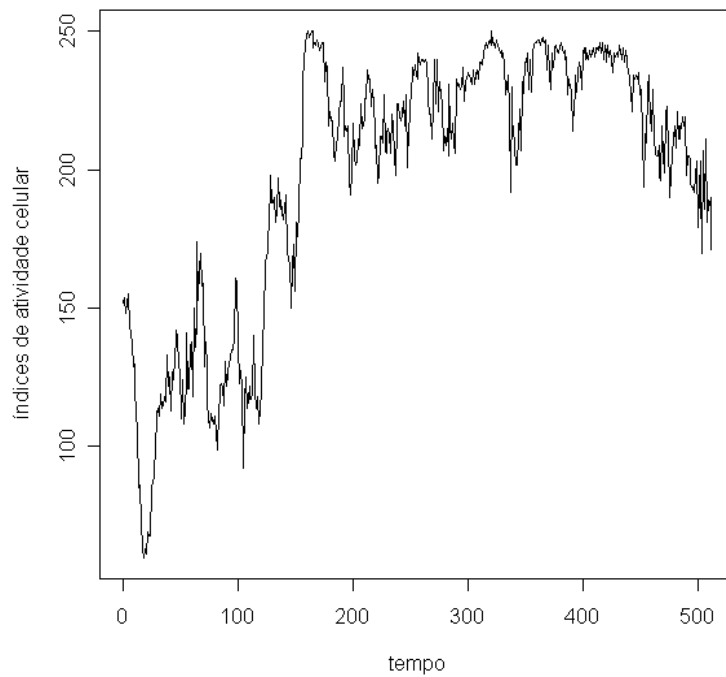
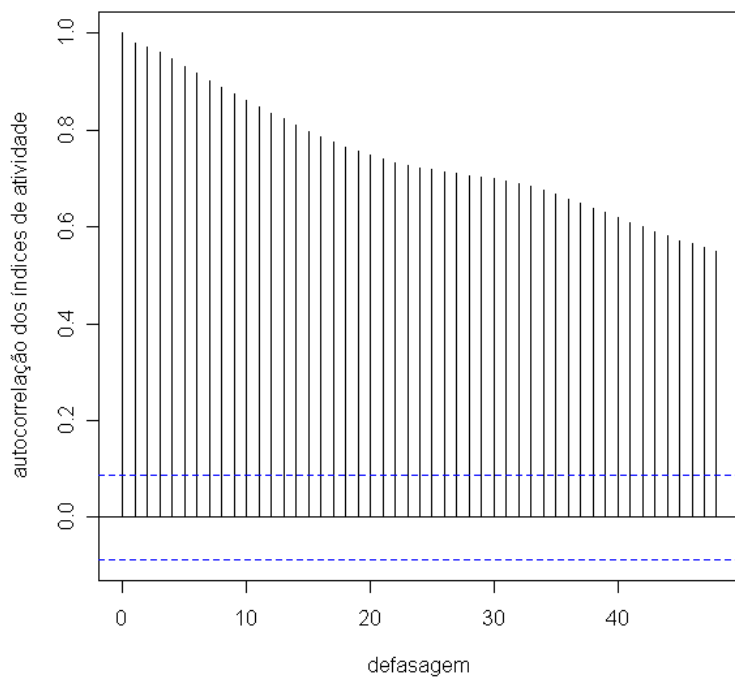


FIGURA 10 Série da linha 56 e sua função de autocorrelação. (Continua)

FIGURA 10, Cont.



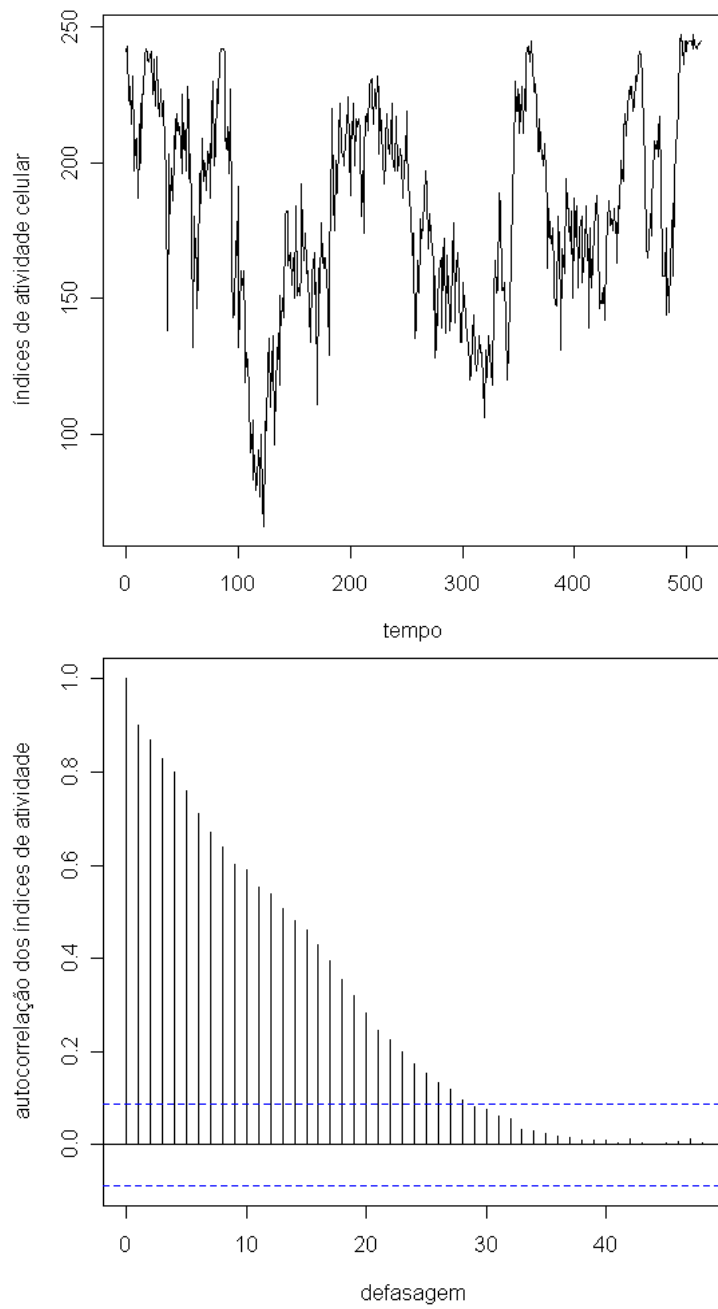


FIGURA 11 Série da linha 289 e da sua respectiva função de autocorrelação.

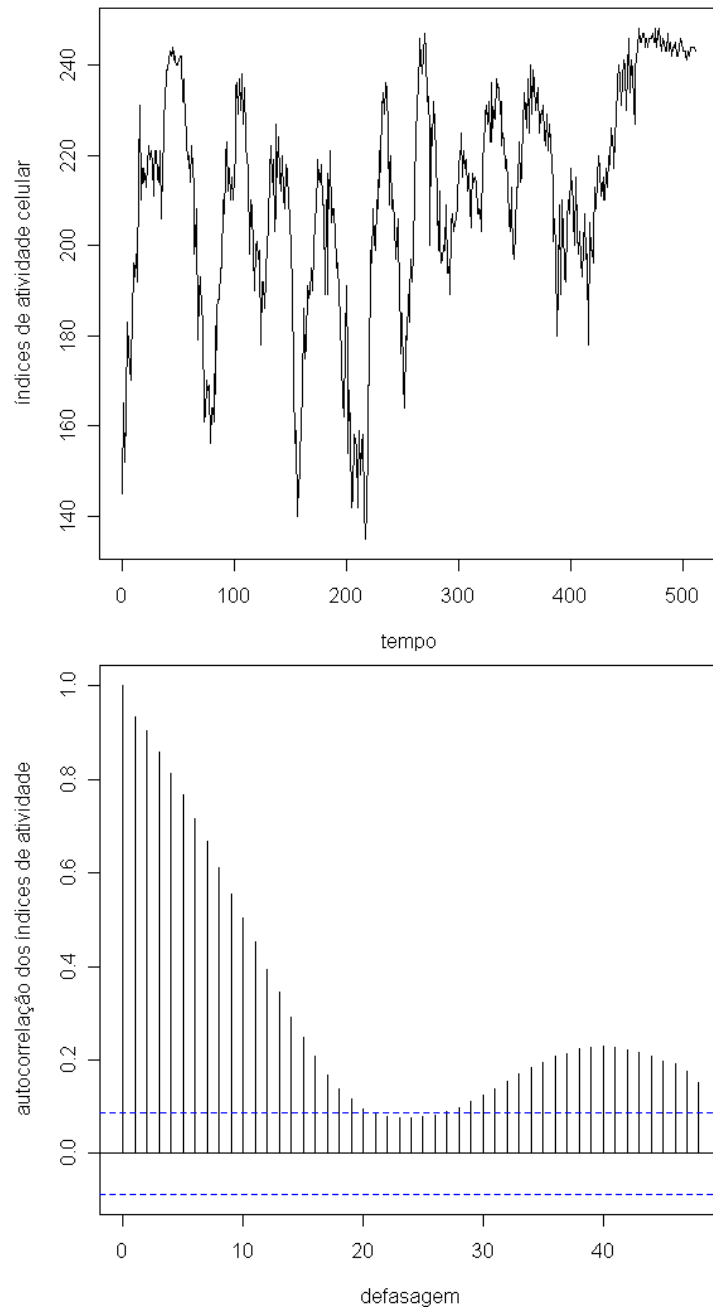


FIGURA 12 Série da linha 59 e da sua respectiva função de autocorrelação.

Os gráficos das Figuras 10, 11 e 12 representam, respectivamente, os índices de atividade celular em relação ao tempo e a função de autocorrelação dos índices de atividade pela defasagem. Nas Figuras 10, 11 e 12, percebe-se que a atividade celular apresenta variações ao longo da história temporal. Estas variações podem ocorrer em virtude de condições biológicas normais por presença de patógenos, perda de água, entre outras. As funções de autocorrelação mostram que as séries não decaem rapidamente para zero, indicando a sua não estacionaridade.

Além disso, é possível notar a presença de indícios da componente tendência para todas as três séries sob análise e que as mesmas aparentam não apresentar sazonalidade.

O teste do sinal (Cox-Stuart) utilizado para verificar a existência da tendência foi aplicado nas séries 56, 289 e 59. Considerando que as séries possuem, cada uma, 512 observações, foi encontrado que  $c = \frac{512}{2} = 256$  e, portanto,  $n = 256$ . Subtraindo-se de  $n$  o número total de empates, as comparações restantes de  $Z_i$  e  $Z_{i+c}$  são todas diferentes.

As estatísticas de Cox-Stuart para  $T$ ,  $t$  e  $n$  encontram-se na Tabela 4.

TABELA 4 Computação de  $T$ ,  $t$  e  $n$  para a verificação da hipótese  $H_0$

Séries	Teste do sinal			$H_0$ : não existe tendência
	T	T	N	
56	199	140,62	255	Rejeita-se $H_0$
289	134	140,09	254	Rejeita-se $H_0$
59	181	140,09	254	Rejeita-se $H_0$

De acordo com os dados da Tabela 4 e considerando o nível de significância de 95%, verifica-se a presença da componente tendência em todas



as séries, confirmando a inspeção visual feita nos gráficos da autocorrelação exibidos nas Figuras 10, 11 e 12.

Entretanto, a realização do teste de Fisher comprovou que, nas séries 56, 289 e 59, não há a presença da componente sazonal.

Nas Figuras 13, 14 e 15 estão representados os gráficos das funções de autocorrelação e autocorrelação parcial para as séries diferenciadas.

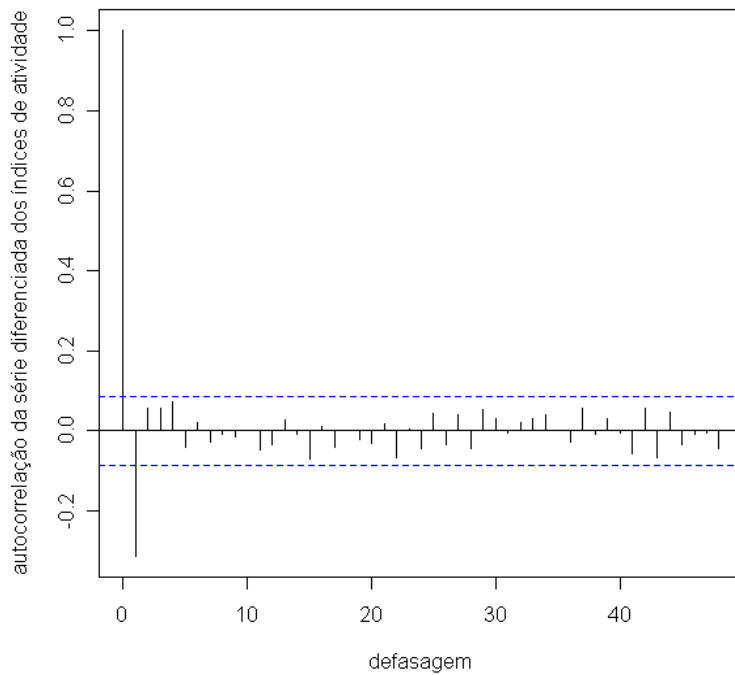
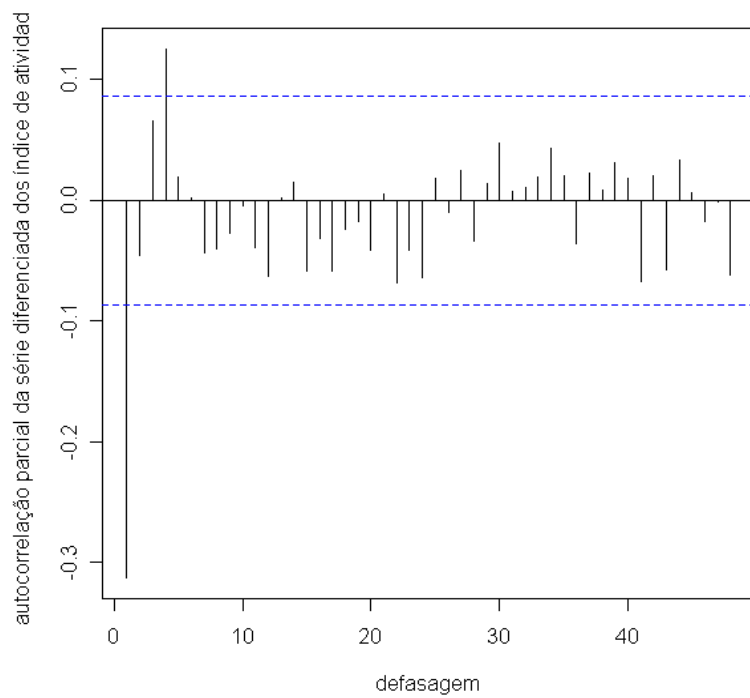


FIGURA 13 Função de autocorrelação e autocorrelação parcial da série 56.  
(Continua)

FIGURA 13, Cont.



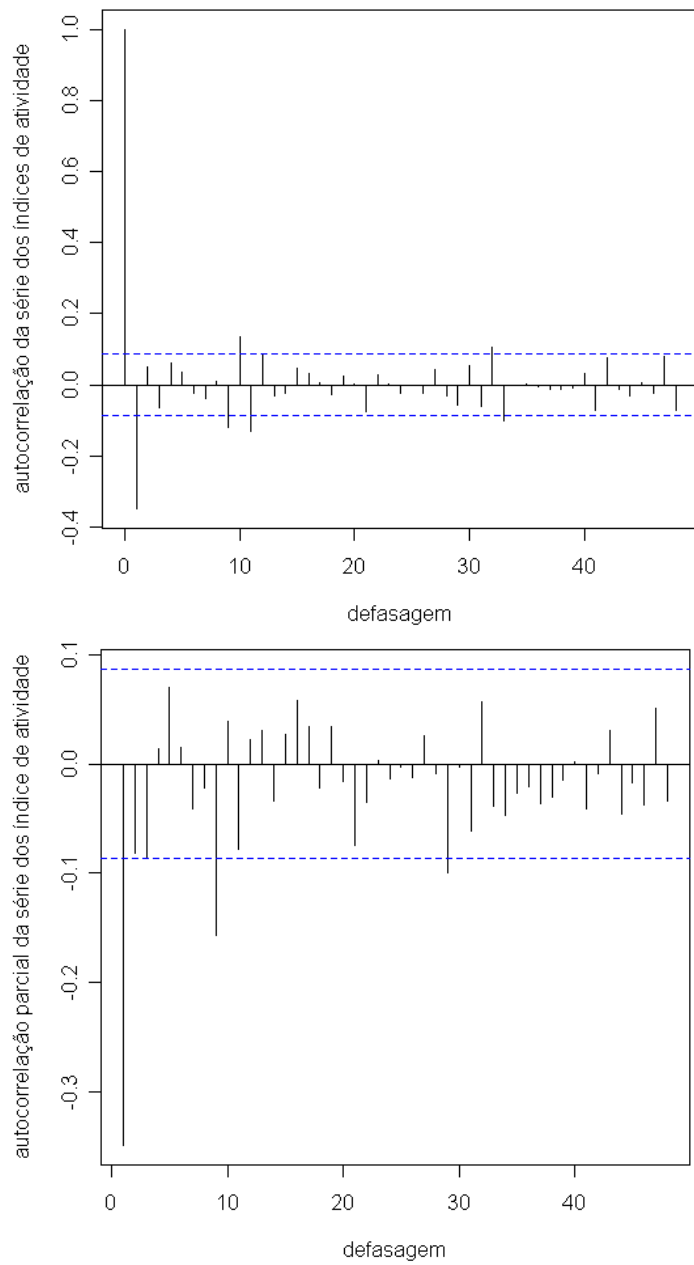


FIGURA 14 Função de autocorrelação e autocorrelação parcial da série 289.

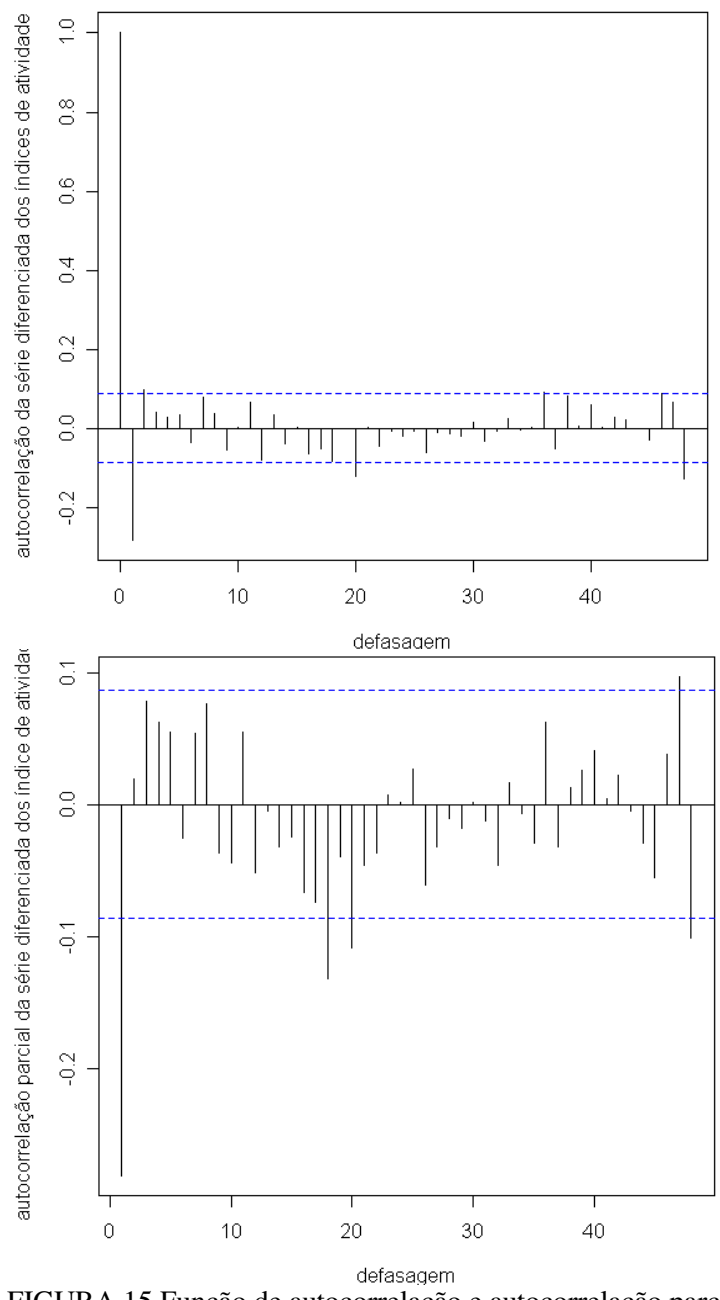


FIGURA 15 Função de autocorrelação e autocorrelação parcial da série 59.

Percebe-se, pelas Figuras 13, 14 e 15, que a linha 56 apresenta um *lag* significativo na *facp*, sugerindo um modelo autorregressivo integrado de ordem 1. Para a linha 289, percebe-se um *lag* significativo na *facp*, sugerindo um modelo autorregressivo integrado de ordem 1. E, para a linha 59, observam-se um *lag* significativo na *facp* e dois *lags* significativos na *fac*, sugerindo um modelo autorregressivo integrado de médias móveis de ordens 1 e 2, respectivamente, de acordo com os dados da Tabela 5.

TABELA 5 Fase de estimação: ordem dos modelos ajustados para as séries do grupo de baixa atividade

	AR(p)	I(d)	MA(q)	Modelo sugerido
Linha56	1	1	0	ARIMA(1,1,0)
Linha289	1	1	0	ARIMA(1,1,0)
Linha59	1	1	2	ARIMA(1,1,2)

Os dados da Tabela 5 demonstram que foram ajustados três modelos da classe ARIMA para as séries 56, 289 e 58 e que as ordens dos modelos diferem de acordo com as funções de autocorrelação e autocorrelação parcial de cada série.

Na Tabela 6 são apresentadas as estimativas para os modelos ajustados para as linhas do grupo de baixa atividade.

TABELA 6 Estimativas dos parâmetros dos modelos ARIMA (1,1,0), ARIMA (1,1,0) e ARIMA (1,1,2), para as séries 56, 289 e 59

Modelos	Parâmetro	Estimativas	Erro padrão
Linha 56	$\phi_1$	-0,3138	0,0421
Linha 289	$\phi_1$	-0,3483	0,0414
Linha 59	$\phi_1$	0,6852	0,2083
	$\theta_1$	-0,9652	0,2056
	$\theta_2$	0,2833	0,0506

Os dados da Tabela 6 demonstram que as estimativas dos parâmetros  $\phi_1$  e  $\theta_1$  para cada série são significativas, uma vez que estão fora do intervalo de duas vezes o módulo do erro padrão.

Na Tabela 7 são exibidas as estatísticas para o teste de Box-Pierce.

TABELA 7 Estatísticas para o teste de Box-Pierce

Modelos	Q (k)	valor-p
Linha 56	44,6128	0,6124
Linha 289	50,084	0,3907
Linha 59	59,012	0,1325

Pela análise da Tabela 7, observa-se que, dos três resíduos analisados, todos constituem ruído branco, já que  $Q(k) < \chi^2$ .

#### 4.2.1 Análise das séries dos grupos de média e de alta atividade

Nas Figuras 16, 17 e 18, como exibido para as séries do grupo de baixa atividade, são apresentados os gráficos das séries dos índices de atividade celular, dos grupos de atividade média e de alta atividade, representados, respectivamente, pelos gráficos das linhas 453, 454 e 455 e pelos gráficos das linhas 319, 320 e 323.

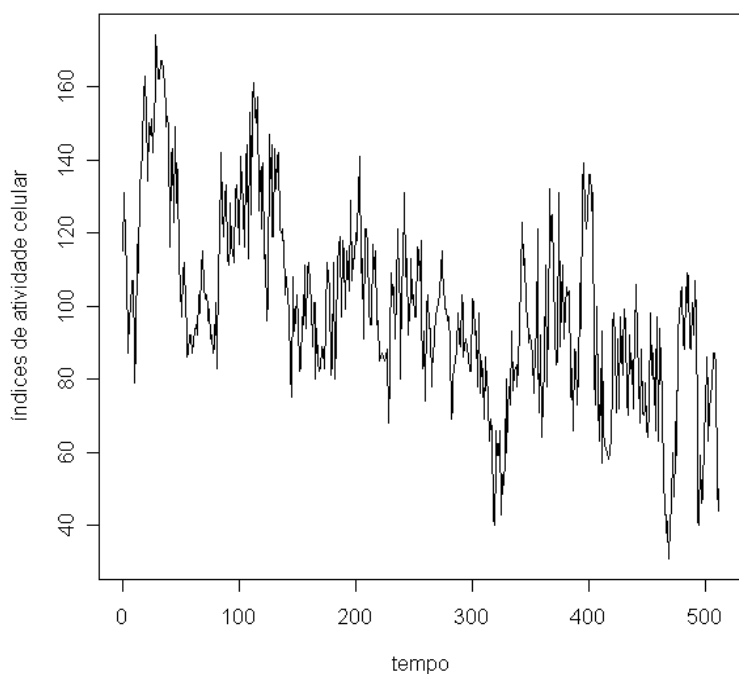
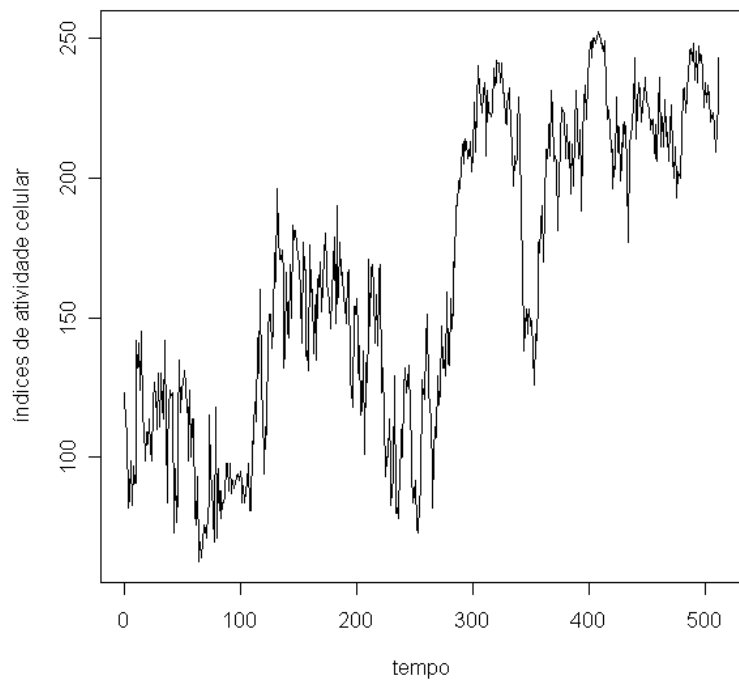


FIGURA 16 Séries 453 e 319 dos grupos de média e alta atividade. (Continua)

FIGURA 16, Cont.





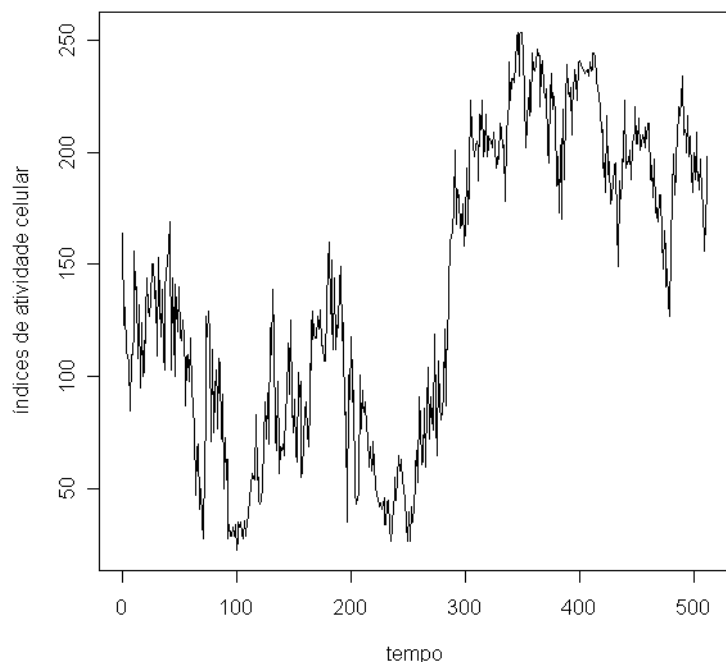
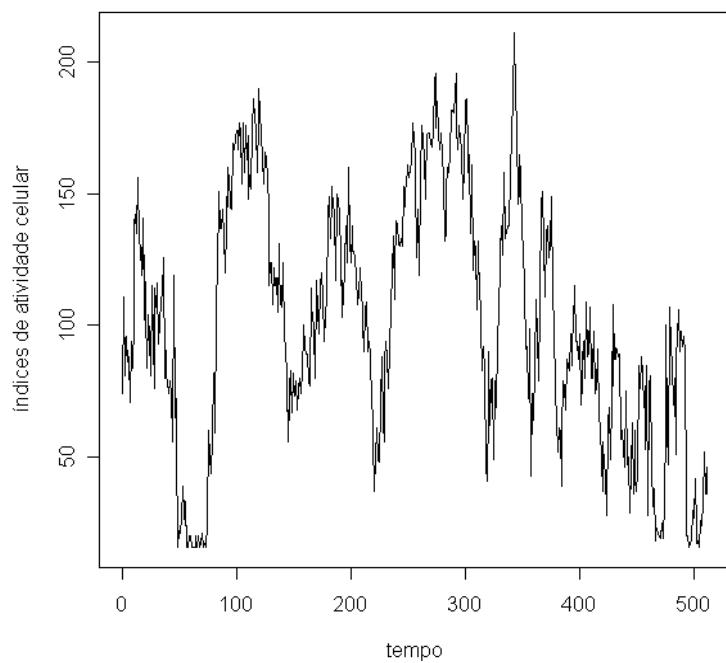


FIGURA 17 Séries 454 e 320 dos grupos de média e alta atividade.

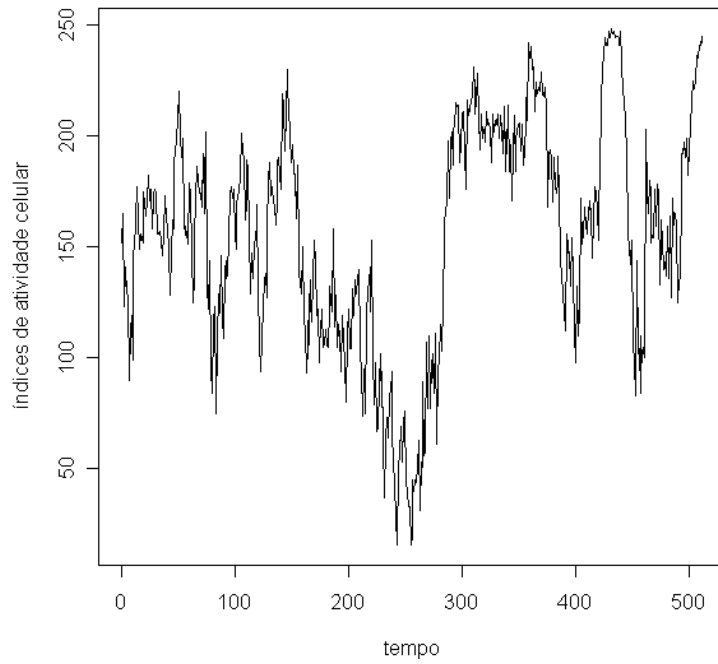
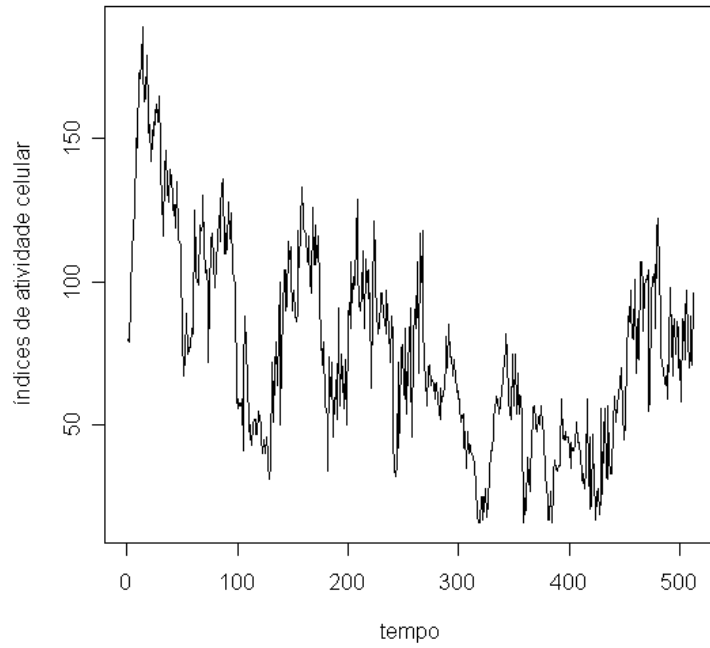


FIGURA 18 Séries 455 e 323 dos grupos de média e alta atividade.

Pelos gráficos das Figuras 16, 17 e 18, é possível intuir a presença da componente tendência como diagnosticado para as séries com baixa atividade. Além disso, mantêm-se as variações da atividade celular ao longo da história temporal, como para as linhas 56, 289 e 59.

O gráfico das funções de autocorrelação (Figuras 19, 20 e 21) mostra um comportamento mais detalhado das séries e esclarece se há ou não a presença desta componente.

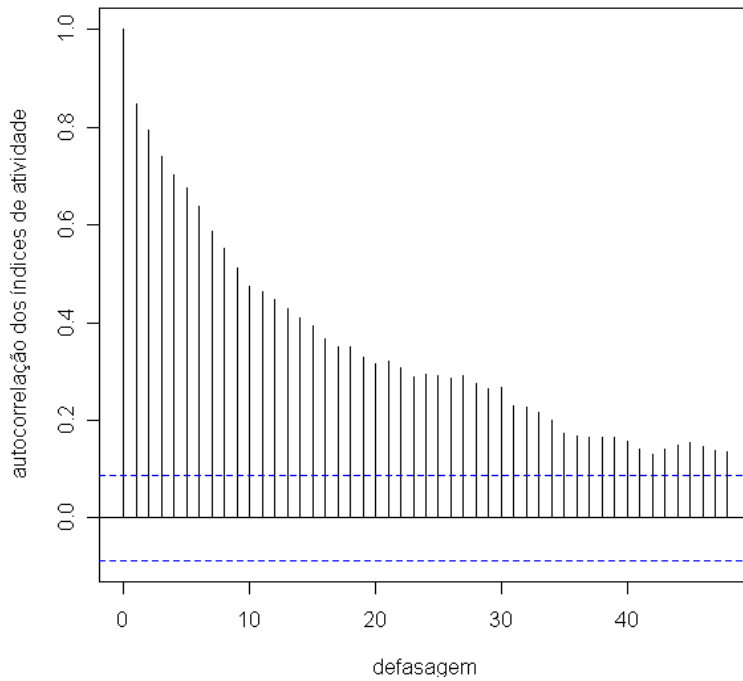
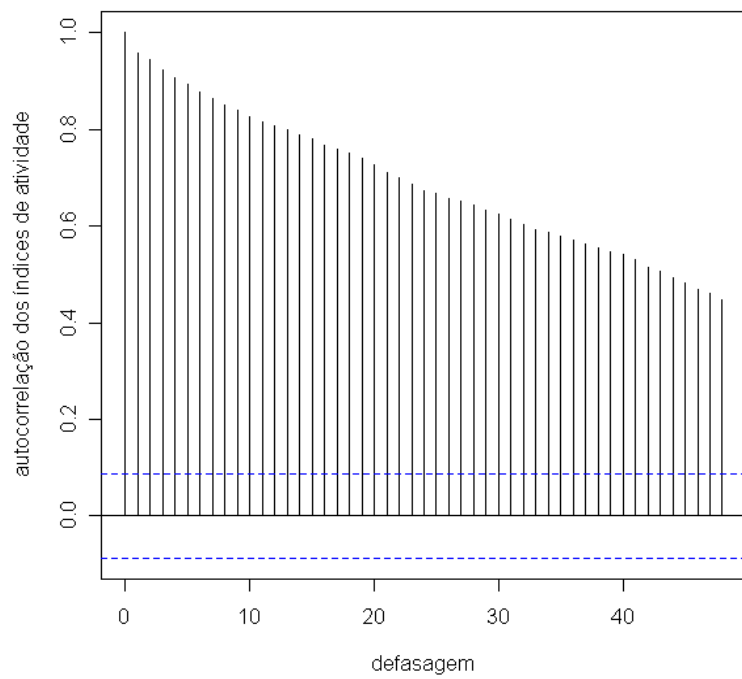


FIGURA 19 Funções de autocorrelação para as séries 453 e 319 dos grupos de média e alta atividade. (Continua)

FIGURA 19, Cont.



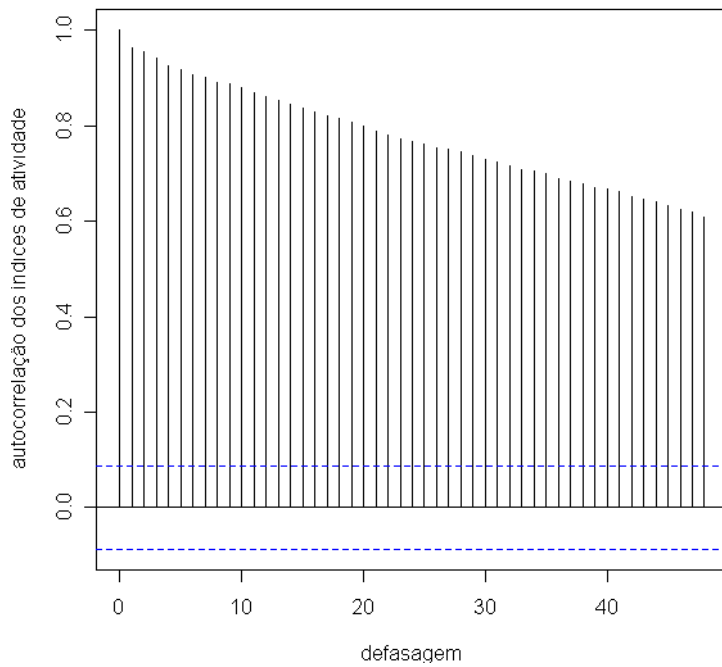
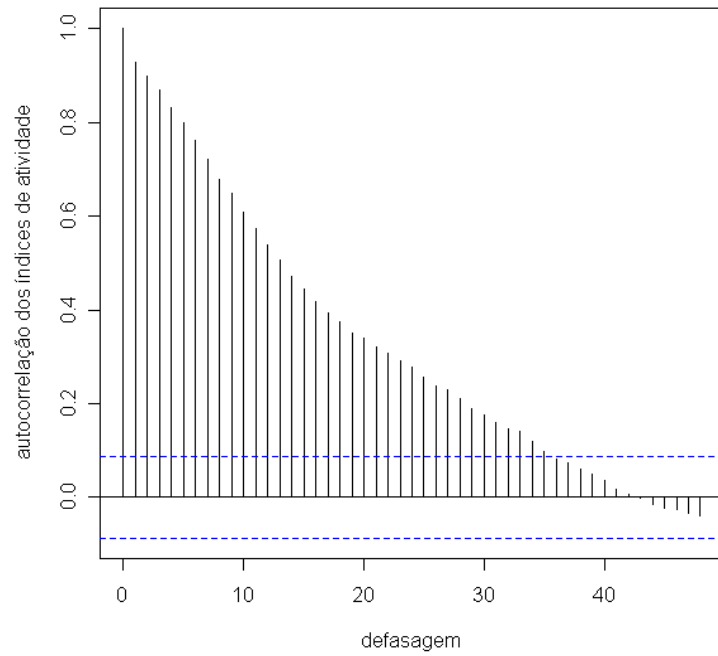


FIGURA 20 Representação gráfica das funções de autocorrelação para as séries 454 e 320, pertencentes, respectivamente, aos grupos de média e alta atividade.

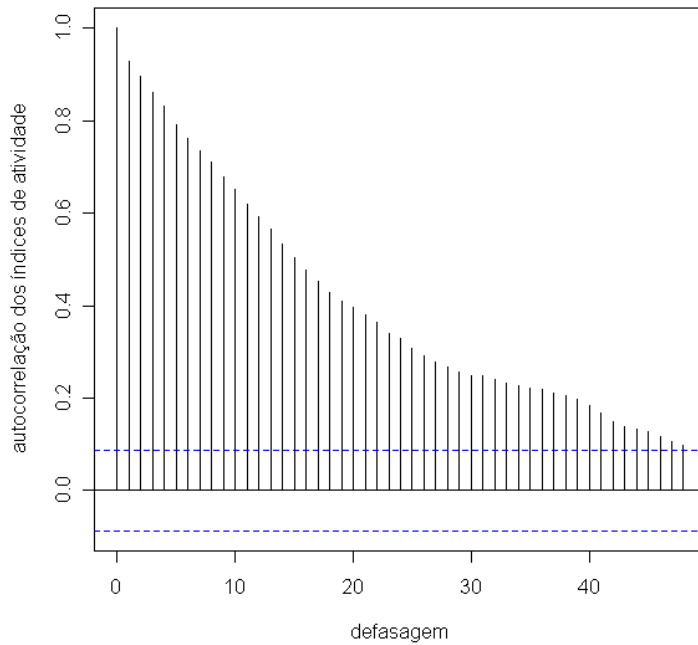
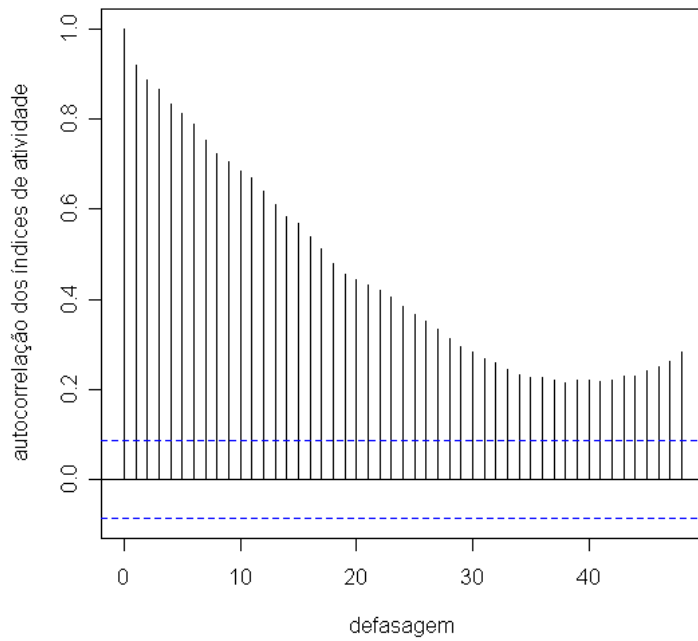


FIGURA 21 Representação gráfica das funções de autocorrelação para as séries 455 e 323, pertencentes, respectivamente, aos grupos de média e alta atividade.

O teste do sinal (Cox-Stuart) foi aplicado para realizar uma verificação analítica da existência ou da componente tendência nas séries. Este teste foi aplicado nas séries 453, 454 e 455, do grupo de atividade média e também nas séries 319, 320 e 323, representantes das linhas de alta atividade. Considerando que todas as séries possuem, cada uma, 512 observações, foi encontrado que  $c = \frac{512}{2} = 256$  e, portanto,  $n = 256$ . Subtraindo-se de  $n$  o número total de empates, as comparações restantes de  $Z_i$  e  $Z_{i+c}$  são todas diferentes.

As estatísticas de Cox-Stuart para  $T$ ,  $t$  e  $n$ , encontram-se na Tabela 8.

TABELA 8 Computação de  $T$ ,  $t$  e  $n$  para a verificação da hipótese  $H_0$

Séries	Teste do sinal			$H_0$ : não existe tendência
	$T$	$t$	$N$	
453	216	140,09	244	Rejeita-se $H_0$
454	118	141,16	256	Rejeita-se $H_0$
455	215	140,09	254	Rejeita-se $H_0$
319	251	140,16	256	Rejeita-se $H_0$
320	226	140,62	255	Rejeita-se $H_0$
323	185	140,62	255	Rejeita-se $H_0$

A realização do teste de Fisher comprovou que, em todas as séries, não há a presença da componente sazonal.

Desse modo, para a estimação das ordens do modelo, foi aplicada uma diferença, nas séries 453, 454 e 455 e nas séries 319, 320 e 323, para a retirada da tendência e construção das funções de autocorrelação e autocorrelação parcial da série diferenciada.

O gráfico da Figura 22 mostra as funções de autocorrelação ( $fac$ ) e autocorrelação parcial ( $facp$ ) das séries diferenciadas do grupo de índices de atividade celular média.

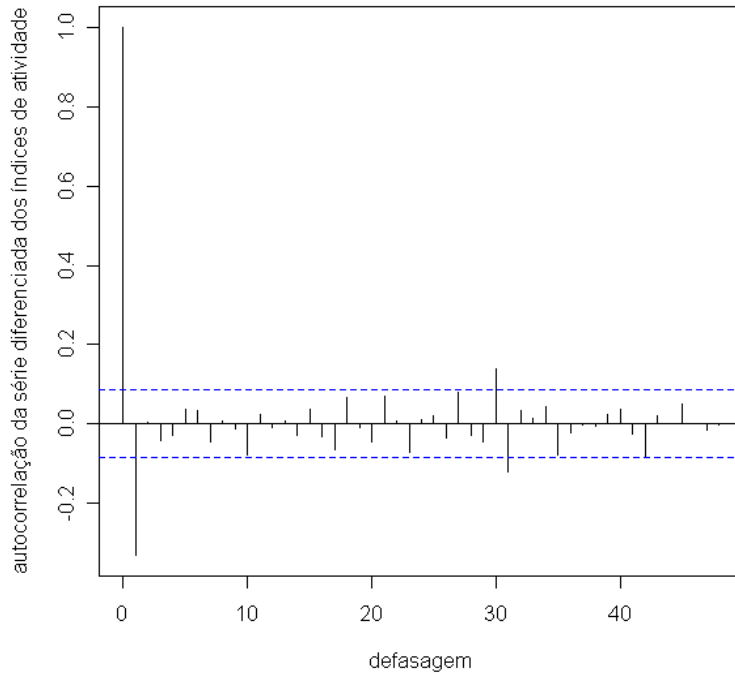
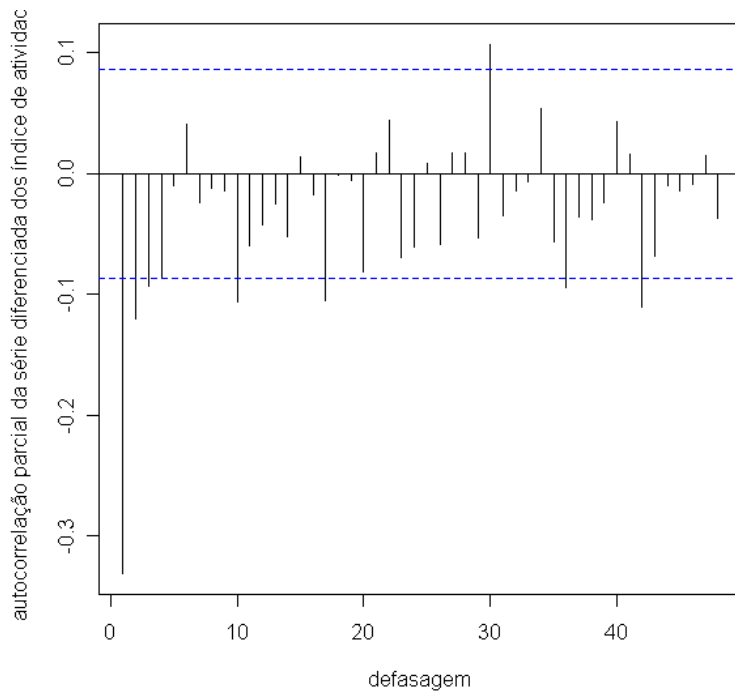


FIGURA 22 Representação gráfica das funções de autocorrelação e autocorrelação parcial para a série 453 do grupo de média atividade. (Continua)



FIGURA 22, Cont.



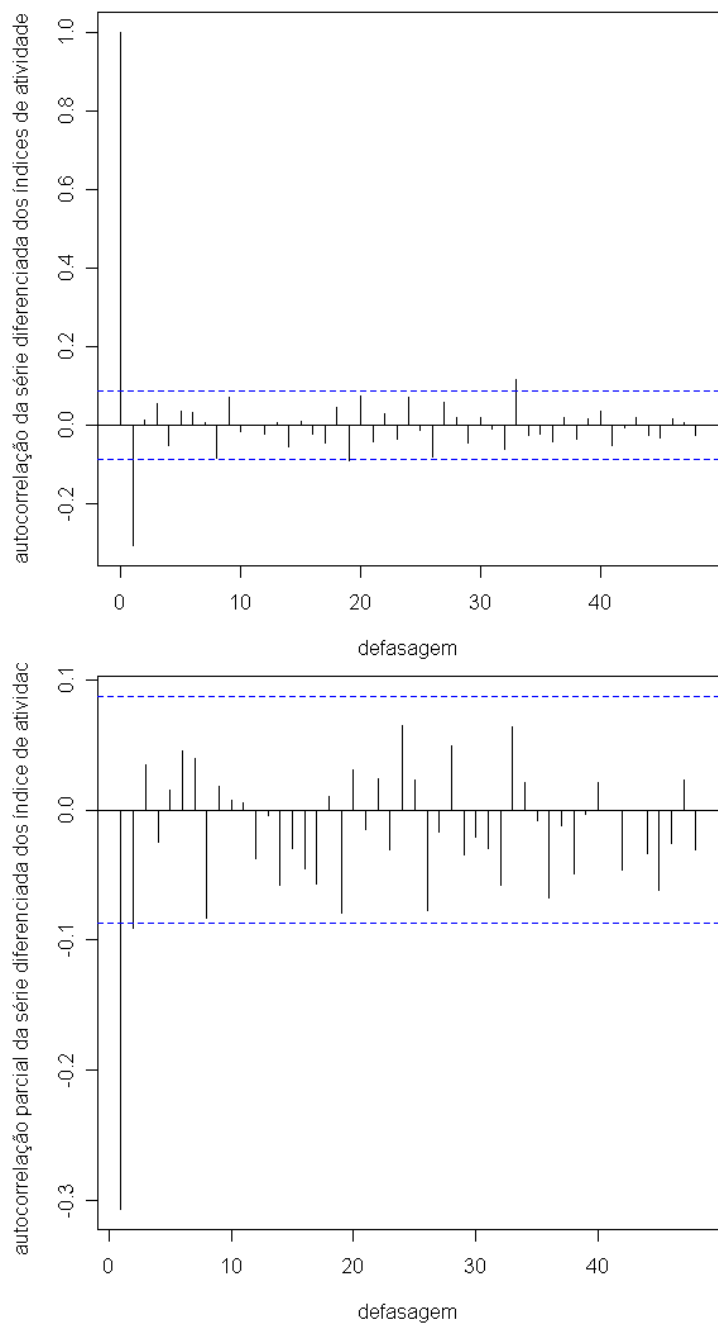


FIGURA 23 Representação gráfica das funções de autocorrelação e autocorrelação parcial para a série 454 do grupo de média atividade.

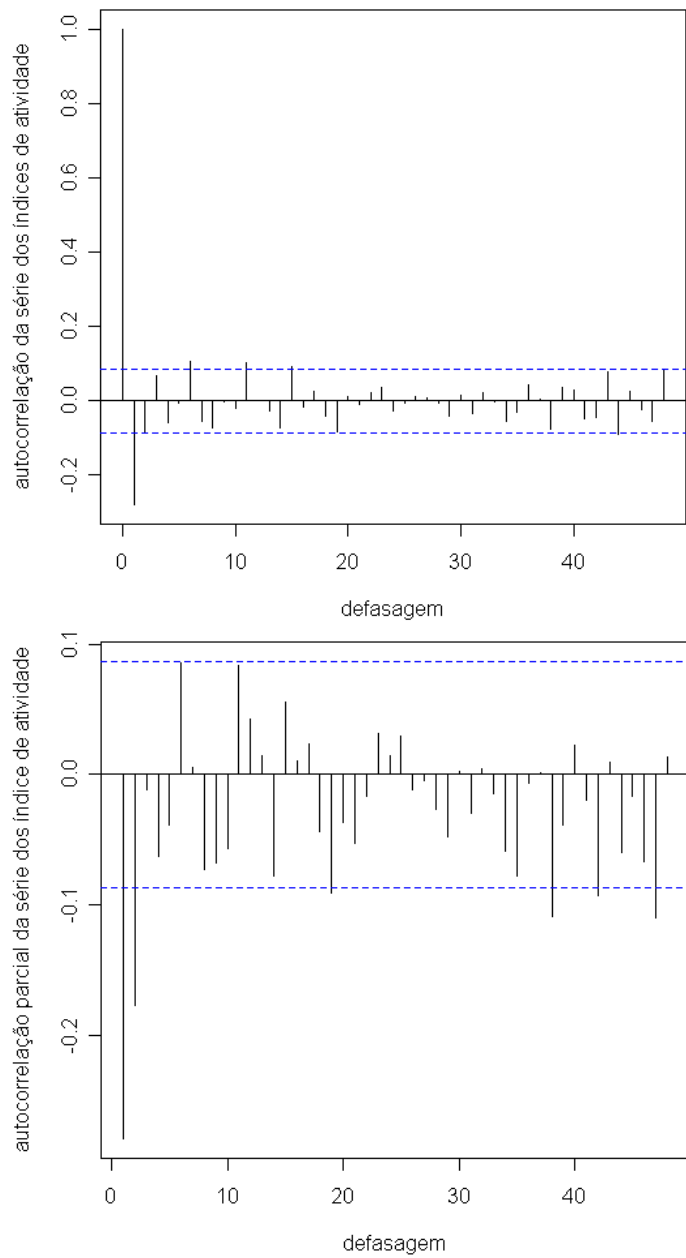


FIGURA 24 Representação gráfica das funções de autocorrelação e autocorrelação parcial para a série 455 do grupo de média atividade.

É possível perceber, pelo gráfico das Figuras 22, 23 e 24, que a linha 453 apresenta três *lags* significativos na *facp* e de um *lag* significativo na *fac*, sugerindo um modelo autorregressivo integrado de ordem 3 de um modelo integrado de médias móveis de ordem 1 ou de um modelo autorregressivo integrado de médias móveis de ordens 3 e 1, respectivamente.

Para a linha 454, observa-se um *lag* significativo na *fac* e outro na *facp*, sugerindo um modelo autorregressivo integrado de ordem 1, um modelo integrado de médias móveis de ordem 1 ou um modelo autorregressivo integrado de médias móveis de ordem 1. E, para a linha 455, observam-se dois *lags* significativos *facp* e um *lag* significativo na *fac*, sugerindo um modelo autorregressivo de ordem 2 ou um modelo de médias móveis de ordem 1, ou, ainda, um modelo autorregressivo de médias móveis de ordens 2 e 1, respectivamente.

As ordens estimadas para as linhas 453, 454 e 455 são mostradas na Tabela 9

TABELA 9 Fase de estimação: ordem dos modelos ajustados para as séries do grupo de atividade média

	AR(p)	I(d)	MA(q)	Modelo Sugerido
Linha 453	3	1	0	ARIMA(3,1,0)
Linha 454	1	1	0	ARIMA(1,1,0)
Linha 455	2	1	0	ARIMA (2,1,0)

Pelos dados da Tabela 9 observa-se que foram ajustados três modelos da classe ARIMA para as séries 453, 454 e 455 e que as ordens dos modelos diferem de acordo com as funções de autocorrelação e autocorrelação parcial de cada série.

As estimativas dos parâmetros para as linhas 453, 454 e 455 são exibidas na Tabela 10.

TABELA 10 Estimativas dos parâmetros dos modelos ARIMA (3,1,0), ARIMA (1,1,0) e ARIMA (2,1,0), para as séries 453, 454 e 455

Modelos	Parâmetro	Estimativas	Erro padrão
Linha 453	$\phi_1$	-0,3815	0,0440
	$\phi_2$	-0,1543	0,0467
	$\phi_3$	-0,0930	0,0443
Linha 454	$\phi_1$	-0,3085	0,0422
Linha 455	$\phi_1$	-0,3305	0,0436
	$\phi_2$	-0,1781	0,0437

De acordo com os dados da Tabela 10, é possível perceber que as estimativas dos parâmetros  $\phi_1$  e  $\theta_1$ , para cada série, são significativas, uma vez que está fora do intervalo de duas vezes o módulo do erro padrão.

Na Tabela 11 são exibidas as estatísticas para o teste de Box-Pierce.

TABELA 11 Estatísticas para o teste de Box-Pierce

Modelos	Q (k)	valor-p
Linhas 453	54,6742	0,236
Linhas 454	44,4657	0,6185
Linhas 455	58,1098	0,1505

Pela análise da Tabela 11, observa-se que os três resíduos constituem ruído branco, já que  $Q(k) < \chi^2$ , em todos os casos.

O gráfico das Figuras 25, 26 e 27 demonstram as funções de autocorrelação (*fac*) e autocorrelação parcial (*facp*) das séries diferenciadas 319, 320 e 323 do grupo de índices de alta atividade.

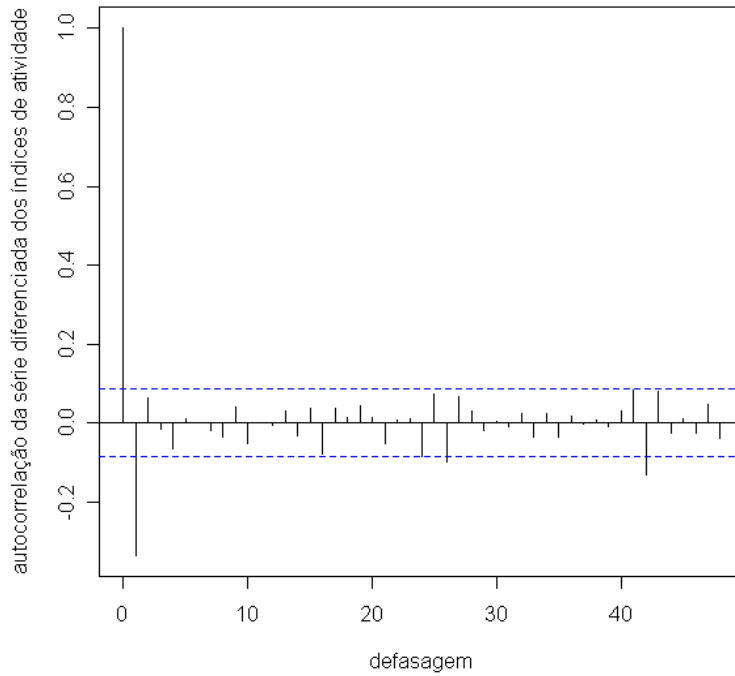
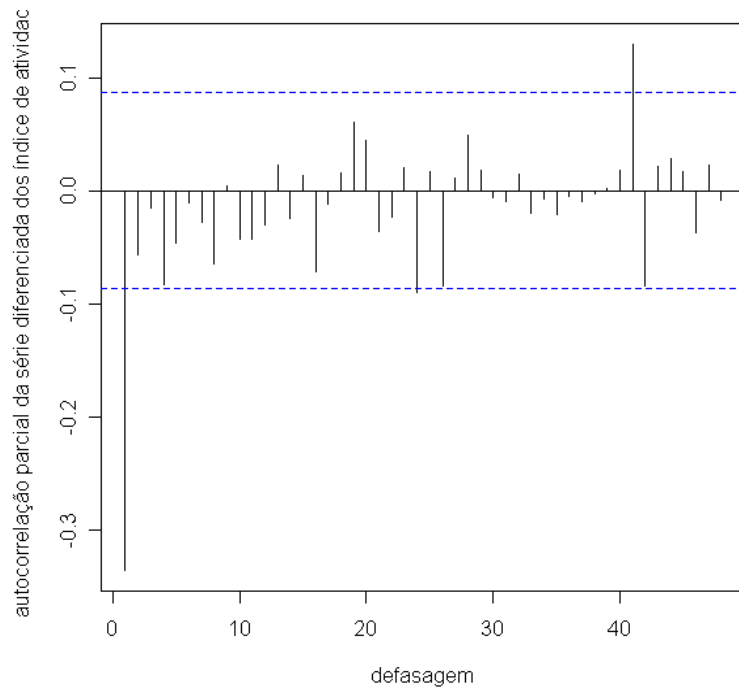


FIGURA 25 Representação gráfica das funções de autocorrelação e autocorrelação parcial para a série 319 do grupo de alta atividade. (Continua)

FIGURA 25, Cont.



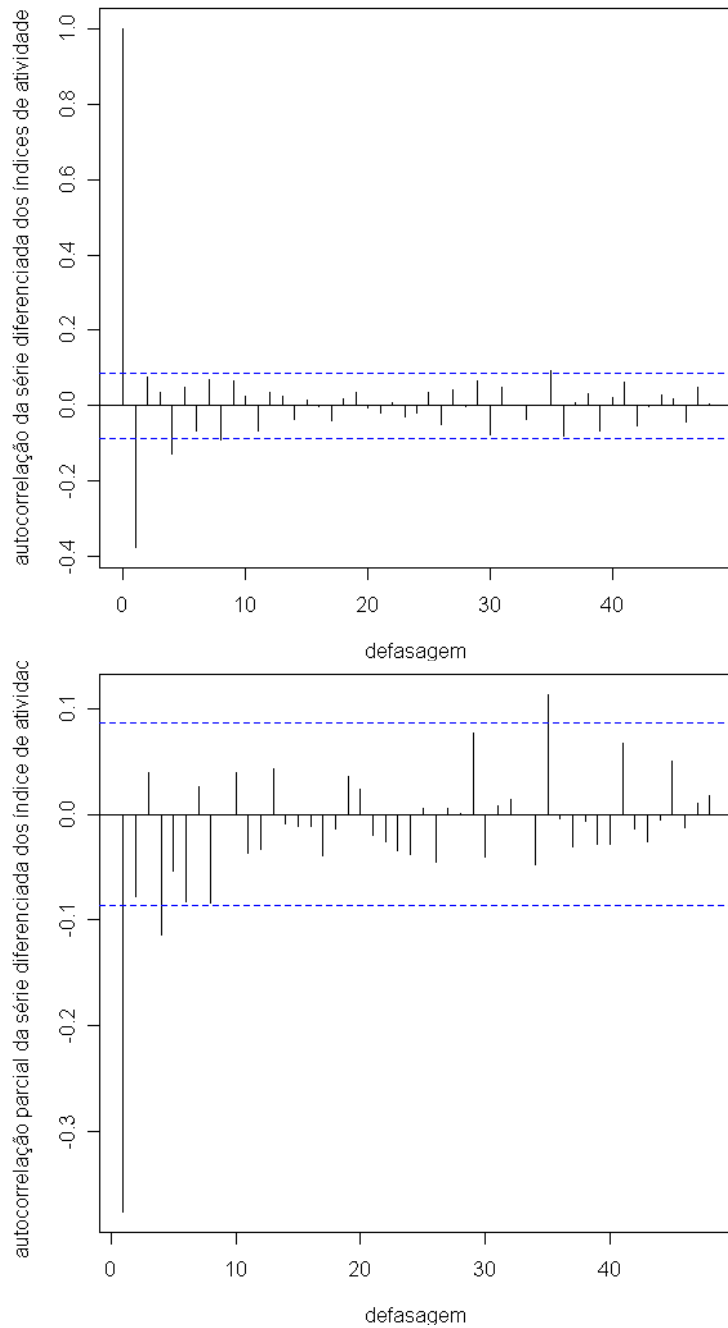


FIGURA 26 Representação gráfica das funções de autocorrelação e autocorrelação parcial para a série 320 do grupo de alta atividade.



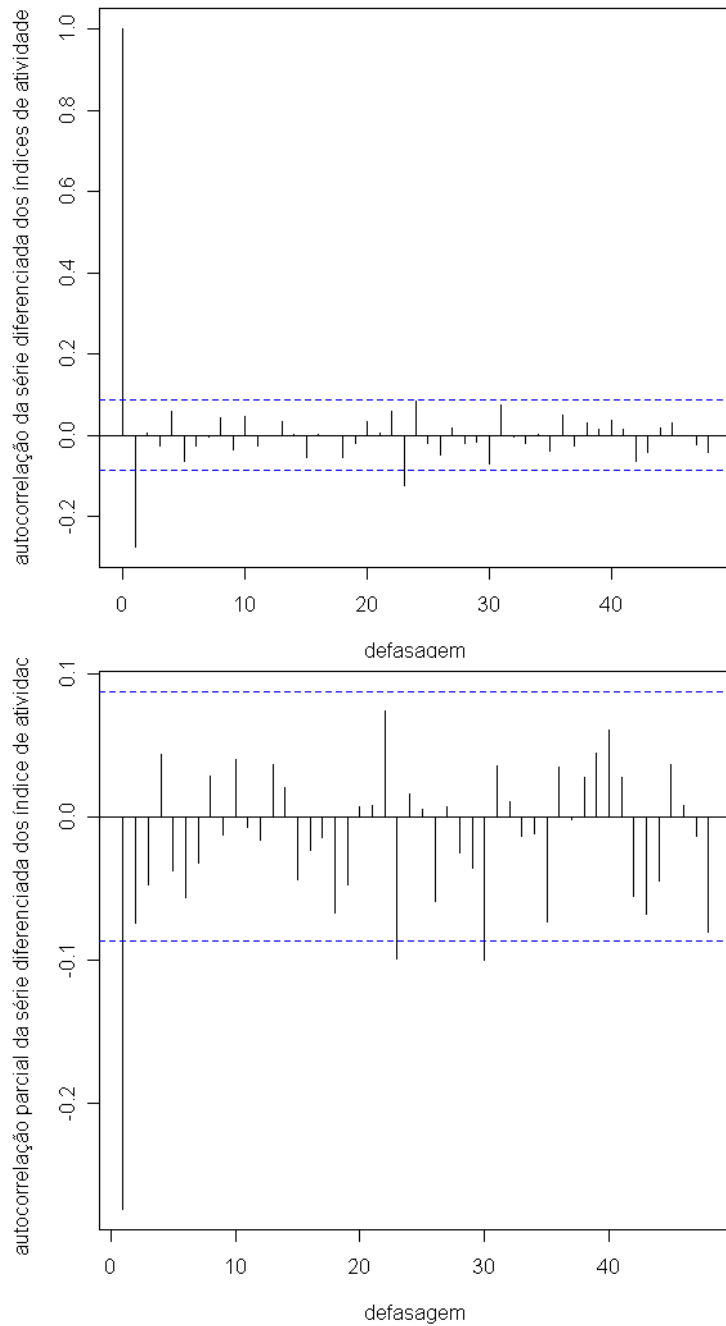


FIGURA 27 Representação gráfica das funções de autocorrelação e autocorrelação parcial para a série 323 do grupo de alta atividade.

Pela análise dos gráficos das Figuras 25, 26 e 27, é possível assumir que a série 319 apresenta um *lag* significativo na *facp* e um *lag* significativo na *fac*, sugerindo um modelo autorregressivo integrado de ordem 1, um modelo integrado de médias móveis de ordem 1 ou um modelo autorregressivo integrado de médias móveis de ordem 1, respectivamente.

Para a linha 320, percebem-se um *lag* significativo na *facp* e outro na *fac*, sugerindo um modelo autorregressivo integrado de ordem 1, um modelo integrado de médias móveis de ordem 1 ou um modelo autorregressivo integrado de médias móveis de ordem 1, respectivamente. E, para a linha 323, observa-se um *lag* significativo na *facp* e outro também na *fac*, sugerindo um modelo autorregressivo integrado de ordem 1, um modelo integrado de médias móveis de ordem 1 ou um modelo autorregressivo integrado de médias móveis de ordem 1, respectivamente, como sugerido para a linha 320. Na Tabela 12 são mostradas as ordens dos modelos ajustados.

TABELA 12 Fase de estimação: ordem dos modelos ajustados para as séries do grupo de atividade média.

	AR(p)	I(d)	MA(q)	Modelo sugerido
Linha 319	0	1	1	ARIMA (0,1,1)
Linha 320	1	1	0	ARIMA (1,1,0)
Linha 323	0	1	1	ARIMA (0,1,1)

Os dados da Tabela 12 demonstram que foram ajustados três modelos da classe ARIMA, para as séries 319, 320 e 321 e que as ordens dos modelos diferem de acordo com as funções de autocorrelação e autocorrelação parcial de cada série.

Na Tabela 13 são apresentadas as estimativas dos parâmetros dos modelos ARIMA (0,1,1), ARIMA (1,1,0) e ARIMA (0,1,1), respectivamente para as linhas 319,320 e 323.

TABELA 13 Estimativas dos parâmetros dos modelos ARIMA (0,1,1), ARIMA (1,1,0) e ARIMA (0,1,1) para as séries 319, 320 e 323.

Modelos	Parâmetro	Estimativas	Erro padrão
Linha 319	$\theta_1$	-0,3497	0,0418
Linha 320	$\phi_1$	-0,3797	0,0412
Linha 323	$\theta_1$	-0,2962	0,0424

Os dados da Tabela 13 demonstram que as estimativas dos parâmetros  $\phi_1$  e  $\theta_1$  para cada série são significativas, uma vez que estão fora do intervalo de duas vezes o módulo do erro padrão.

Na Tabela 14 são exibidas as estatísticas para o teste de Box-Pierce.

TABELA 14 Estatísticas para o teste de Box-Pierce

Modelos	Q (k)	valor-p
Linha 319	39,2401	0,8121
Linha 320	47,2427	0,5038
Linha 323	38,5136	0,8343

Pela análise dos dados da Tabela 14, observa-se que os três resíduos constituem ruído branco, já que  $Q(k) < \chi^2$ , em todos os casos.

## 5 CONCLUSÕES

As técnicas de séries temporais e de análise multivariada foram aplicadas com sucesso ao conjunto de células vivas. A aplicação destas técnicas possibilitou reduzir a dimensão das estruturas, matrizes STS, facilitando a interpretação das análises pela construção de clusters compostos de um número menor de informações.

Além disso, foi possível, com uma análise exploratória de cada massa de dados, observar o comportamento dos três grupos formados dentro da primeira matriz STS e mostrar, por meio do gráfico de dispersão, a configuração das linhas de acordo com os seus índices de atividade.

Na tomada de três linhas de cada grupo formado foi possível avaliar fatores como a presença de tendência, variações sazonais e presença de ciclos.

Fazendo-se a comparação dos grupos gerados em cada STS, foi possível perceber que os mesmos apresentam o mesmo comportamento tendencioso de queda do momento de inércia com o decorrer do tempo. Além disso, que fatores, como o grau de abertura da câmera, que caracteriza o sistema óptico utilizado, apresentam influência relevante na geração dos resultados.

Em suma, o trabalho foi válido por permitir exemplificar e detalhar, por meio das técnicas de séries temporais, o comportamento de um conjunto de matrizes STS ao longo do tempo e por possibilitar a realização da análise de cluster como ferramenta que permite reduzir o custo de interpretação dos resultados gerados nas análises.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, T. W. **An introduction to multivariate statistical analysis**. 2. ed. New York: J. Wiley, 1971. 674 p.
- ARIZAGA, R.; GRUMEL, E. E.; CAP, N.; TRIVI, M. Following the drying of spray paints using space and time contrast of Dynamic *Speckle*. **Journal of Coatings Technology**, Philadelphia, v. 3, n. 4, p. 295-296, Oct. 2006.
- ARIZAGA, R.; TRIVI, M.; RABAL, H. *Speckle* time evolution characterization by the co-occurrence matrix analysis. **Optics & Laser Technology**, London, v. 31, n. 2, p. 163-169, 1999.
- BERGKVIST, A. **Biospeckle-based study of the line profile of light scattered in strawberries**. 1997. 62 f. Dissertation (Master in Science) - Faculty of Technology at Lund University, La Plata.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis: forecasting and control**. San Francisco: Holden-Day, 1970. 575 p.
- BRAGA JR., R. A. **Biospeckle**: desenvolvimento de uma tecnologia aplicada à análise de sementes. 2000. 117 p. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual de Campinas, Campinas.
- BRAGA, R. A.; SILVA, W. S.; SAFADI, T.; NOBRE, C. M. B. Time history *speckle* pattern under statistical view. **Optics Communications**, Amsterdam, n. 281, p. 2443-2448, 2008.
- BRIERS, J. D. “*speckle*” fluctuations and biomedical optics: implications and applications. **Optical Engineering**, Redondo Beach, v. 32, n. 2, p. 277-283, 1993.
- BRIERS, J. D. Laser *speckle* contrast imaging for measuring blood flow. **Applied Optics**, New York, v. 37, n. 1, p. 1-2, 2007.
- DAINTY, J. C. Laser *speckle* and related phenomena. **Applied Optics**, New York, v. 23, n. 16, p. 2661, 1984.
- FERREIRA, D. F. **Estatística multivariada**. Lavras: UFLA, 2008. 661 p.

GALE, M. F. R.; LANDAU, M. R.; HOGERT, E. N.; GAGGIOLI, N. G.  
Changing surfaces: a theoretical and experimental approach. **Journal of Optics A: Pure and Applied Optics**, Bristol, v. 6, n. 2, p. 187-192, 2004.

HAIR, J. F. **Análise multivariada de dados**. 5. ed. Porto Alegre: Bookman, 2005. 593 p.

HENÃO, R.; RABAL, H. J.; TAGLIAFERRI, A.; TORROBA, R.  
Determination of the zero-order fringe position in digital *speckle* pattern interferometry. **Applied Optics**, New York, v. 36, n. 10, p. 2066-2069, 1997.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3. ed. Englewood Cliffs: Prentice-Hall, 1992. 642 p.

MANLY, B. F. J. **Multivariate statistical methods: a primer**. 2. ed. London: Chapman & Hall, 1995. 215 p.

McGARIGAL, K.; CUSHMAN, S.; STAFFORD, S. **Multivariate statistics for wildlife and ecology research**. New York: Springer, 2002. 283 p.

MILLIGAN, G. W. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. **Psychometrika**, Williamsburg, v. 45, n. 3, p. 325-342, 1980.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: UFMG, 2005. 295 p.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: E. Blucher, 2004. 439 p.

NASCIMENTO, A. L. do; COSTA JÚNIOR, A. T.; RABELO, G. F.; BRAGA JÚNIOR, R. A. Desenvolvimento de um modelo para o Biospeckle na análise de sementes de feijão (*Phaseolus vulgaris* L.). **Ciência e Agrotecnologia**, Lavras, v. 31, n. 2, p. 456-461, mar./abr. 2007.

NOBRE, C. M. B. **Análise espectral do biospeckle laser através do momento de inércia, entropia e cross-spectrum**. 2008. 197 p. Dissertação (Mestrado em Engenharia de Sistemas) - Universidade Federal de Lavras, Lavras.

OULAMARA, A.; TRIBILLON, G.; DUVERNOY, J. Biological activity measurement on botanical specimen surfaces using a temporal decorrelation effect of laser *Speckle*. **Optics and Lasers in Engineering**, London, v. 36, n. 2, p. 165-179, 1989.

PIRES, G. P.; TAGIAFERRI, A. A.; THAUMATURGO, C. **Interferometria óptica para ensaios não destrutivos**. Disponível em: <<http://www.materia.coppe.ufrj.br/mirror/sarra/artigos/artigo10069/index.html>>. Acesso em: 10 ago. 2007.

POMARICO, J. A.; DI ROCCO, H. O.; ALVAREZ, L.; LANUSSE, C.; MOTTIER, L.; SAUMELL, C.; ARIZAGA, R.; RABAL, H.; TRIVI, M. *Speckle* interferometry applied to pharmacodynamic studies: evaluation of parasite motility. **European Biophysics Journal**, New York, v. 33, n. 8, p. 694-699, 2004.

RABAL, H. J.; ARIZAGA, R.; CAP, L. N.; GRUMEL, E. E.; TRIVI, M. Numerical model for dynamic *speckle*: an approach using the movement of the scatterers. **Journal of Optics A: Pure and Applied Optics**, Bristol, v. 5, n. 5, p. 381-385, 2003.

RABAL, H. J.; BRAGA JUNIOR, R. A.; TRIVI, M.; RABELO, G. F. **Dynamic laser Speckle and applications**. New York: CRC, 2008. v. 1, 282 p.

RABELO, G. F. **Avaliação da aplicação do *speckle* dinâmico no monitoramento da qualidade da laranja**. 2000. 149 p. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual de Campinas, Campinas.

SILVA, E. R.; MURAMATSU, M. O fenômeno do *speckle* como introdução à metrologia óptica no laboratório didático. **Revista Brasileira de Ensino de Física**, São Paulo, v. 29, n. 2, p. 283-286, 2007.

XU, Z.; JOENATHAN, C.; KHORANA, B. M. Temporal and spatial proprieties of the time-varying *speckles* of botanical specimens. **Optical Engineering**, Redondo Beach, v. 34, n. 5, p. 1487-1502, 1995.