



LUCIENE RESENDE GONÇALVES

**SISTEMA DE ALARME ÓTIMO PARA O
MODELO TARSO COM APLICAÇÃO NA
FERRUGEM DO CAFÉ**

LAVRAS-MG

2011

LUCIENE RESENDE GONÇALVES

**SISTEMA DE ALARME ÓTIMO PARA O MODELO TARSO COM
APLICAÇÃO NA FERRUGEM DO CAFÉ**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientadora
Dra. Thelma Sáfydi

**LAVRAS-MG
2011**

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

Gonçalves, Luciene Resende.

Sistema de alarme ótimo para o modelo TARSO com aplicação na ferrugem do café / Luciene Resende Gonçalves. – Lavras : UFLA, 2011.

121 p. : il.

Tese (Doutorado) - Universidade Federal de Lavras, 2011.

Orientador: Thelma Sáfyadi.

Bibliografia.

1. Inferência bayesiana. 2. Molhamento foliar. 3. Temperatura média. 4. Threshold. 5. Variável climática. I. Universidade Federal de Lavras. II. Título.

CDD – 519.542

LUCIENE RESENDE GONÇALVES

**SISTEMA DE ALARME ÓTIMO PARA O MODELO TARSO COM
APLICAÇÃO NA FERRUGEM DO CAFÉ**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 15 de julho de 2011.

Dr. Augusto Ramalho de Moraes UFLA

Dr. Denismar Alves Nogueira UFLA

Dr. Renato Ribeiro de Lima UFLA

Dr. Rodrigo Luz da Cunha UFLA

Dra. Thelma Sáfadi
Orientadora

**LAVRAS-MG
2011**

Dedico à minha família
meus pais,
Edson e Márcia,
minha querida
avó Terexinha,
e aos meus irmãos,
Christiane, Márcio e
Lilane.

AGRADECIMENTOS

A Deus, por estar sempre presente, me guiando e dando forças para vencer os obstáculos.

A minha querida família pelos constantes incentivos.

À Universidade Federal de Lavras, pela oportunidade de aprimoramento acadêmico.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

Ao apoio, confiança, dedicação e amizade da orientadora, Thelma Sáfadi.

Aos membros da banca examinadora, pelas críticas e sugestões a este trabalho.

Aos professores do Departamento de Ciências Exatas, pelos conhecimentos transmitidos.

Ao Leonardo Bísvaro Japiassú da Fundação Procafé pela concessão dos dados utilizados nas aplicações deste trabalho.

Aos funcionários do Departamento de Ciências Exatas, sempre bem dispostos a ajudar.

Aos amigos Anderson e Deive que muito contribuíram para a realização deste trabalho.

À amizade conquistada em todas as repúblicas pelas quais passei.

À amizade e companheirismo dos amigos: Ademária, Anderson, Carlos, Claudiney, Fabrício, Geila, Gordo, Helder, Patrícia Siqueira, Renata, Taciana e dos amigos permanentes que fiz em Lavras. Vocês, sem dúvida, fizeram a diferença!

A todos, que direta ou indiretamente, contribuíram para a realização deste trabalho, muito obrigada!

RESUMO

O café ocupa posição de destaque na economia brasileira sendo o país um dos maiores produtores e exportadores de café do mundo. Além disso ocupa, em consumo, o segundo lugar. No entanto, a cultura cafeeira é muito suscetível ao ataque de doenças como a ferrugem, existindo registros de incidência na maioria das lavouras cultivadas. Os prejuízos podem chegar a comprometer 50% da produção. Dessa forma, pesquisas envolvendo esse fator limitante para a produção são intensas e frequentes, indicando os fatores ambientais como responsáveis no alastramento de epidemias, que não ocorrem se eles não forem favoráveis e atuarem simultaneamente. Em razão desses fatos, os objetivos desse trabalho são estender o modelo TARSO com dois regimes para duas variáveis independentes de forma a criar uma função para ajuste via inferência bayesiana no programa R e aplicar a metodologia de sistemas de alarme ótimos para processos TARSO de ordem p na criação de mais um mecanismo de previsão para a ferrugem do café. A extensão do modelo foi exemplificada para um TARSO de ordem $p = 1$ com uma variável independente ajustado via inferência bayesiana. Estudos de simulação foram feitos para testar o desempenho da extensão proposta, bem como aplicações a dados reais de índices de ferrugem utilizando uma variável independente e duas variáveis independentes. A análise com a variável independente temperatura média resultou em uma defasagem $d = 1$ e na contribuição da temperatura média na formação da ferrugem para valores de temperatura abaixo do valor limiar $r = 22,3^{\circ}C$. Já na análise utilizando as duas variáveis independentes, temperatura média e molhamento foliar, a defasagem também foi $d = 1$ e o índice de ferrugem sendo respondido, para temperaturas inferiores ao limiar $r = 21,7^{\circ}C$, pela temperatura média; para valores superiores a este limiar pelo molhamento foliar. A metodologia do alarme consistiu em prever se um nível de incidência qualquer da doença cruzou superiormente um nível u preestabelecido. Estudos de simulação foram feitos e os resultados da aplicação mostraram que o sistema se comporta de maneira eficiente em pontos onde a catástrofe ainda não ocorreu, não tendo o mesmo comportamento quando já ocorreu. E também que o sistema fornece probabilidades mais altas em níveis de cruzamento mais altos.

Palavras-chave: Ferrugem. Inferência bayesiana. Limiar. Sistemas de alarme. Variável climática.

ABSTRACT

Coffee crop occupies a featured position in the Brazilian economy and the country is the largest coffee producers and exporters in the world. Also, Brazil is in the second place when it considered the consumption. However, the coffee culture is very susceptible to attack diseases such as rust, existing incidence records in most of crops grown. The damage can compromise 50% of production. Thus, researches involving the limiting factor for production are intense and frequent, indicating the environmental factors as responsible for the spread epidemics, which do not occur if they are not friendly and act simultaneously. Because of these facts, the purposes of this study were to extend TARSO model with two regimes and one independent variable for two independent variables to create a function for adjusting via Bayesian inference in the R program. Besides of, applying the optimal alarm systems methodology for TARSO processes of p order in the development more a forecasting mechanism to coffee rust. In TARSO analysis, simulation studies were made to test the performance of the proposed extension, as well as applications the actual data to rust rates, using an independent variable and two independent variables. The analysis with the independent variable temperature average resulted in a lag $d=1$ and the average temperature contribution in the rust formation to temperature values below the threshold value $r = 22,3^{\circ}C$. In the analysis using two independent variables, average temperature and leaf wetness, the lag was also $d=1$ and the rust rate for temperatures below the threshold $r = 21,7^{\circ}C$ was answered by the average temperature. And for values above, the rust rate was answered by leaf wetness. The alarm methodology was exemplified for a TARSO of order $p=1$ with an independent variable being this model adjusted via Bayesian inference. This analysis consisted to provide if any level of disease incidence crossed upper a pre-established level. Simulation studies were made and the application results showed that the system behaves efficiently at points where the catastrophe has not occurred, not having the same behavior when it occurred. And also that the system provides higher probability in higher crossing levels.

Keywords: Rust. Bayesian inference. Threshold. Alarm systems. Climatic variable.

LISTA DE FIGURAS

Capítulo 1	15
Figura 1 Triângulo representativo da interação entre patógeno, hospedeiro e ambiente em uma doença de planta (AGRIOS, 1988)	20
Figura 2 Representação das interações entre os fatores que influenciam a severidade das doenças do cafeeiro (ZAMBOLIM et al., 1997)	21
Figura 3 Folhas de café atacadas com ferrugem: (a) folha com manchas isoladas; (b) folha toda coberta; (c) folha sob o efeito da aplicação de fungicida sistêmico; (d) folha sob o efeito da aplicação de fungicida sistêmico granulado via solo (ZAMBOLIM et al., 1997)	22
Capítulo 2	37
Figura 1 Representação gráfica da incidência da ferrugem (em %), observada no período de outubro de 1999 a agosto de 2007, em lavoura de alta produção, na região de Varginha, MG	66
Figura 2 Representação gráfica da temperatura média mensal (em $^{\circ}C$) e respectivo ponto limiar ($r = 21,7^{\circ}C$), lado esquerdo, e do molhamento foliar (em %), no período de outubro de 1999 a agosto de 2007, na região de Varginha, MG	67
Capítulo 3	88
Figura 1 Valores simulados em % de um processo TARSO(2;1,1,1,1)	107
Figura 2 Valores das probabilidades ($y_t \in G$) calculadas pela Expressão (6) em $t=4752$	108
Figura 3 Limites inferiores e superiores das regiões de alarme para as observações do grid em $t=4752$	109
Figura 4 Índices de ferrugem e temperatura média com valor <i>threshold</i> $r = 21,5^{\circ}C$	113
Figura 5 Valores de probabilidade e regiões de alarme do ponto $t = 133$	118

LISTA DE TABELAS

Capítulo 1	15
Capítulo 2	37
Tabela 1 Resultado da análise de regressão de mínimos quadrados para a sazonalidade presente na série de temperatura média.	68
Tabela 2 Estimativas bayesianas, erro padrão e fator de convergência, $\sqrt{\hat{R}}$, para cada parâmetro do modelo TARSO(2;1,1,0,1,0,1) ajustado com duas variáveis independentes.	70
Tabela 3 Resultado da análise de regressão de mínimos quadrados para a sazonalidade presente na série de temperatura média.	71
Tabela 4 Estimativas bayesianas, erro padrão e fator de convergência, $\sqrt{\hat{R}}$, para cada parâmetro do modelo TARSO(2;1,1,1,0) com uma variável independente.	73
Tabela 5 Estimativas bayesianas, erro padrão e fator de convergência, $\sqrt{\hat{R}}$, para cada parâmetro do modelo TARSO(2;1,0,1,0) com uma variável independente.	74
Capítulo 3	88
Tabela 1 Valores reais e simulados dos parâmetros do modelo TARSO(2;1,1,1,1).	106
Tabela 2 Nível de cruzamento, \bar{k}^* e região de alarme dos pontos analisados no sistema de alarme.	110
Tabela 3 Características operacionais do alarme em diferentes pontos para o modelo TARSO(2;1,1,1,1).	112
Tabela 4 Valores simulados do processo TARSO(2;1,1,1,1) em $t, t + 1, t + 2$	113
Tabela 5 Estimativas do modelo TARSO(2;1,0,1,1).	114
Tabela 6 Nível de cruzamento, \bar{k}^* , e região de alarme dos pontos de cruzamento analisados na construção do alarme.	115
Tabela 7 Características operacionais do alarme em $t = 12, t = 39$ e $t = 133$ do processo TARSO(2;1,0,1,1).	117
Tabela 8 Índices de ferrugem em $t, t + 1$ e $t + 2$	118

SUMÁRIO

	CAPÍTULO 1 Introdução geral	15
1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO	19
2.1	Doenças de plantas	19
2.1.1	Ferrugem do cafeeiro	20
2.2	Séries temporais	27
2.2.1	Modelos para séries temporais	27
2.2.2	Séries estacionárias	27
2.3	Inferência bayesiana	28
2.3.1	Teorema de Bayes	29
2.3.2	Distribuição a priori	31
2.3.3	Distribuição preditiva	32
	REFERÊNCIAS	34
	CAPÍTULO 2 Modelo TARSO: uma aplicação na ferrugem do café	37
1	INTRODUÇÃO	39
2	METODOLOGIA	42
2.1	Modelo TARSO com $k=2$ regimes	42
2.2	Análise bayesiana do modelo TARSO com 2 regimes	47
2.2.1	Análise a posteriori considerando uma priori própria	48
2.2.2	Análise a posteriori considerando a priori de Jeffreys	50
2.2.3	Densidade preditiva	51
2.3	Rotina computacional em R para o modelo TARSO	61
3	DADOS	64
4	RESULTADOS E DISCUSSÃO	65
4.1	Aplicação na ferrugem do cafeeiro	65
4.1.1	Modelo TARSO com duas variáveis independentes	65
4.1.2	Modelo TARSO com uma variável independente	71
5	CONCLUSÕES	76
	REFERÊNCIAS	77
	ANEXO	79
	CAPÍTULO 3 Sistema de alarme ótimo para processos TARSO .	88
1	INTRODUÇÃO	90
2	METODOLOGIA	94
2.1	Conceitos básicos em eventos de predição	94
2.2	Sistemas de alarme ótimos para processos TARSO com 2 regimes	98
2.3	Sistema de alarme ótimo para o TARSO(2;1,1,1,1) com 2 regimes	100

3	RESULTADOS E DISCUSSÃO	106
3.1	Estudo de simulação	106
3.2	Aplicação aos dados de ferrugem do café	113
4	CONCLUSÕES	118
	REFERÊNCIAS	120

CAPÍTULO 1

Introdução geral

1 INTRODUÇÃO

A chegada da cultura do café ao Brasil data de 1727. O produto se adaptou tão bem ao clima brasileiro que, a partir da década de 1850, já se encontrava bastante difundido entre as terras brasileiras. Com pouco tempo se tornou um marco na economia do país tornando-se a mais importante fonte de receitas e de divisas externas durante muitas décadas.

De acordo com Ferrugem... (2009), o Brasil é o maior produtor, maior exportador e 2º mercado consumidor de café do mundo, sendo a espécie *Coffea arabica* L. a principal commodity de exportação do país (AGRIANUAL, 2000). A maioria dos cultivares de café disponíveis para plantio, no entanto, são suscetíveis a algumas pragas e doenças. Isso causa grande preocupação aos produtores porque pode levar a baixas produtividades, consequentemente a prejuízos na produção e na economia dos cafeicultores. Dentre as doenças pode-se citar a ferrugem, cercosporiose e phoma; dentre as pragas tem-se o bicho mineiro, ácaro vermelho e broca.

Embora haja registros da incidência de todas as pragas e doenças citadas, a ferrugem do cafeeiro, cujo agente etiológico é o *Hemileia vastatrix* Berk. e Br, é considerada como sendo a principal doença da cultura em todo o mundo. De acordo com Zambolim et al. (1997), foi constatada pela primeira vez no Brasil, em

janeiro de 1970, no sul da Bahia; quatro meses depois, a doença já era encontrada na maioria das lavouras de café cultivadas no país. Os prejuízos na produção atingem cerca de 35%, em média, nas regiões onde as condições climáticas são favoráveis à doença. Sob estiagem prolongada, as perdas na produção podem chegar a mais de 50%.

Inicialmente essa doença era controlada com produtos protetores, principalmente aqueles à base de cobre (GODOY; BERGAMIN FILHO; SALGADO, 1997). Entretanto, o mal uso desses produtos resultam na resistência de determinados patógenos e, no caso exclusivo do café, observa-se que o uso contínuo de fungicidas sistêmicos associado a manejos nutricionais inadequados favorecem surtos de uma outra doença muito importante no ciclo da cultura que é a cercosporiose (*Cercospora coffeicola* Berk. Cook.) (ZAMBOLIM et al., 1997).

Diante da demanda atual por uma agricultura mais racional e menos agressiva ao meio ambiente, existe a tendência de aumentar a importância de sistemas que orientem os agricultores quanto ao reconhecimento das doenças importantes, seu monitoramento sistemático e o conhecimento da importância dos sistemas de aviso ou alerta fitossanitários (VALE; ZAMBOLIM; JESUS JÚNIOR, 2001).

Conforme Vale, Zambolim e Jesus Júnior (2001), o sistema de previsão de doença tem por objetivo auxiliar o agricultor na tomada de decisão a respeito de quando iniciar as medidas de controle de doenças e sua necessidade efetiva. A continuidade em se desenvolver novos sistemas de previsão de doença e uma melhor compreensão dos programas que já estão em andamento é uma aplicação desafiante para a epidemiologia de plantas.

As frequentes pesquisas envolvendo a ferrugem do café apontam os fatores ambientais como determinantes no alastramento de epidemias, que não ocorrem se eles não forem favoráveis e atuarem simultaneamente. Dessa forma, o objetivo

desse trabalho é utilizar uma extensão do modelo TARSO para desenvolver mais um mecanismo de previsão para detecção de uma catástrofe ocasionada pela ferrugem do cafeeiro embasado em informações sobre o clima com a finalidade de otimizar a utilização de produtos químicos. O modelo TARSO é um tipo de modelo de séries temporais que permite associar a incidência da ferrugem com várias variáveis climáticas.

O trabalho está organizado em três capítulos com duas contribuições ao estudo da ferrugem e é organizado como segue abaixo.

No capítulo 1, tem-se o referencial teórico com alguns conceitos básicos necessários ao desenvolvimento dos capítulos 2 e 3. Há uma breve revisão sobre a ferrugem do café, séries temporais e análise bayesiana. Sobre a ferrugem são evidenciados, principalmente, os fatores determinantes no alastramento da doença. Em séries temporais descreve-se um breve conceito do termo e a série estacionária. Na seção inferência bayesiana são descritas a distribuição a posteriori e a distribuição preditiva.

No capítulo 2, é apresentada a primeira contribuição ao estudo da ferrugem. Inicialmente, é apresentada uma revisão do modelo TARSO com 2 regimes e uma variável independente. Uma proposta de extensão do modelo com 2 regimes para 2 variáveis independentes é feita e uma função, para ajuste, é criada no programa R. A importância do TARSO está no fato de permitir que uma série temporal seja descrita em regimes; acima e abaixo de um limiar, como ocorre na ferrugem. Se, por exemplo, a temperatura média mensal estiver acima de um determinado valor, a ferrugem responde de uma forma e se estiver abaixo responde de outra.

No capítulo 3, é feita uma descrição dos conceitos teóricos básicos envolvendo a metodologia de sistema alarme ótimo de acordo com Antunes, Amaral-

Turkman e Turkman (2003). A segunda contribuição consiste em propor mais um mecanismo de previsão, por meio dessa metodologia, para a ferrugem do cafeeiro a partir de um modelo TARSO com 2 regimes e com uma variável independente ajustado via inferência bayesiana. A técnica consiste em prever se um nível de incidência qualquer da doença cruza superiormente um nível u preestabelecido.

Vale ressaltar que as ambas contribuições podem ser estendidas ao estudo de pragas e doenças de qualquer outra cultura que seja influenciada por fatores climáticos. Estudos de simulação foram feitos nos capítulos 2 e 3 e exemplificados com aplicação a dados reais.

2 REFERENCIAL TEÓRICO

2.1 Doenças de plantas

As plantas ficam doentes quando sofrem alterações anormais, em uma ou mais de suas funções, como a fotossíntese. Para Meira (2008), as doenças infecciosas são aquelas que resultam de infecção da planta por patógeno. São caracterizadas pela habilidade do patógeno (organismo vivo) de crescer e se multiplicar rapidamente nas plantas doentes e também de se propagar para plantas saudáveis. Além da planta (hospedeiro) e do patógeno, um conjunto de condições ambientais dentro de um intervalo favorável também deve ocorrer para que a doença se desenvolva. Zambolim et al. (1997) ressaltam que o grande desafio no manejo integrado das doenças consiste em se prever com antecedência quando determinada doença poderá causar dano econômico a determinada cultura. Sabe-se que as epidemias só ocorrerão se:

- o ambiente for favorável e perdurar o tempo suficiente para que ocorram vários ciclos do patógeno;
- houver raça virulenta e grande quantidade de inóculo do patógeno presente na área;
- a planta for suscetível e cultivada em grande escala.

O estudo das relações entre patógeno, hospedeiro e ambiente pode auxiliar na compreensão da ocorrência de epidemias e, conseqüentemente, permitir a aplicação de medidas de controle mais adequadas.

As interações dos três componentes de uma doença podem ser representadas por um triângulo, conforme apresentado na Figura 1.

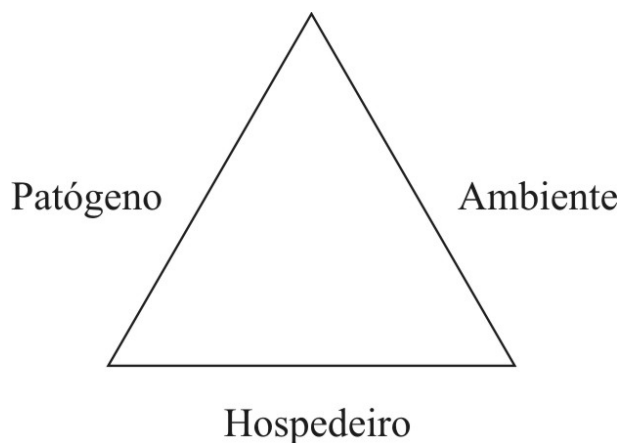


Figura 1 Triângulo representativo da interação entre patógeno, hospedeiro e ambiente em uma doença de planta (AGRIOS, 1988)

Se as plantas são resistentes ao patógeno, o lado do triângulo do hospedeiro (e a quantidade de doença) é pequeno ou inexistente. Quanto mais virulento, abundante e ativo, maior é o lado do triângulo do patógeno e maior a quantidade potencial da doença. Também, quanto mais favoráveis as condições ambientais, que ajudam o patógeno ou que reduzem a resistência do hospedeiro, maior é o lado do triângulo referente ao ambiente (AGRIOS, 1988).

No que se refere às doenças do cafeeiro, além dos três fatores responsáveis pelo surgimento de epidemias (patógeno, hospedeiro e ambiente), há que se considerar o tipo de solo, disponibilidade de nutrientes, a densidade de plantas por área e o estado nutricional, de acordo com o esquema apresentado na Figura 2.

2.1.1 Ferrugem do cafeeiro

A ferrugem do cafeeiro (*Coffea arabica* L.) constitui ainda um fator limitante na cultura do café, sendo uma doença foliar causadora de desfolha, comprometendo a produção futura (CUNHA; MENDES; CHALFOUN, 2004). Os sintomas da ferrugem podem ser observados na face inferior das folhas, na qual

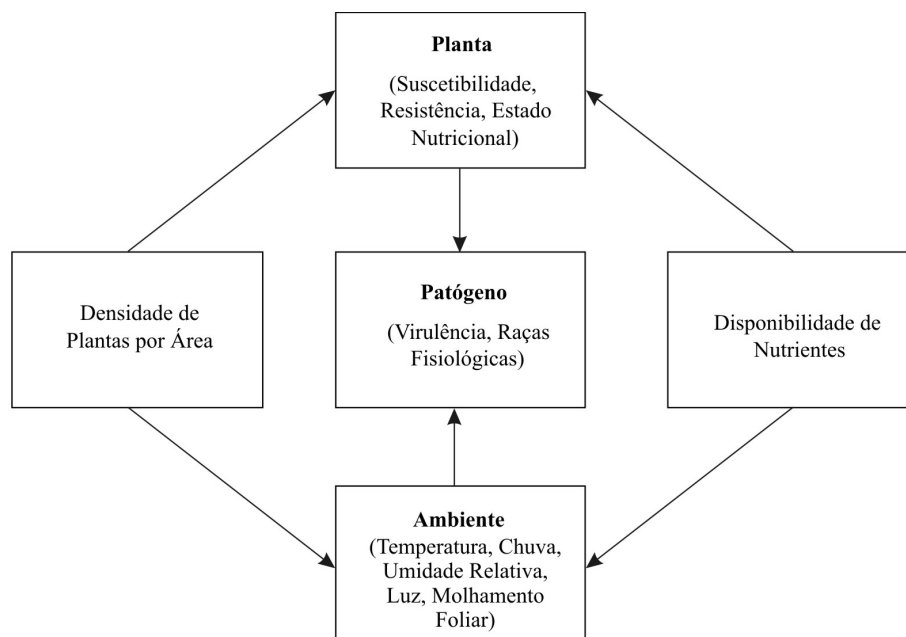


Figura 2 Representação das interações entre os fatores que influenciam a severidade das doenças do cafeeiro (ZAMBOLIM et al., 1997)

aparecem manchas de coloração amarelo-pálida, inicialmente pequenas, com 1 mm a 3 mm de diâmetro, que evoluem atingindo até 2 cm de diâmetro, quando então apresentam aspecto pulverulento (uredosporos) e coloração amarelo alaranjada característica da doença. Na face superior das folhas observam-se manchas cloróticas amareladas, correspondendo aos limites da pústula na face inferior, que posteriormente necrosam. Na Figura 3 são apresentadas duas folhas de café com ferrugem.

Segundo Montoya e Chaves (1974), conhecer os fatores que determinam a taxa de progresso da ferrugem é de grande importância, uma vez que eles condicionam a distribuição da doença, a sua incidência e a severidade.

Zambolim et al. (1997) afirmaram que, nos anos agrícolas de baixa produção, a doença não é severa. Porém, nos anos agrícolas de alta produção da lavoura

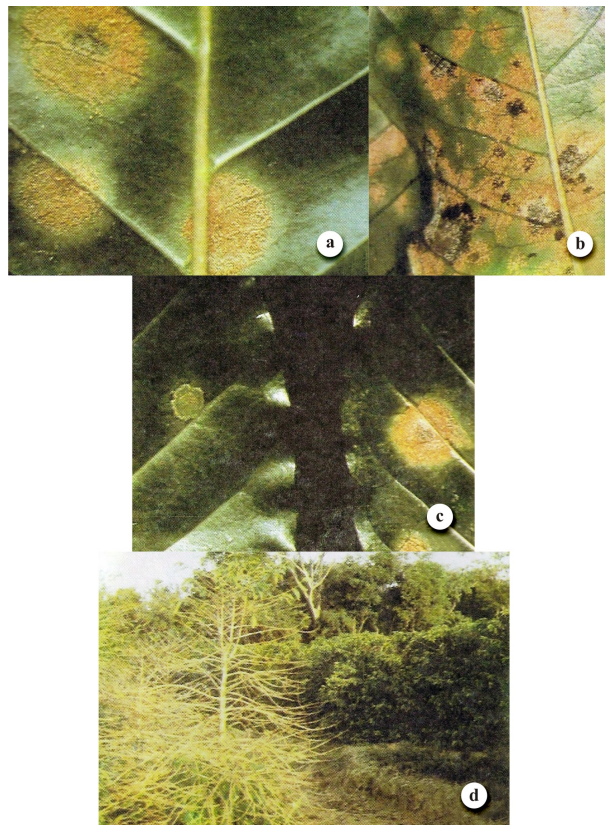


Figura 3 Folhas de café atacadas com ferrugem: (a) folha com manchas isoladas; (b) folha toda coberta; (c) folha sob o efeito da aplicação de fungicida sistêmico; (d) folha sob o efeito da aplicação de fungicida sistêmico granulado via solo (ZAMBOLIM et al., 1997)

é que se deve tomar maiores cuidados na proteção das plantas contra a incidência da ferrugem. Em tais situações, a temperatura, umidade, molhamento foliar e a chuva serão os fatores climáticos que, aliados ao índice da doença no campo (incidência da doença), deverão ser monitorados para se determinar o início, intervalo, número e até a escolha do tipo de fungicida a ser empregado nas aplicações.

Assim sendo, os modelos de previsão da ferrugem devem levar em consideração todos esses fatores para que seja tomada decisão sobre o controle químico

da doença por meio de pulverizações. A doença nos anos agrícolas de alta produção das plantas inicia-se em dezembro-janeiro e, de março a abril, aumentando em escala logarítmica até atingir o pico em junho, para depois decrescer a partir daí, devido às baixas temperaturas, à queda de folhas provocada pela colheita, à senescência natural e também devido ao fato de que a grande severidade da doença provoca intensa desfolha das plantas. No ano agrícola seguinte, como a produção será baixa, a doença não ultrapassará 30% a 35% de incidência, mesmo sob condições favoráveis do clima.

De acordo com Moraes (1983), epidemias não ocorrem se as condições ambientais não forem favoráveis, por mais que as condições relacionadas com o hospedeiro e o patógeno sejam favoráveis. Vários fatores do ambiente, atuando simultaneamente, interferem na ocorrência da ferrugem do cafeeiro. Variações regionais de incidência e os prejuízos causados pela ferrugem podem ser atribuídos principalmente aos fatores climáticos, através da coincidência de condições de temperatura e de umidade (intensidade, duração e frequência das chuvas; orvalho; e período de molhamento) favoráveis à doença, no período de outubro a março.

Segundo Kushalappa (1989), enquanto a superfície da folha está molhada, a temperatura é um dos fatores mais importantes na determinação da quantidade de esporos a ser germinada e penetrada. Para Zambolim et al. (2002), a temperatura variando de 22°C a 24°C é ótima para a germinação e a penetração do fungo nos estômatos de folhas da café. Temperaturas superiores a 30°C e inferiores a 14°C são limitantes para a infecção (KUSHALAPPA; AKUTSU; LUDWIG, 1983). Montoya e Chaves (1974) estimaram em $23,7^{\circ}\text{C}$ a temperatura ótima de germinação sobre mudas de cafeeiro.

Para Zambolim et al. (2002), a doença é disseminada a grandes distâncias pelo vento, podendo atingir lavouras situadas a milhares de quilômetros de

distância. Dentro da planta, a chuva exerce papel importante na disseminação do patógeno, pelo respingo das gotas de uma folha para outra. Este fato pode explicar a razão pela qual a doença é notada em focos nas lavouras, para depois disseminar-se de planta a planta.

Os uredosporos requerem água no estado líquido para germinar e penetrar nos estômatos situados na face inferior das folhas. A temperatura ótima para sua germinação está entre $21^{\circ}C$ e $23^{\circ}C$, na ausência de luz direta. Portanto, em regiões cafeeiras com temperaturas médias inferiores a $18^{\circ}C$ e superiores a $28^{\circ}C$, a doença pode não causar danos econômicos na produção, embora os sintomas ainda possam ser visíveis em algumas folhas das plantas.

Na adoção de medidas de controle da doença deve-se levar em consideração a carga pendente, a intensidade de enfolhamento das plantas e o clima. Geralmente, o período chuvoso marca o início de infecção das plantas ao ataque da ferrugem. Normalmente, o período infeccioso da ferrugem estende-se de dezembro a maio, podendo chegar até junho em algumas regiões, indicando portanto a época adequada para se efetuar o controle preventivo da doença.

Admite-se que o controle da ferrugem do café é bem sucedido quando a incidência da doença atinge no máximo 20% na época da colheita dos frutos.

Várias medidas podem ser utilizadas no controle da ferrugem, visando sempre a aliar o sucesso no controle da doença aos aspectos de segurança ambiental e do trabalhador (CUNHA; MENDES; CHALFOUN, 2004). Segundo Zambolim et al. (1997), o controle da ferrugem do cafeeiro pode ser feito pelo princípio da proteção com uso de fungicidas de contato, sendo os cúpricos os mais efetivos, ou fungicidas sistêmicos via foliar ou via solo, formulados com inseticidas sistêmicos ou separadamente. O controle da ferrugem também pode ser feito empregando-se uma mistura de nutrientes que apresenta efeito fungicida deno-

minada calda viçosa. Vários resultados demonstraram que a produtividade pode aumentar mais de 200% em relação às plantas não tratadas e cerca de 100% em relação aos produtos sistêmicos via foliar e solo. Entretanto, vale ressaltar que a aplicação indiscriminada de fungicida sistêmico misturado a inseticidas via solo numa mesma área ano após ano pode, a partir do terceiro ou quarto ano, modificar a fisiologia da planta a tal ponto que a lavoura pode manifestar reflexos negativos na produção. Essa mistura leva a um esgotamento das plantas após vários anos de aplicação.

Na Figura 3, podem ser observadas folhas de café infectadas por ferrugem sob o efeito de aplicação de fungicida sistêmico e fungicida sistêmico granulado, respectivamente.

A opção pelo controle da doença via foliar pode ser vinculada à intensidade de ataque da ferrugem nas plantas. Se a porcentagem de folhas doentes estiver em torno de 5%, recomenda-se iniciar o controle da doença com fungicidas de contato (à base de cobre); se ultrapassar 5% chegando até 12%, recomenda-se a aplicação de fungicidas sistêmicos.

Dada a importância da ferrugem do cafeeiro, vários trabalhos referentes ao estudo dessa doença podem ser encontrados na literatura. Pinto et al. (2002) avaliaram o potencial de redes neurais com dados da incidência da ferrugem do cafeeiro, coletados em Lavras de 13/02/1998 a 20/04/2001, como método alternativo aos sistemas fundamentais para descrever a epidemia da ferrugem do cafeeiro (*Coffea arabica* L.). As variáveis climáticas que foram utilizadas para construir as redes e elaborar equações de regressão foram precipitação, número de dias com e sem precipitação pluvial, umidade relativa média do ar, horas de insolação, temperaturas máxima, média e mínima, lâmina de água de irrigação e produção. Séries temporais da incidência da doença também foram empregadas na elaboração de

redes.

Meira (2008) desenvolveu uma árvore de decisão para analisar a epidemia da ferrugem do cafeeiro. A árvore demonstrou seu potencial como modelo simbólico e interpretável, levando assim à identificação das fronteiras de decisão e da lógica contidas nos dados, o que auxiliou na compreensão de quais variáveis e como as interações dessas variáveis condicionaram o progresso da doença no campo. Nesse estudo, as variáveis explicativas mais importantes foram a temperatura média nos períodos de molhamento foliar, a carga pendente de frutos, a média das temperaturas máximas diárias no período de incubação e a umidade relativa do ar. Desenvolveram modelos de alerta considerando taxas de infecção calculadas a partir de avaliações mensais de incidência da ferrugem agrupadas em três classes: TX1 - redução ou estagnação; TX2 - crescimento moderado e TX3 crescimento acelerado.

Garçon et al. (2004) avaliaram o controle da ferrugem do cafeeiro (*Coffea arabica* L.) baseando-se no valor de severidade, calculado em função das variáveis meteorológicas, molhamento foliar e temperatura média durante o período de molhamento.

Cunha, Mendes e Chalfoun (2004) avaliaram um melhor ajuste do controle da ferrugem do cafeeiro (*Coffea arabica* L.) durante a fase de pré-colheita e verificaram os seus efeitos sobre a produção e preservação do enfolhamento testando fungicidas preventivos à base de cobre (oxicloreto de cobre e sulfato de cobre) aplicados isoladamente e associados ao sistêmico epoxiconazole; utilizaram também o produto sistêmico aplicado isoladamente. Quando a incidência da doença era baixa, os fungicidas cúpricos aplicados preventivamente, isoladamente ou associados ao epoxiconazole foram eficientes no controle da ferrugem, preservando o enfolhamento e proporcionando boa produtividade.

2.2 Séries temporais

Uma série temporal é um conjunto de observações coletadas de forma sequencial, ao longo do tempo. A dependência entre as observações é o que caracteriza as aplicações das técnicas de séries temporais, já que as metodologias estatísticas clássicas, para serem aplicadas, exigem independência dos dados. Vale ressaltar que, além do tempo, uma série pode ser função de outra variável, como, por exemplo, espaço, profundidade, etc.

2.2.1 Modelos para séries temporais

Conforme Morettin e Toloi (2006), os modelos utilizados para descrever séries temporais são processos estocásticos, controlados por leis probabilísticas. Em séries temporais, um processo estocástico é definido como sendo uma coleção de variáveis aleatórias sequenciadas no tempo e definidas em um conjunto de pontos T , que pode ser contínuo ou discreto. A variável aleatória no tempo t é denotada por Y_t , em que $t = 0, \pm 1, \pm 2, \dots \in T$.

2.2.2 Séries estacionárias

Uma das suposições mais frequentes que diz respeito a uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Existem, tecnicamente, duas formas de estacionariedade: fraca (ou ampla, ou de segunda ordem) e estrita (ou forte). A estacionariedade forte é uma propriedade muito exigente e, em geral, de difícil verificação. Assim, faz sentido definir um conceito de estacionariedade baseado nos momentos de uma série temporal, mais precisamente nos momentos de primeira e segunda ordens. Isto é dado pela estacionariedade fraca. Maiores detalhes de um processo estocástico fracamente esta-

cionário ou estacionário de segunda ordem podem ser encontrados em Morettin e Toloi (2006).

A maior parte das séries temporais, na prática, são não-estacionárias e componentes como tendência $(Td)_t$ e sazonalidade S_t impedem o estágio estacionário de uma série. A tendência pode ser entendida como aumento ou diminuição gradual das observações ao longo do tempo e a sazonalidade indica possíveis flutuações ocorridas sempre em períodos menores ou iguais a doze meses. Considerando tais componentes, uma série pode ser escrita de forma aditiva como

$$Y_t = (Td)_t + S_t + a_t, \quad t = 1, \dots, n \quad (1)$$

em que a_t é a componente aleatória que mostra oscilações aleatórias irregulares.

A Equação (1) pode ser estacionarizada, isto é, as componentes tendência e sazonalidade podem ser estimadas e subtraídas da série de modo que a série é reescrita como

$$a_t = Y_t - (\hat{Td})_t - \hat{S}_t. \quad (2)$$

a_t na Equação (2) é então a série estacionária a partir da qual é desenvolvido qualquer modelo de séries temporais.

2.3 Inferência bayesiana

A abordagem bayesiana, extensivamente utilizada, só foi possível graças à iniciativa de Richard Price que publicou, em 1763, a obra póstuma do Rev. Thomas Bayes intitulada "An Essay Towards Solving a Problem in the Doctrine of Chances". Mas foi só em 1939 que Harold Jeffreys (JEFFREYS, 1939), reagindo contra a posição clássica, ressurgiu com o bayesianismo na resolução de problemas

até então sem solução.

Para Gelman et al. (2004), a característica essencial da inferência bayesiana é o uso explícito de probabilidades para quantificação de incertezas de uma quantidade de interesse θ . O verdadeiro valor de θ é desconhecido e a ideia é tentar reduzir este desconhecimento. Isso faz com que a inferência bayesiana esteja enraizada no conceito de probabilidade subjetiva, que mede o grau de confiança que alguém deposita no acontecimento de um determinado evento do espaço amostral. A especificação da incerteza sobre θ é feita por meio de uma densidade $P(\theta)$, denominada densidade a priori. De acordo com Paulino, Turkman e Murteira (2003), por ser uma probabilidade subjetiva, a incerteza a respeito de θ assume diferentes graus, o que conduz a modelos probabilísticos diferentes.

O processo de inferência bayesiana tem como base o teorema de Bayes e sua descrição é dada segundo Gelman et al. (2004).

2.3.1 Teorema de Bayes

A fim de fazer declarações de probabilidade sobre θ dado os dados y , começa-se com um modelo oriundo de uma distribuição de probabilidade conjunta para θ e y . A probabilidade conjunta ou função densidade $P(\theta, y)$ pode ser escrita como um produto de duas densidades que são frequentemente referidas como distribuição a priori $P(\theta)$ e distribuição dos dados $P(y|\theta)$,

$$P(\theta, y) = P(\theta)P(y|\theta).$$

Condicionando no valor dos dados y e usando a propriedade básica de probabilidade condicional conhecida como regra de Bayes, tem-se a densidade a

posteriori

$$P(\theta|y) = \frac{P(\theta, y)}{P(y)} = \frac{P(\theta)P(y|\theta)}{P(y)}, \quad (3)$$

em que $P(y) = \int P(\theta)P(y|\theta)d\theta$ para θ contínuo. Uma forma equivalente da Expressão (3) é omitir o fator $P(y)$ que não depende de θ e com y fixado pode ser considerado uma constante. Assim, $P(\theta|y)$ é o produto $P(\theta)P(y|\theta)$, ou seja,

$$P(\theta|y) \propto P(\theta)P(y|\theta). \quad (4)$$

Esta simples expressão encapsula o núcleo da técnica de inferência bayesiana: desenvolver o modelo $P(\theta, y)$ e o desempenho computacional necessário para resumir $P(\theta|y)$ na forma apropriada.

Na Expressão (4), $P(y|\theta)$ pode ser considerado como uma função não de y mas de θ . Quando assim considerada é chamada função de verossimilhança de θ dado y sendo denotada por $\mathcal{L}(\theta|y)$. Dessa forma, a fórmula de Bayes pode ser escrita como

$$P(\theta|y) \propto P(\theta)\mathcal{L}(\theta|y).$$

Em outras palavras o teorema de Bayes diz que a distribuição de probabilidade a posteriori para θ dado y é proporcional ao produto da distribuição a priori $P(\theta)$ pela verossimilhança de θ dado y .

Para Box e Tiao (1992), a função de verossimilhança $\mathcal{L}(\theta|y)$ desempenha um papel importante na fórmula de Bayes, pois é a função através da qual os dados y modificam o conhecimento a priori de θ , ou seja, representa a informação sobre θ advinda dos dados.

2.3.2 Distribuição a priori

A distribuição a priori desempenha um papel importante na análise bayesiana. Para Box e Tiao (1992), ela representa o que se sabe sobre os parâmetros desconhecidos antes de os dados estarem disponíveis e pode ser usada para representar o conhecimento a priori ou a ignorância relativa. Em problemas de inferência científica, normalmente se fosse possível, seria bom se os dados falassem por si. Por isso, geralmente é apropriado conduzir a análise como se um estado de ignorância relativa existisse a priori.

Segundo O'Hagan e Luce (2003), a crítica fundamental na abordagem bayesiana é o fato da informação a priori ser subjetiva, já que diferentes pesquisadores possuem informações a priori diferentes, conduzindo a diferentes distribuições a priori. Isso faz a distribuição a posteriori e todas as inferências derivadas dela, subjetivas. Nesse sentido toda a abordagem bayesiana é subjetiva.

Em muitas situações há um relativo consenso de qual a forma funcional que a distribuição a priori deve ter e o problema da sua completa especificação resume-se a especificar os seus parâmetros, chamados de hiperparâmetros (PAULINO; TURKMAN; MURTEIRA, 2003). A especificação da distribuição a priori deve ser tal que conduza à simplicidade da derivação analítica das distribuições a posteriori, o que pode ser alcançado por meio da atribuição do que se chama de famílias "conjugadas": quando a distribuição a posteriori segue a mesma forma paramétrica da distribuição a priori, esta é chamada de "conjugada". Segundo Gelman et al. (2004), a família conjugada é matematicamente conveniente porque, dessa forma, a distribuição a posteriori segue uma forma paramétrica conhecida. Formalmente conjugação pode ser definida como: se \mathcal{F} é uma classe de distribuições amostrais $P(y|\theta)$, e \mathcal{P} é uma classe de distribuições a priori para θ , então a

classe \mathcal{P} é conjugada de \mathcal{F} se

$$P(\theta|y) \in \mathcal{P} \text{ para todo } P(\cdot|\theta) \in \mathcal{F} \text{ e } P(\cdot) \in \mathcal{P}.$$

O interesse está em tomar \mathcal{P} como sendo o conjunto de todas as densidades que têm a mesma forma funcional da verossimilhança.

Quando as distribuições a priori não têm qualquer base populacional, elas podem ser difíceis de serem construídas, e é desejável que as distribuições a priori desempenhem um papel mínimo na distribuição a posteriori. Tais distribuições são às vezes chamadas distribuições a priori de referência, e a densidade a priori é descrita como vaga, difusa ou não informativa (GELMAN, 2004). Uma abordagem às vezes usada para definir distribuições a priori não informativas foi introduzida por Jeffreys (1961), cujo princípio leva a definir a densidade a priori não informativa como $P(\theta) \propto [J(\theta)]^{\frac{1}{2}}$, em que $J(\theta)$ é a informação de Fisher para θ . Esse princípio é chamado princípio da invariância de Jeffreys.

2.3.3 Distribuição preditiva

Inferência sobre uma observação desconhecida é frequentemente chamada inferência preditiva. Depois dos dados y serem observados pode-se prever \tilde{y} desconhecido. A distribuição de \tilde{y} é chamada distribuição a posteriori preditiva. Posteriori porque ela é condicional nos y observados e preditiva porque ela é a predição para um \tilde{y} que ainda não foi observado. A distribuição de \tilde{y} é dada por

$$\begin{aligned} P(\tilde{y}|y) &= \int P(\tilde{y}, \theta|y) d\theta \\ &= \int P(\tilde{y}|\theta, y) P(\theta|y) d\theta \\ &= \int P(\tilde{y}|\theta) P(\theta|y) d\theta. \end{aligned}$$

A segunda e a terceira linha mostram a distribuição preditiva a posteriori como uma média de predições condicionais sobre a distribuição a posteriori de θ . A última equação segue porque y e \tilde{y} são condicionalmente independentes dado θ . Quando não é possível (ou ineficiente) amostrar θ diretamente de $P(\theta|y)$, a simulação por cadeias de Markov é uma boa alternativa. Amostra-se iterativamente de forma que, a cada passo do processo, espera-se amostrar de uma distribuição mais próxima de $P(\theta|y)$ (GELMAN, 2004).

REFERÊNCIAS

- AGRIANUAL. **Anuário da agricultura brasileira**. São Paulo: FNP Consultoria e Comércio, 2000. 434 p.
- AGRIOS, G. N. **Plant pathology**. 3rd ed. San Diego: Academic, 1988. 803 p.
- ANTUNES, M.; AMARAL-TURKMAN, M. A.; TURKMAN, K. F. F. A bayesian approach to event prediction. **Journal of Time Series Analysis**, Clevedon, v. 24, n. 6, p. 631–646, Dec. 2003.
- BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. New York: J. Wiley Classics, 1992. 608 p.
- CUNHA, R. L.; MENDES, A. N. G.; CHALFOUN, S. M. Controle químico da ferrugem do cafeeiro (*Coffea arabica* L.) e seus efeitos na produção e preservação do enfolhamento. **Ciência e Agrotecnologia**, Lavras, v. 28, n. 5, p. 990-996, set./out. 2004.
- FERRUGEM do café (*Hemileia coffeicola*). Disponível em: <<http://www.revistacafeicultura.com.br/index.php?tipo=ler&mat=3563>>. Acesso em: 25 mar. 2009.
- GARÇON, C. L. P. et al. Controle da ferrugem do cafeeiro com base no valor de severidade. **Fitopatologia Brasileira**, Brasília, v. 29, n. 5, p. 486-491, set./out. 2004.
- GELMAN, A. et al. **Bayesian data analysis**. 2nd ed. London: Chapman and Hall/CRC, 2004. 695 p.
- GODOY, C. V.; BERGAMIN FILHO, A.; SALGADO, C. L. Doenças do cafeeiro. In: KIMATI, H. et al. (Ed.) **Manual de Fitopatologia: doenças das plantas cultivadas** 3. ed. São Paulo: Ceres, 1997. v. 2, p. 184-200.
- JEFFREYS, H. **Theory of probability**. London: Oxford University, 1939. 470 p.

JEFFREYS, H. **Theory of probability**. 3rd ed. London: Oxford University, 1961. 470 p.

KUSHALAPPA, A. C. Biology and epidemiology. In: KUSHALAPPA, A. C.; ESKES, A. B. **Coffe rust: epidemiology, resistance, and management**. Boca Raton: CRC, 1989. p. 13-80.

KUSHALAPPA, A. C.; AKUTSU, M.; LUDWIG, A. Application or survival ratio for monocyclic process of *Hemileia Vastatrix* in predicting coffe rust infection rates. **Phytopathology**, Saint Paul, v. 73, n. 1, p. 96-103, 1983.

MEIRA, C. A. A. **Processo de descoberta de conhecimento em base de dados para análise e o alerta de doenças de culturas agrícolas e sua aplicação na ferrugem do cafeeiro**. 2008. 198 p. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual de Campinas, Campinas, 2008.

MORAES, S. A. **A ferrugem do cafeeiro: importância, condições predisponentes, evolução e situação no Brasil**. Campinas: Instituto Agrônômico, 1983. 50 p. (IAC. Circular, 119).

MONTOYA, R. H.; CHAVES, G. M. Influência da temperatura e da luz na germinação, infectividade e período de geração de *Hemileia vastatrix* Berk e Br. **Experientiae**, Viçosa, MG, v. 18, n. 11, p. 239-266, 1974.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: E. Blücher, 2006. 538 p.

O'HAGAN, A.; LUCE, B. R. **A primer on bayesian statistics in health economics and outcomes research**. London: MEDTAP International, 2003. Disponível em: <<http://www.bayesian-initiative.com>>. Acesso em: 26 nov. 2010.

PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Fundação Calouste Gulbekian, 2003. 444 p.

PINTO, A. C. S. et al. Descrição da epidemia da ferrugem do cafeeiro com redes neuronais. **Fitopatologia Brasileira**, Brasília, v. 27, n. 5, p. 517-524, set./out. 2002.

VALE, F. X. R.; ZAMBOLIM, L.; JESUS JÚNIOR, W. C. **Epidemiologia aplicada ao controle de doenças de plantas**. Viçosa, MG: UFV, 2001. 182 p.

ZAMBOLIM, L. et al. Café (*Coffea arábica* L.): controle de doenças: doenças causadas por fungos, bactérias e vírus. In: VALE, F. X. R. do; ZAMBOLIM, L. (Ed.). **Controle de doenças de plantas: grandes culturas**. Viçosa, MG: UFV, 1997. v. 1, p. 83-139.

———. Epidemiologia e controle integrado da ferrugem-do-cafeeiro. In: ZAMBOLIM, L. (Ed.). **O estado da arte de tecnologias na produção de café**. Viçosa, MG: Suprema, 2002. p. 369-449.

CAPÍTULO 2

Modelo TARSO: uma aplicação na ferrugem do café

RESUMO

O modelo TARSO é um dos modelos não-lineares de séries temporais mais bem concebidos. A classe *threshold* TARSO modela séries temporais influenciadas por vários fatores, como acontece com as séries de fluxo, séries de níveis de água, níveis de doenças de plantas influenciadas por variáveis climáticas, dentre diversas outras situações. É um modelo descrito completamente pela estrutura dos dados. Neste capítulo foi feita breve descrição teórica e a análise bayesiana do modelo foi desenvolvida. Por meio dessa abordagem, o objetivo foi criar uma função, no programa R, para ajuste de séries que podem ser modeladas pelo TARSO. Um estudo de simulação foi feito para comprovar sua eficiência bem como aplicações a dados reais de incidência de ferrugem utilizando uma e duas variáveis independentes. A análise com a variável independente temperatura média resultou em uma defasagem $d = 1$ e na contribuição da temperatura média na formação da ferrugem para valores de temperatura abaixo do valor limiar $r = 22,3^{\circ}C$. Já na análise utilizando as duas variáveis independentes, temperatura média e molhamento foliar, a defasagem também foi $d = 1$ e o índice de ferrugem sendo respondido para temperaturas inferiores ao limiar $r = 21,7^{\circ}C$ pela temperatura média; para valores superiores a este limiar pelo molhamento foliar. Vários valores limiares no intervalo de $20^{\circ}C$ a $24^{\circ}C$ foram testados e também várias defasagens, mas apenas os valores citados conduziram a um modelo mais completo.

Palavras-chave: Limiar. Temperatura média. Molhamento foliar.

ABSTRACT

TARSO model is one of time series nonlinear models better designed. The threshold TARSO class models time series influenced by various factors, like the flow series, water levels series, levels of plant diseases influenced by climatic variables, among many other situations. This model is completely described by the data structure. In this chapter was made brief theoretical description and the Bayesian analysis of model was developed. Through this approach, the aim was to create a function, in the R program, for adjusting the series that can be modeled by TARSO. A simulation study was done to prove its efficiency as well as applications to actual data of rust incidence using one and two independent variables. The analysis with the independent variable average temperature resulted in a lag $d = 1$ and in the average temperature contribution in the formation of rust to temperature values below the threshold value $r = 22,3^{\circ}\text{C}$. In the analysis using the two independent variables, average temperature and leaf wetness, the lag was also $d = 1$ and the rust rate to temperatures below the threshold $r = 21,7^{\circ}\text{C}$ was responded by the average temperature. For values above this threshold, the rust was responded by leaf wetness. Several threshold values in range of 20°C to 24°C were tested and also several lags, but only cited values led to a more complete model.

Keywords: Threshold. Average temperature. Leaf wetness.

1 INTRODUÇÃO

Um modelo linear de séries temporais falha ao detectar certas propriedades tais como ciclos limites, frequências dependentes de amplitude e fenômenos de salto. Na tentativa de superar esta limitação, Tong (1978, 1983) propôs os modelos não-lineares e Tong e Lim (1980) criaram uma alternativa para séries temporais periódicas. Dentre os diversos modelos não lineares propostos para séries com tais propriedades pode-se dizer que os mais bem concebidos são o "*threshold*", os bilineares e os autorregressivos com coeficientes aleatórios. Exemplos de áreas de aplicação podem ser encontrados em Tong (1990), tais como ecologia (estudos de população), física solar, entomologia (dinâmica de população), economia, finanças, saúde, poluição, geofísica e hidrologia.

A classe *threshold*, especificamente, pode ser encontrada em Tong (1990). Nesta classe, o mais popular é o modelo autorregressivo com limiar (*TAR-Threshold autoregressive*) $TAR(k, p)$. Teoricamente um modelo não-linear pode ser separado em várias seções lineares usando diferentes valores *thresholds* de acordo com o grau de não-linearidade das séries temporais. Os valores *thresholds* são determinados para assegurar a estabilidade do sistema. Um modelo desse tipo é representado por

$$Y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} Y_{t-i} + a_t^{(1)} & , \text{ se } Z_t < r_1 \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} Y_{t-i} + a_t^{(2)} & , \text{ se } r_1 \leq Z_t < r_2 \\ \vdots & \\ \phi_{k0} + \sum_{i=1}^{p_k} \phi_{ki} Y_{t-i} + a_t^{(k)} & , \text{ se } r_{k-1} \leq Z_t \end{cases}$$

em que k é o número de regimes, $p_i, i = 1, \dots, k$ é a ordem da autorregressão

da equação do regime i e $p = \max(p_1, \dots, p_k)$ é a ordem do modelo. Em cada regime k , o modelo não-linear é apenas um processo ordinário autorregressivo de ordem p - $AR(p)$. Os *thresholds* são números reais $-\infty = r_0 < r_1 < \dots < r_{k-1} < r_k = +\infty$ e $Z_t = Z(Y_1, \dots, Y_{t-1})$ é a variável *threshold* que especifica a mudança no regime.

Um outro processo que faz parte da classe *threshold* foi proposto por Tsay (1998). Nesse processo, o autor considera um modelo de regressão *threshold* multivariada sendo chamado *threshold* autorregressivo misto, ou ainda, "*open loop threshold autoregressive*", abreviadamente TARSO.

O modelo TARSO surgiu para modelar séries temporais influenciadas por vários fatores, como acontece com as séries de fluxo, séries de níveis de água, níveis de doenças de plantas influenciadas por variáveis climáticas, dentre diversas outras situações. É um modelo descrito completamente pela estrutura dos dados.

Dentre algumas aplicações do modelo TARSO pode-se citar Sáfadi e Morettin (2001), que utilizaram o enfoque bayesiano na análise do modelo de regressão multivariada "*open loop*"(TARSO) por meio de uma priori própria e da priori de Jeffreys. Exemplificaram a metodologia considerando as séries de número de óbitos causados por problemas cardíacos e de temperatura mínima diária na cidade de São Paulo.

Jian et al. (1998) utilizaram o modelo TARSO para descrever a dinâmica do fluxo de água subterrânea na província de Shanxi, noroeste da China. O parâmetro *threshold* permitiu relacionar diferentes processos de precipitação, e o parâmetro de defasagem relacionou o tempo defasado entre eventos de precipitação e crescimento observado do fluxo de primavera. Os resultados na análise mostraram que o modelo descreveu com acurácia a série temporal não-linear.

A partir do trabalho de Sáfadi e Morettin (2001), é proposta uma extensão

do modelo. Serão utilizadas duas variáveis independentes na criação de uma função para modelagem bayesiana do TARSO no programa R (R DEVELOPMENT CORE TEAM, 2009). Uma aplicação à série de índices da ferrugem do cafeeiro influenciada pelas variáveis climáticas temperatura média e molhamento foliar será feita para mostrar como a metodologia funciona. Para isso, na seção 2, encontra-se descrita a metodologia do capítulo. Em 2.1 descreve-se a estrutura do modelo TARSO para 2 regimes, bem como a sua função de verossimilhança. Em 2.2 é desenvolvida a inferência bayesiana utilizando uma priori própria e a priori de Jeffreys, a partir da qual é desenvolvida a distribuição preditiva. Em 2.3 é proposta uma descrição detalhada de uma função para ajuste do modelo TARSO. Na seção 3 está descrito, em detalhes, o conjunto de observações a serem utilizadas no ajuste do modelo TARSO feito a partir da função proposta. Na seção 4 está os resultados e discussão de uma aplicação da função proposta utilizando 2 situações: TARSO com duas variáveis independentes e TARSO com uma variável independente.

2 METODOLOGIA

2.1 Modelo TARSO com k=2 regimes

Um modelo *threshold* com dois regimes e valor *threshold* r é o modelo mais simples (TONG; LIM, 1980). A extensão de dois regimes para regimes múltiplos e com mais de um valor *threshold* é trivial. O modelo de uma série temporal Y_t , em que Y_t é determinado por seus valores históricos e afetado pelos valores históricos da série temporal X_t , é descrito a seguir.

Sejam $\{Y_t$ e $X_t, t = 0, \pm 1, \pm 2, \dots\}$ um processo estocástico satisfazendo

$$Y_t = \phi_{l_0} + \sum_{i=1}^{p_l} \phi_{l_i} Y_{t-i} + X_t + a_t^{(l)}, \quad \text{se } Z_{t-d} \in R_l \quad (1)$$

para $l = 1, \dots, k$ regimes e as suposições:

- (i) $\{a_t^{(l)}\}$ é uma sequência de variáveis aleatórias i.i.d. $\sim N(0, \tau_l^{-1})$ em que a precisão $\tau_l > 0$ e $\tau_l^{-1} = \sigma_l^2 = \text{Var}(a_t^{(l)})$ é desconhecida, para cada l ;
- (ii) R_l é um intervalo da reta \mathbb{R} tal que $R_i \cap R_j = \emptyset$, se $i \neq j$ e $\bigcup_{l=1}^k R_l = \mathbb{R}$, isto é, $\{R_1, \dots, R_k\}$ é uma partição da reta com $k - 1$ variáveis "patamar";
- (iii) d é um parâmetro de retardo pertencente ao conjunto $\{1, \dots, d_0\}$;
- (iv) X_t é uma função de variáveis independentes, isto é, $X_t = f(T_t, U_t, \dots)$. X_t é dada por $X_t = \sum_{\nu=0}^{m_l} \beta_{l\nu} T_{t-\nu} + \sum_{u=0}^{q_l} \theta_{lu} U_{t-u} + \dots$;
- (v) Z_{t-d} determina os regimes do modelo e pode ser qualquer variável que compõe a soma de X_t , ou seja, T_{t-d}, U_{t-d} , etc.. Z_{t-d} é chamada variável "patamar, limiar ou *threshold*", independente de $a_t^{(l)}$, para todo l e todo t ;

(vi) as k seqüências são supostas serem independentes entre si.

Considerando (para simplificação da descrição) um modelo com $k = 2$ regimes e X_t como sendo função de duas variáveis independentes T_t e U_t , a Equação (1) é escrita como

$$Y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} Y_{t-i} + \sum_{\nu=0}^{m_1} \beta_{1\nu} T_{t-\nu} + \sum_{u=0}^{q_1} \theta_{1u} U_{t-u} + a_t^{(1)} & , \text{ se } T_{t-d} \leq r \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} Y_{t-i} + \sum_{\nu=0}^{m_2} \beta_{2\nu} T_{t-\nu} + \sum_{u=0}^{q_2} \theta_{2u} U_{t-u} + a_t^{(2)} & , \text{ se } T_{t-d} > r \end{cases}, \quad (2)$$

sendo r o parâmetro de limiar e T_t escolhida, por conveniência, para ser a variável *threshold*. Essa variável pode ser qualquer outra que compõe a função X_t .

A representação da Equação (2) diz que Y_t segue dois diferentes processos com diferentes ordens $(p_1, p_2, m_1, m_2, q_1, q_2)$ que dependem do valor da variável limiar T_{t-d} . Denota-se por $\text{TARSO}(k = 2; p_1, p_2, m_1, m_2, q_1, q_2)$ ou $\text{TARSO}(2; p_1, p_2, m_1, m_2, q_1, q_2)$. A notação seguida neste trabalho será $\text{TARSO}(2; p_1, p_2, m_1, m_2, q_1, q_2)$.

Supondo que as ordens (p_i, m_i, q_i) , $i = 1, 2$ sejam conhecidas, os parâmetros do modelo são (γ_i, τ_i) em que $\gamma_i = (\phi_{i0}, \phi_{i1}, \dots, \phi_{ip_i}, \beta_{i0}, \beta_{i1}, \dots, \beta_{im_i}, \theta_{i0}, \theta_{i1}, \dots, \theta_{iq_i})'$, $i = 1, 2$. Fazendo $X_{it} = (1, Y_{t-1}, Y_{t-2}, \dots, Y_{t-p_i}, T_t, T_{t-1}, \dots, T_{t-m_i}, U_t, U_{t-1}, \dots, U_{t-q_i})'$, com $i = 1, 2$, o Modelo (2) pode ser reescrito matricialmente como

$$Y_t = \begin{cases} \gamma_1' X_{1t} + a_t^{(1)} & , \text{ se } T_{t-d} \leq r; \\ \gamma_2' X_{2t} + a_t^{(2)} & , \text{ se } T_{t-d} > r. \end{cases} \quad (3)$$

Sejam $D = \{Y_t, T_t, U_t, t = p + 1, \dots, n\}$ o conjunto de todas as obser-

vações, $p = \max\{p_1, p_2, m_1, m_2, q_1, q_2\}$, o valor máximo no qual o modelo é condicionado e n_1 e n_2 o número de observações em cada regime. Condicionando nas p -primeiras observações a função de verossimilhança condicional pode ser aproximada por

$$\mathcal{L}(\Theta|d, r, D) \propto \tau_1^{\frac{n_1}{2}} \tau_2^{\frac{n_2}{2}} \exp \left\{ -\frac{\tau_1}{2} \sum_{t=p+1}^n (Y_t - \gamma'_1 X_{1t})^2 - \frac{\tau_2}{2} \sum_{t=p+1}^n (Y_t - \gamma'_2 X_{2t})^2 \right\}, \quad (4)$$

em que $\sum_{t=p+1}^n (Y_t - \gamma'_1 X_{1t})^2$ é a soma em t no regime 1, isto é, se $T_{t-d} \leq r$, $\sum_{t=p+1}^n (Y_t - \gamma'_2 X_{2t})^2$ é a soma em t no regime 2, ou seja, se $T_{t-d} > r$ e $\Theta = (\gamma_1, \gamma_2, \tau_1, \tau_2)$.

Denotando por $\sum_t (Y_t - \gamma'_i X_{it})^2$ a soma em $t = p + 1$ no regime $i = 1, 2$ tem-se que,

$$\begin{aligned} \sum_t (Y_t - \gamma'_i X_{it})^2 &= \sum_t (Y_t - \gamma'_i X_{it})(Y_t - \gamma'_i X_{it})' = \sum_t Y_t^2 - \sum_t Y_t X_{it}' \gamma_i \\ &\quad - \gamma_i' \sum_t X_{it} Y_t + \gamma_i' \sum_t X_{it} X_{it}' \gamma_i. \end{aligned} \quad (5)$$

Utilizando a notação de matrizes, os somatórios da Expressão (5) podem ser reescritos e conseqüentemente a função de verossimilhança (4). Para isto sejam

$$\sum_t X_{it} Y_t = \mathbf{B}_i^{(m)} = \begin{pmatrix} \mathbf{B}_{1i}^{(m)} \\ \mathbf{B}_{2i}^{(m)} \\ \mathbf{B}_{3i}^{(m)} \end{pmatrix}$$

e

$$\sum_t^m X_{it}X'_{it} = \mathbf{A}_i^{(m)} = \begin{pmatrix} \mathbf{A}^{(m)}_{11} & \mathbf{A}^{(m)}_{12} & \mathbf{A}^{(m)}_{13} \\ \mathbf{A}^{(m)}_{21} & \mathbf{A}^{(m)}_{22} & \mathbf{A}^{(m)}_{23} \\ \mathbf{A}^{(m)}_{31} & \mathbf{A}^{(m)}_{32} & \mathbf{A}^{(m)}_{33} \end{pmatrix},$$

em que

$$\mathbf{B}_{1i}^{(m)} = \begin{pmatrix} \sum_t^m Y_t \\ \sum_t^m Y_t Y_{t-1} \\ \cdot \\ \cdot \\ \sum_t^m Y_t Y_{t-p_i} \end{pmatrix}, \quad \mathbf{B}_{2i}^{(m)} = \begin{pmatrix} \sum_t^m Y_t T_t \\ \sum_t^m Y_t T_{t-1} \\ \cdot \\ \cdot \\ \sum_t^m Y_t T_{t-m_i} \end{pmatrix},$$

$$\mathbf{B}_{3i}^{(m)} = \begin{pmatrix} \sum_t^m Y_t U_t \\ \sum_t^m Y_t U_{t-1} \\ \cdot \\ \cdot \\ \sum_t^m Y_t U_{t-q_i} \end{pmatrix}$$

e

$$\begin{aligned}
\mathbf{A}^{(\mathbf{m})^i}_{11} &= \begin{pmatrix} \sum_t^m 1 & \sum_t^m Y_{t-1} & \dots & \sum_t^m Y_{t-p_i} \\ \sum_t^m Y_{t-1} & \sum_t^m Y_{t-1}Y_{t-1} & \dots & \sum_t^m Y_{t-1}Y_{t-p_i} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_t^m Y_{t-p_i} & \sum_t^m Y_{t-1}Y_{t-p_i} & \dots & \sum_t^m Y_{t-p_i}Y_{t-p_i} \end{pmatrix}, \\
\mathbf{A}^{(\mathbf{m})^i}_{12} &= \begin{pmatrix} \sum_t^m T_t & \sum_t^m T_{t-1} & \dots & \sum_t^m T_{t-m_i} \\ \sum_t^m Y_{t-1}T_t & \sum_t^m Y_{t-1}T_{t-1} & \dots & \sum_t^m Y_{t-1}T_{t-m_i} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_t^m Y_{t-p_i}T_t & \sum_t^m Y_{t-p_i}T_{t-1} & \dots & \sum_t^m Y_{t-p_i}T_{t-m_i} \end{pmatrix}, \\
\mathbf{A}^{(\mathbf{m})^i}_{13} &= \begin{pmatrix} \sum_t^m U_t & \sum_t^m U_{t-1} & \dots & \sum_t^m U_{t-q_i} \\ \sum_t^m Y_{t-1}U_t & \sum_t^m Y_{t-1}U_{t-1} & \dots & \sum_t^m Y_{t-1}U_{t-q_i} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_t^m Y_{t-p_i}U_t & \sum_t^m Y_{t-p_i}U_{t-1} & \dots & \sum_t^m Y_{t-p_i}^m U_{t-q_i} \end{pmatrix}, \\
\mathbf{A}^{(\mathbf{m})^i}_{22} &= \begin{pmatrix} \sum_t^m T_t T_t & \sum_t^m T_t T_{t-1} & \dots & \sum_t^m T_t T_{t-m_i} \\ \sum_t^m T_{t-1} T_t & \sum_t^m T_{t-1} T_{t-1} & \dots & \sum_t^m T_{t-1} T_{t-m_i} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_t^m T_{t-m_i} T_t & \sum_t^m T_{t-m_i} T_{t-1} & \dots & \sum_t^m T_{t-m_i} T_{t-m_i} \end{pmatrix},
\end{aligned}$$

$$\mathbf{A}^{(\mathbf{m})^i}_{23} = \begin{pmatrix} \sum_t^m T_t U_t & \sum_t^m T_t U_{t-1} & \dots & \sum_t^m T_t U_{t-q_i} \\ \sum_t^m T_{t-1} U_t & \sum_t^m T_{t-1} U_{t-1} & \dots & \sum_t^m T_{t-1} U_{t-q_i} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_t^m T_{t-m_i} U_t & \sum_t^m T_{t-m_i} U_{t-1} & \dots & \sum_t^m T_{t-m_i} U_{t-q_i} \end{pmatrix},$$

$$\mathbf{A}^{(\mathbf{m})^i}_{33} = \begin{pmatrix} \sum_t^m U_t U_t & \sum_t^m U_t U_{t-1} & \dots & \sum_t^m U_t U_{t-q_i} \\ \sum_t^m U_{t-1} U_t & \sum_t^m U_{t-1} U_{t-1} & \dots & \sum_t^m U_{t-1} U_{t-q_i} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_t^m U_{t-q_i} U_t & \sum_t^m U_{t-q_i} U_{t-1} & \dots & \sum_t^m U_{t-q_i} U_{t-q_i} \end{pmatrix},$$

$\mathbf{A}^{(\mathbf{m})^i}_{21} = (\mathbf{A}^{(\mathbf{m})^i}_{12})'$, $\mathbf{A}^{(\mathbf{m})^i}_{31} = (\mathbf{A}^{(\mathbf{m})^i}_{13})'$ e $\mathbf{A}^{(\mathbf{m})^i}_{32} = (\mathbf{A}^{(\mathbf{m})^i}_{23})'$. Com isso, fazendo $m = n$, a função de verossimilhança (4) torna-se

$$\mathcal{L}(\Theta|d, r, D) \propto \tau_1^{\frac{n_1}{2}} \tau_2^{\frac{n_2}{2}} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} \left[C^{(n)} - \mathbf{B}^{(\mathbf{n})'}_i \mathbf{A}^{(\mathbf{n})^{-1}}_i \mathbf{B}^{(\mathbf{n})}_{i+} + (\gamma_i - \mathbf{A}^{(\mathbf{n})^{-1}}_i \mathbf{B}^{(\mathbf{n})}_{i+})' \mathbf{A}^{(\mathbf{n})}_i (\gamma_i - \mathbf{A}^{(\mathbf{n})^{-1}}_i \mathbf{B}^{(\mathbf{n})}_{i+}) \right] \right\}, \quad (6)$$

sendo $C^{(m)} = \sum_t^m Y_t^2$.

2.2 Análise bayesiana do modelo TARSO com 2 regimes

Nesta subseção será desenvolvida a distribuição a posteriori do modelo dado em (3) considerando uma priori própria e a priori imprópria de Jeffreys (1961). A densidade preditiva será desenvolvida utilizando a priori de Jeffreys (1961). As condicionais completas para cada parâmetro serão utilizadas na cons-

trução de uma rotina computacional em R para ajuste do modelo TARSO. A incidência da ferrugem do cafeeiro associada às variáveis climáticas temperatura média e molhamento foliar ilustrarão o funcionamento da rotina desenvolvida. A densidade preditiva será utilizada no próximo capítulo para criação de um sistema de alarme.

2.2.1 Análise a posteriori considerando uma priori própria

A posteriori conjunta $P(\Theta|r, d, D)$ do modelo TARSO apresentado em (3) é dada por

$$P(\Theta|r, d, D) \propto \mathcal{L}(\Theta|r, d, D)P(\gamma_1, \gamma_2, \tau_1, \tau_2)$$

em que $\mathcal{L}(\Theta|r, d, D)$ é a função de verossimilhança (6). Neste caso, r e d são supostamente conhecidos, contudo poderiam ser desconhecidos e assim uma priori seria associada a cada um. $P(\gamma_1, \gamma_2, \tau_1, \tau_2)$ ou $P(\Theta)$ é a priori conjunta atribuída aos parâmetros γ_i e τ_i , $i = 1, 2$. A priori conjunta é o produto de prioris próprias independentes dadas por:

- γ_1 e γ_2 - independentes com distribuição Normal $(\gamma_{0i}, \tau_i^{-1}\mathbf{Q}_i^{-1})$, $i = 1, 2$.
- τ_1 e τ_2 - independentes com distribuição Gama $\left(\frac{\alpha_i}{2}, \frac{\beta_i}{2}\right)$, $i = 1, 2$, em que

γ_{0i} , \mathbf{Q}_i , α_i e β_i , $i = 1, 2$ são hiperparâmetros supostamente conhecidos, sendo $\alpha_i > 0$, $\beta_i > 0$, $\gamma_{0i} \in R^{p_i+m_i+q_i+1}$ e \mathbf{Q}_i é uma matriz definida positiva de dimensão $(p_i + m_i + q_i + 1) \times (p_i + m_i + q_i + 1)$.

Dessa forma, a distribuição a priori conjunta dos parâmetros é dada por

$$P(\Theta) \propto \tau_1^{\frac{\alpha_1+P_1}{2}-1} \tau_2^{\frac{\alpha_2+P_2}{2}-1} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} [\beta_i + (\gamma_i - \gamma_{0i})' \mathbf{Q}_i (\gamma_i - \gamma_{0i})] \right\}, \quad (7)$$

em que $P_i = p_i + m_i + q_i + 1, i = 1, 2$.

A distribuição a posteriori de $(\Theta|r, d, D)$ é uma normal-gama.

Demonstração:

Combinando (6) com (7) obtem-se,

$$P(\Theta|d, r, D) \propto \tau_1^{\frac{n_1}{2}} \tau_2^{\frac{n_2}{2}} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} \boldsymbol{\mu} \right\} \times \tau_1^{\frac{\alpha_1+P_1}{2}-1} \tau_2^{\frac{\alpha_2+P_2}{2}-1} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} [\beta_i + (\gamma_i - \gamma_{0i})' \mathbf{Q}_i (\gamma_i - \gamma_{0i})] \right\},$$

em que

$$\boldsymbol{\mu} = \left[C^{(n_i)} - \mathbf{B}^{(\mathbf{n}_i)'} \mathbf{A}^{(\mathbf{n}_i)^{-1}} \mathbf{B}^{(\mathbf{n}_i)} + (\gamma_i - \mathbf{A}^{(\mathbf{n}_i)^{-1}} \mathbf{B}^{(\mathbf{n}_i)})' \mathbf{A}^{(\mathbf{n}_i)} (\gamma_i - \mathbf{A}^{(\mathbf{n}_i)^{-1}} \mathbf{B}^{(\mathbf{n}_i)}) \right].$$

Expandindo as expressões $\boldsymbol{\mu}$ e $(\gamma_i - \gamma_{0i})' \mathbf{Q}_i (\gamma_i - \gamma_{0i})$ tem-se

$$P(\Theta|d, r, D) \propto \tau_1^{\frac{n_1+\alpha_1+P_1}{2}-1} \tau_2^{\frac{n_2+\alpha_2+P_2}{2}-1} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} \left[\beta_i + C^{(n_i)} + \right. \right. \\ \left. \left. - \mathbf{B}^{(\mathbf{n}_i)'} \mathbf{A}^{(\mathbf{n}_i)^{-1}} \mathbf{B}^{(\mathbf{n}_i)} + \gamma_i' \mathbf{A}^{(\mathbf{n}_i)} \gamma_i - \gamma_i' \mathbf{B}^{(\mathbf{n}_i)} - \mathbf{B}^{(\mathbf{n}_i)'} \gamma_i + \right. \right. \\ \left. \left. + \mathbf{B}^{(\mathbf{n}_i)'} \mathbf{A}^{(\mathbf{n}_i)^{-1}} \mathbf{B}_i + \gamma_i' \mathbf{Q}_i \gamma_i - \gamma_i' \mathbf{Q}_i \gamma_{0i} - \gamma_{0i}' \mathbf{Q}_i \gamma_i + \right. \right. \\ \left. \left. + \gamma_{0i}' \mathbf{Q}_i \gamma_{0i} \right] \right\}.$$

Colocando termos comuns em evidência, obtem-se

$$\begin{aligned}
P(\Theta|d, r, D) &\propto \tau_1^{\frac{n_1+\alpha_1+P_1}{2}-1} \tau_2^{\frac{n_2+\alpha_2+P_2}{2}-1} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} \left[\beta_i + C^{(n_i)} + \right. \right. \\
&\quad + \gamma'_{0i} \mathbf{Q}_i \gamma_{0i} + \gamma'_i (\mathbf{A}^{(n_i)}_i + \mathbf{Q}_i) \gamma_i - \gamma'_i (\mathbf{B}^{(n_i)}_i + \mathbf{Q}_i \gamma_{0i}) + \\
&\quad \left. \left. - (\mathbf{B}^{(n_i)}_i)' + \gamma'_{0i} \mathbf{Q}_i \right) \gamma_i \right\}.
\end{aligned}$$

Completando quadrado em γ_i , obtem-se a normal-gama

$$\begin{aligned}
P(\Theta|d, r, D) &\propto \tau_1^{\frac{n_1+\alpha_1+P_1}{2}-1} \tau_2^{\frac{n_2+\alpha_2+P_2}{2}-1} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} \left[\beta_i + C^{(n_i)} + \right. \right. \\
&\quad + \gamma'_{0i} \mathbf{Q}_i \gamma_{0i} - (\mathbf{B}^{(n_i)}_i + \mathbf{Q}_i \gamma_{0i})' (\mathbf{A}^{(n_i)}_i + \mathbf{Q}_i)^{-1} \\
&\quad (\mathbf{B}^{(n_i)}_i + \mathbf{Q}_i \gamma_{0i}) + (\gamma_i - (\mathbf{A}^{(n_i)}_i + \mathbf{Q}_i)^{-1} \\
&\quad (\mathbf{B}^{(n_i)}_i + \mathbf{Q}_i \gamma_{0i}))' (\mathbf{A}^{(n_i)}_i + \mathbf{Q}_i) (\gamma_i - (\mathbf{A}^{(n_i)}_i + \mathbf{Q}_i)^{-1} \\
&\quad \left. \left. (\mathbf{B}^{(n_i)}_i + \mathbf{Q}_i \gamma_{0i})) \right] \right\}. \tag{8}
\end{aligned}$$

As distribuições condicionais completas para γ_i e τ_i são obtidas da expressão (8) e são dadas por

$$\begin{aligned}
\tau_i | \gamma_i, d, r, D &\sim \text{Gama} \left(\frac{n_i + \alpha_i}{2}, \frac{1}{2} (\beta_i + \boldsymbol{\mu}) \right) \text{ e} \\
\gamma_i | \tau_i, d, r, D &\sim \text{Normal} \left((\mathbf{A}^{(n_i)}_i + \mathbf{Q}_i)^{-1} (\mathbf{B}^{(n_i)}_i + \mathbf{Q}_i \gamma_{0i}) \right. \\
&\quad \left. [(\tau_i (\mathbf{A}^{(n_i)}_i + \mathbf{Q}_i))]^{-1} \right), \quad i = 1, 2.
\end{aligned}$$

2.2.2 Análise a posteriori considerando a priori de Jeffreys

No caso de se ter pouca ou nenhuma informação a priori, pode-se utilizar a priori de Jeffreys,

$$P(\Theta) \propto \tau_1^{-1} \tau_2^{-1}.$$

Neste caso, a distribuição a posteriori é também uma normal-gama dada por

$$P(\Theta|d, r, D) \propto \tau_1^{\frac{n_1}{2}-1} \tau_2^{\frac{n_2}{2}-1} \exp \left\{ \sum_{i=1}^2 -\frac{\tau_i}{2} \boldsymbol{\mu} \right\}. \quad (9)$$

Para obter a Expressão (9) é suficiente tomar $\mathbf{Q} \rightarrow \mathbf{0}$, $\beta \rightarrow 0$ e $\alpha_i = -(p_i + m_i + q_i + 1)$ na posteriori conjunta obtida na expressão (8).

As novas condicionais completas são imediatas e dadas por

$$\begin{aligned} \tau_i | \gamma_i, d, r, D &\sim \text{Gama} \left(\frac{n_i - (p_i + m_i + q_i + 1)}{2}, \frac{1}{2} \boldsymbol{\mu} \right) \text{ e} \\ \gamma_i | \tau_i, d, r, D &\sim \text{Normal}(\mathbf{A}^{(n_i)}_i^{-1} \mathbf{B}^{(n_i)}_i, (\tau_i \mathbf{A}^{(n_i)}_i)^{-1}), \quad i = 1, 2. \end{aligned}$$

2.2.3 Densidade preditiva

A densidade preditiva será desenvolvida para o regime i , em que $i = 1, 2$. Considerando uma série de tamanho $t = n$, seja $P(y_{n+h}|d, r, y_1, y_2, \dots, y_{n+h-1})$ a densidade preditiva de uma observação futura y_{n+h} para qualquer $h \in \mathbb{N}$. Combinando a verossimilhança de j observações, Y_{n+1}, \dots, Y_{n+j} com a posteriori obtida na Expressão (9) obtém-se a distribuição conjunta das observações Y_{n+1}, \dots, Y_{n+j} e dos parâmetros. Integrando-se em relação aos parâmetros obtém-se que a preditiva de $(Y_{n+1}, \dots, Y_{n+j}|d, r, D)$ é um produto de distribuições t-Student univariadas, isto é,

$$P(y_{n+1}, \dots, y_{n+j}|d, r, D) = \prod_{h=1}^j P(y_{n+h}|d, r, y_1, y_2, \dots, y_{n+h-1}).$$

Tomando $\mathbf{y} = (d, r, y_1, y_2, \dots, y_{n+h-1})$ tem-se que

$$P(y_{n+h}|\mathbf{y}) \propto \frac{1}{2} \left[\sum_{t=p+1}^{n+h} Y_t^2 - \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) \right]^{-\frac{n_i+h+1}{2}}. \quad (10)$$

Em notação de matrizes, fazendo $m = n + h$, a expressão (10) torna-se

$$P(y_{n+h}|\mathbf{y}) \propto \frac{1}{2} \left[C^{(n+h)} - (\mathbf{B}_i^{(n+h)})' (\mathbf{A}_i^{(n+h)})^{-1} (\mathbf{B}_i^{(n+h)}) \right]^{-\frac{n_i+h+1}{2}}. \quad (11)$$

De acordo com Broemeling (1985), a distribuição t-Student obtida em (11) é não-central com $\nu = n_i + h$ graus de liberdade, média $E^{(h)}$ e variância $V^{(h)}$. Ou seja,

$$P(y_{n+h}|\mathbf{y}) \sim St(\nu, E^{(h)}, V^{(h)})$$

sendo

$$E^{(h)} = \frac{\left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h}}{1 - X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h}}$$

e

$$V^{(h)} = \left[\frac{\sum_{t=p+1}^{n+h-1} Y_t^2 - \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right)^{-1}}{\left(1 - X_{in+h}' \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right)^{-1} X_{in+h} \right)} + \frac{\left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)}{\left(1 - X_{in+h}' \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right)^{-1} X_{in+h} \right)} - (E^{(h)})^2 \right] \frac{1}{\nu}.$$

Demonstração:

A distribuição preditiva bayesiana é a densidade condicional das observações futuras dadas as observações passadas. Dessa forma, a demonstração da Expressão (11) é desenvolvida por meio do produto entre a posteriori dada pela Expressão (9) e pela função de verossimilhança aproximada de y_{n+h} condicionada às observações passadas, isto é, $\mathcal{L}(y_{n+h}|\gamma_i, \tau_i, d, r, D)$. O produto é desenvolvido e após o completamento de quadrado e integração em γ_i , integra-se em relação a τ_i para obter a Expressão (11). Assim, tem-se

$$\begin{aligned} P(y_{n+h}, \gamma_i, \tau_i | \mathbf{y}) &\propto P(\gamma_i, \tau_i | d, r, D) \times \mathcal{L}(y_{n+h} | \gamma_i, \tau_i, d, r, D) \\ &\propto \tau_i^{\frac{n_i}{2}} \exp \left\{ -\frac{\tau_i}{2} \sum_{t=p+1}^n (Y_t - \gamma_i' X_{it})^2 \times \tau_i^{-1} \times \tau_i^{\frac{h}{2}} \right. \\ &\quad \left. \exp \left\{ -\frac{\tau_i}{2} (Y_{n+h} - \gamma_i' X_{in+h})^2 \right\} \right\} \\ &\propto \tau_i^{\frac{n_i+h-1}{2}} \exp \left\{ -\frac{\tau_i}{2} \sum_{t=p+1}^{n+h} (Y_t - \gamma_i' X_{it})^2 \right\} \end{aligned}$$

$$\begin{aligned}
P(y_{n+h}, \gamma_i, \tau_i | \mathbf{y}) &\propto \tau_i^{\frac{n_i+h-1}{2}} \exp \left\{ -\frac{\tau_i}{2} \sum_{t=p+1}^{n+h} (Y_t - \gamma_i' X_{it})(Y_t - \gamma_i' X_{it})' \right\} \\
&\propto \tau_i^{\frac{n_i+h-1}{2}} \exp \left\{ -\frac{\tau_i}{2} \left(\sum_{t=p+1}^{n+h} Y_t^2 - \sum_{t=p+1}^{n+h} Y_t X_{it}' \gamma_i + \right. \right. \\
&\quad \left. \left. - \sum_{t=p+1}^{n+h} \gamma_i' X_{it} Y_t + \sum_{t=p+1}^{n+h} \gamma_i' X_{it} X_{it}' \gamma_i \right) \right\} \\
&\propto \tau_i^{\frac{n_i+h-1}{2}} \exp \left\{ -\frac{\tau_i}{2} \left[\sum_{t=p+1}^{n+h} Y_t^2 + \gamma_i' \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right) \gamma_i + \right. \right. \\
&\quad \left. \left. - \gamma_i' \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) - \left(\sum_{t=p+1}^{n+h} Y_t X_{it}' \right) \gamma_i \right] \right\}.
\end{aligned}$$

Completando-se quadrado em γ_i tem-se,

$$\begin{aligned}
P(y_{n+h}, \gamma_i, \tau_i | \mathbf{y}) &\propto \tau_i^{\frac{n_i+h-1}{2}} \exp \left\{ -\frac{\tau_i}{2} \left(\sum_{t=p+1}^{n+h} Y_t^2 \right) \right\} \exp \left\{ -\frac{\tau_i}{2} \left[\right. \right. \\
&\quad \left(\gamma_i - \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right)^{-1} \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) \right)' \\
&\quad \sum_{t=p+1}^{n+h} X_{it} X_{it}' \left(\gamma_i - \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right)^{-1} \right. \\
&\quad \left. \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) \right) - \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right)' \\
&\quad \left. \left. \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right)^{-1} \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) \right] \right\}.
\end{aligned}$$

Integrando-se em relação a γ_i , obtém-se

$$P(y_{n+h}, \tau_i | \mathbf{y}) \propto \tau_i^{\frac{n_i+h-1}{2}} \exp\left\{-\frac{\tau_i}{2} \left[\sum_{t=p+1}^{n+h} Y_t^2 + \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) \right]\right\}.$$

Completando a expressão acima para uma distribuição Gama e integrando em relação a τ_i , tem-se

$$P(y_{n+h} | \mathbf{y}) \propto \frac{1}{2} \left[\sum_{t=p+1}^{n+h} Y_t^2 - \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) \right]^{-\frac{n_i+h+1}{2}}. \quad (12)$$

A Expressão (12), em notação de matrizes, torna-se

$$P(y_{n+h} | \mathbf{y}) \propto \frac{1}{2} \left[C^{(n+h)} - (\mathbf{B}_i^{(n+h)})' (\mathbf{A}_i^{(n+h)})^{-1} (\mathbf{B}_i^{(n+h)}) \right]^{-\frac{n_i+h+1}{2}} \quad (13)$$

que é a Expressão dada em (11).

Então, para qualquer h , a distribuição preditiva de $(Y_{n+h} | \mathbf{y})$ dada pela Expressão (13) é uma distribuição t-Student não-central com $\nu = n_i + h$ graus de liberdade, média $E^{(h)}$ e variância $V^{(h)}$.

Em notação de matrizes, $E^{(h)}$ e $V^{(h)}$ são dadas por

$$E^{(h)} = \frac{(\mathbf{B}_i^{(n+h-1)})' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}}{(1 - X'_{in+h} (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h})} \quad (14)$$

e

$$V^{(h)} = \left\{ \frac{C^{(n+h-1)} - (\mathbf{B}_i^{(n+h-1)})' (\mathbf{A}_i^{(n+h)})^{-1} (\mathbf{B}_i^{(n+h-1)})}{1 - X'_{in+h} (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}} - \left(\frac{(\mathbf{B}_i^{(n+h)})' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}}{1 - X'_{in+h} (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}} \right)^2 \right\} \frac{1}{(n_i + h)}, \quad (15)$$

sendo n_i o número de elementos no regime i .

Demonstração

A demonstração de $E^{(h)}$ e $V^{(h)}$, Equações (14) e (15) respectivamente, são feitas a partir da Expressão (10). As somas dessa expressão são desmembradas e após uma série de operações algébricas a demonstração é concluída. Assim,

$$\begin{aligned} P(y_{n+h}|\mathbf{y}) &\propto \frac{1}{2} \left[\sum_{t=p+1}^{n+h-1} Y_t^2 + Y_{n+h}^2 - \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t + \right. \right. \\ &\quad \left. \left. + X_{in+h} Y_{n+h} \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t + \right. \right. \\ &\quad \left. \left. + X_{in+h} Y_{n+h} \right) \right]^{-\frac{n_i+h+1}{2}} \\ P(y_{n+h}|\mathbf{y}) &\propto \frac{1}{2} \left[\sum_{t=p+1}^{n+h-1} Y_t^2 + Y_{n+h}^2 - \left(\left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \right. \right. \\ &\quad \left. \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} + Y'_{n+h} X'_{in+h} \right. \\ &\quad \left. \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \right] \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t + \right. \\ &\quad \left. \left. + X_{in+h} Y_{n+h} \right) \right]^{-\frac{n_i+h+1}{2}} \end{aligned}$$

$$\begin{aligned}
P(y_{n+h}|\mathbf{y}) &\propto \frac{1}{2} \left\{ \sum_{t=p+1}^{n+h-1} Y_t^2 + Y_{n+h}^2 - \left[\left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \right. \right. \\
&\quad \left. \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right) + \right. \\
&\quad + \left. \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} (X_{in+h} Y_{n+h}) + \right. \\
&\quad + \left. Y'_{n+h} X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right) + \right. \\
&\quad + \left. Y'_{n+h} X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \right. \\
&\quad \left. (X_{in+h} Y_{n+h}) \right\}^{-\frac{n_i+h+1}{2}} \\
P(y_{n+h}|\mathbf{y}) &\propto \frac{1}{2} \left\{ \sum_{t=p+1}^{n+h-1} Y_t^2 - \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \right. \\
&\quad \left. \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right) + Y'_{n+h} Y_{n+h} - Y'_{n+h} X'_{in+h} \right. \\
&\quad \left. \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h} Y_{n+h} - Y'_{n+h} X'_{in+h} \right. \\
&\quad \left. \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right) + \right. \\
&\quad - \left. \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \right. \\
&\quad \left. (X_{in+h} Y_{in+h}) \right\}^{-\frac{n_i+h+1}{2}} \\
P(y_{n+h}|\mathbf{y}) &\propto \frac{1}{2} \left\{ \sum_{t=p+1}^{n+h-1} Y_t^2 - \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \right. \\
&\quad \left. \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right) + Y'_{n+h} (1 - X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \right. \right.
\end{aligned}$$

$$\begin{aligned}
& X_{in+h}Y_{n+h} - Y'_{n+h}X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} \\
& \left(\sum_{t=p+1}^{n+h-1} X_{it}Y_t \right) - \left(\sum_{t=p+1}^{n+h-1} X_{it}Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} \\
& \left(X_{in+h}Y_{n+h} \right) \left\}^{-\frac{n_i+h+1}{2}} \\
P(y_{n+h}|\mathbf{y}) & \propto \frac{1}{2} \left\{ \sum_{t=p+1}^{n+h-1} Y_t^2 - \left(\sum_{t=p+1}^{n+h-1} X_{it}Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} \right. \\
& \left. \left(\sum_{t=p+1}^{n+h-1} X_{it}Y_t \right) + (Y_{n+h} - (1 - X'_{in+h} \right. \\
& \left. \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} X_{in+h})^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it}Y_t \right)' \right. \\
& \left. \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} X_{in+h} \right)' (1 - X'_{in+h} \\
& \left. \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} X_{in+h}) (Y_{n+h} - (1 + \right. \\
& \left. - X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} X_{in+h})^{-1} \right. \\
& \left. \left(\sum_{t=p+1}^{n+h-1} X_{it}Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it}X'_{it} \right)^{-1} X_{in+h} \right) + \\
& \left. - E^{(h)'} \Sigma E^{(h)} \right\}^{-\frac{n_i+h+1}{2}},
\end{aligned}$$

sendo

$$\begin{aligned}
 E^{(h)'} \Sigma E^{(h)} &= X'_{in+h} \left(\sum_t^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_t^{n+h-1} X_{it} Y_t \right) (1 - X'_{in+h} \\
 &\quad \left(\sum_t^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h})^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \\
 &\quad \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h}).
 \end{aligned}$$

Daí,

$$P(y_{n+h} | \mathbf{y}) \propto \frac{1}{2} \left[F + (Y_{n+h} - E^{(h)})' \Sigma (Y_{n+h} - E^{(h)}) \right]^{-\frac{n_i+h+1}{2}}$$

com

$$\begin{aligned}
 E^{(h)} &= \left(1 - X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h} \right)^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \\
 &\quad \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h}), \\
 \Sigma &= \left(1 - X'_{in+h} \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} X_{in+h} \right) \mathbf{e} \\
 F &= \sum_{t=p+1}^{n+h-1} Y_t^2 - \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \\
 &\quad \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right) - E^{(h)'} \Sigma E^{(h)}.
 \end{aligned}$$

Dessa forma

$$\begin{aligned}
P(y_{n+h}|\mathbf{y}) &\propto F \frac{1}{2} [1 + (Y_{n+h} - E^{(h)})' \Sigma F^{-1} \\
&\quad (Y_{n+h} - E^{(h)})]^{-\frac{n_i+h+1}{2}} \\
&\propto \frac{1}{2} [1 + \frac{1}{\nu} (Y_{n+h} - E^{(h)})' \Sigma F^{-1} \nu \\
&\quad (Y_{n+h} - E^{(h)})]^{-\frac{n_i+h+1}{2}}
\end{aligned}$$

sendo $\nu = n_i + h$. Ou ainda,

$$P(y_{n+h}|\mathbf{y}) \propto \frac{1}{2} \left[1 + \frac{\Sigma F^{-1} \nu}{\nu} (Y_{n+h} - E^{(h)})^2 \right]^{-\frac{\nu+1}{2}}$$

em que

$$\begin{aligned}
\Sigma F^{-1} \nu &= \frac{1 - X'_{in+h} (\sum_{t=p+1}^{n+h} X_{it} X'_{it})^{-1} X_{in+h}}{\boldsymbol{\mu}_1} - \\
&\quad \frac{(1 - X'_{in+h} (\sum_{t=p+1}^{n+h} X_{it} X'_{it})^{-1} X_{in+h})^2}{(\sum_{t=p+1}^{n+h} X_{it} Y_t)' (\sum_{t=p+1}^{n+h} X_{it} X'_{it})^{-1} X_{in+h}} \nu
\end{aligned}$$

e

$$\frac{F}{\Sigma \nu} = \left[\frac{\boldsymbol{\mu}_1}{(1 - X'_{in+h} (\sum_{t=p+1}^{n+h} X_{it} X'_{it})^{-1} X_{in+h})} - \left(\frac{(\sum_{t=p+1}^{n+h-1} X_{it} Y_t)' (\sum_{t=p+1}^{n+h} X_{it} X'_{it})^{-1} X_{in+h}}{1 - X'_{in+h} (\sum_{t=p+1}^{n+h} X_{it} X'_{it})^{-1} X_{in+h}} \right)^2 \right] \frac{1}{\nu}$$

com

$$\boldsymbol{\mu}_1 = \sum_{t=p+1}^{n+h-1} Y_t^2 - \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X'_{it} \right)^{-1} \left(\sum_{t=p+1}^{n+h-1} X_{it} Y_t \right).$$

Portanto

$$V^{(h)} = \frac{F}{\Sigma\nu}.$$

Centralizando a distribuição dada em (12), obtem-se

$$W_{n+h} = [V^{(h)}]^{-\frac{1}{2}}(Y_{n+h} - E^{(h)}). \quad (16)$$

2.3 Rotina computacional em R para o modelo TARSO

A rotina elaborada para modelagem bayesiana do modelo TARSO foi desenvolvida no software R (R DEVELOPMENT CORE TEAM, 2009). O modelo considerado foi aquele dado pela Expressão (2), sendo

$$Y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} Y_{t-i} + \sum_{\nu=0}^{m_1} \beta_{1\nu} T_{t-\nu} + \sum_{u=0}^{q_1} \theta_{1u} U_{t-u} + a_t^{(1)} & , \text{ se } T_{t-d} \leq r \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} Y_{t-i} + \sum_{\nu=0}^{m_2} \beta_{2\nu} T_{t-\nu} + \sum_{u=0}^{q_2} \theta_{2u} U_{t-u} + a_t^{(2)} & , \text{ se } T_{t-d} > r \end{cases}$$

em que

1. $\tau_i^{-1} = \sigma_i^2$ é a precisão, sendo $\tau_i > 0$ para $i = 1, 2$;
2. T_{t-d} é uma das variáveis independentes e também variável *threshold*; d o valor de retardo e r o valor de limiar;
3. $\gamma_i = (\phi_{i0}, \phi_{i1}, \dots, \phi_{ip_i}, \beta_{i0}, \beta_{i1}, \dots, \beta_{im_i}, \theta_{i0}, \theta_{i1}, \dots, \theta_{iq_i})'$ é o vetor de parâmetros supondo as ordens (p_i, m_i, q_i) conhecidas, $i = 1, 2$ e $X_{it} = (1, Y_{t-1}, \dots, Y_{t-p_i}, T_t, T_{t-1}, \dots, T_{t-m_i}, \dots, U_t, U_{t-1}, \dots, U_{t-q_i})'$ a matriz de delineamento.

A inferência bayesiana desse modelo foi desenvolvida na seção (2.2), na qual obteve-se as distribuições condicionais completas dos parâmetros, dadas por

$$\gamma_i | \tau_i, d, r, D \sim Normal[\mathbf{A}^{(n)}_i^{-1} \mathbf{B}^{(n)}_i, (\tau_i \mathbf{A}^{(n)}_i)^{-1}], \quad i = 1, 2 \quad (17)$$

e

$$\tau_i | \gamma_i, d, r, D \sim Gama \left[\frac{n_i - (p_i + m_i + 1)}{2}, \frac{1}{2} \left(C^{(n)} - \mathbf{B}'_i \mathbf{A}^{(n)}_i^{-1} \mathbf{B}^{(n)}_i + (\gamma_i - \mathbf{A}^{(n)}_i^{-1} \mathbf{B}^{(n)}_i)' \mathbf{A}^{(n)}_i (\gamma_i - \mathbf{A}^{(n)}_i^{-1} \mathbf{B}^{(n)}_i) \right) \right], \quad i = 1, 2. \quad (18)$$

Por meio dessas condicionais, implementou-se a função para estimar os parâmetros do modelo TARSO via cadeias de Markov.

A elaboração da função seguiu os seguintes passos:

1. Foi gerado o TARSO a partir de 2 conjuntos de séries estacionárias T e U sendo cada uma constituída de 1000 observações de um modelo ARIMA. T e U são as séries independentes e T é também uma variável *threshold*;
2. estabelecimento dos vetores de parâmetros γ_i e τ_i , defasagem d e limiar r ;
3. geração de 1000 observações da série dependente Y utilizando as informações dos 2 passos anteriores;
4. divisão das observações de Y , T e U em 2 regimes pela variável *threshold* T segundo o critério: se $T_{t-d} \leq r$, Y_t , T_t e U_t pertencem ao regime 1 e se $T_{t-d} > r$, Y_t , T_t e U_t pertencem ao regime 2 para $\{t = 1, \dots, n\}$;
5. determinação das matrizes de delineamento dos regimes 1 e 2 para estimação bayesiana de γ_i e τ_i por meio das Equações (17) e (18);
6. diagnóstico bayesiano.

A rotina elaborada para ajuste do modelo TARSO encontra-se no anexo deste capítulo.

3 DADOS

Os dados a serem utilizados na comprovação da eficiência da função criada referem-se ao acompanhamento mensal da incidência de ferrugem no cafeeiro em (%), temperatura média e molhamento foliar. Esses dados foram coletados por Japiassú et al. (2007) no período de outubro de 1999 a agosto de 2007, na fazenda experimental da Fundação Procafé, localizada em Varginha, MG. As observações de incidência da ferrugem e temperatura média para o período de setembro de 2007 a fevereiro de 2011 foram obtidas no *site* da Fundação Procafé (2011). Para o molhamento foliar, neste período, os dados não foram disponibilizados.

A coleta dos dados de incidência da ferrugem é feita no mês de setembro, de cada ano, tomando uma lavoura de café plantada em espaçamento do tipo adensado (2,5 metros entre linhas e 0,5 metros entre plantas resultando em uma densidade média de 8000 plantas/ha). A lavoura possui uma alta carga pendente de frutos (com mais de 30 sacas beneficiadas/ha). Durante o ano agrícola (setembro a agosto), não há controle da doença na lavoura observada.

O processo de amostragem foi realizado de acordo com Chalfoun (1998): coleta de 100 folhas do terço médio das plantas da lavoura, no terceiro ou quarto par de folhas; contagem do número de folhas com lesões de ferrugem; e determinação da incidência (percentual de folhas atacadas).

Os dados mensais de temperatura média em ($^{\circ}C$) e molhamento foliar em (%) são resultantes da média das médias observadas diariamente.

4 RESULTADOS E DISCUSSÃO

A aplicação na ferrugem foi feita utilizando o modelo TARSO para duas variáveis independentes e uma variável dependente. Foram considerados diferentes limiares r no intervalo de $20^{\circ}C$ a $24^{\circ}C$ e diferentes defasagens d . A razão para esta escolha intervalar do limiar r baseia-se em valores referenciados na literatura como ideais ao progresso da doença.

4.1 Aplicação na ferrugem do cafeeiro

4.1.1 Modelo TARSO com duas variáveis independentes

Essa primeira análise foi feita utilizando incidência da ferrugem observada em lavoura com espaçamento largo e com alta carga pendente de frutos (lapaf). As variáveis independentes foram as séries climáticas de temperatura média e molhamento foliar. As observações aqui referem-se apenas ao período de outubro de 1999 a agosto de 2007 devido à falta das observações subsequentes de molhamento foliar, totalizando 95 dados. O objetivo é apenas exemplificar a metodologia já que as observações não estão atualizadas. Todas as séries são periódicas, com período de 365 dias, e foram dessazonalizadas por regressão.

Nas Figuras 1 e 2 são apresentados, respectivamente, os gráficos da variável de saída lapaf (Y_t) e das variáveis de entrada temperatura média (T_t) e molhamento foliar (U_t). A variável T_t é também a variável *threshold* com limiar $r = 21,7^{\circ}C$. O limiar divide as observações em dois grupos que podem ser chamados de regimes. No gráfico da temperatura média, Figura 2, o limiar está representado pela reta horizontal. Para cada regime é ajustado um modelo.

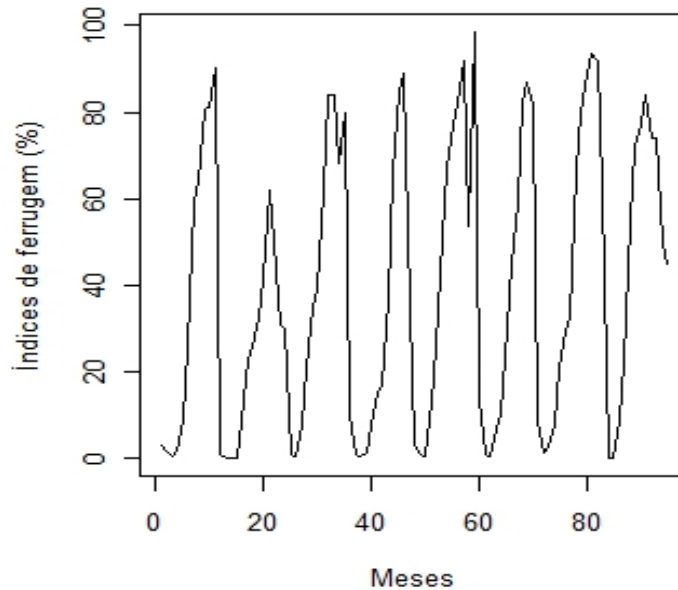


Figura 1 Representação gráfica da incidência da ferrugem (em %), observada no período de outubro de 1999 a agosto de 2007, em lavoura de alta produção, na região de Varginha, MG

A razão para escolha de $r = 21,7^{\circ}C$ está no fato de ser o valor de temperatura que conduziu a um modelo mais completo entre os vários valores testados no intervalo de $20^{\circ}C$ a $24^{\circ}C$. Nesse intervalo, segundo a literatura, ocorrem as condições favoráveis ao progresso da doença.

No ajuste do modelo TARSO, as séries são dessazonalizadas e dessa forma o valor *threshold* $r = 21,7^{\circ}C$ se transforma em 0,4110. Isso é feito subtraindo a temperatura $21,7^{\circ}C$ que ocorre em $t = 51$ (dezembro de 2003) do valor da sazonalidade estimada em $t = 51$ dada por

$$S_t = \mu + \sum_{j=1}^6 \left[\alpha_j \cos \frac{(2\Pi jt)}{12} + \lambda_j \sin \frac{(2\Pi jt)}{12} \right]. \quad (19)$$

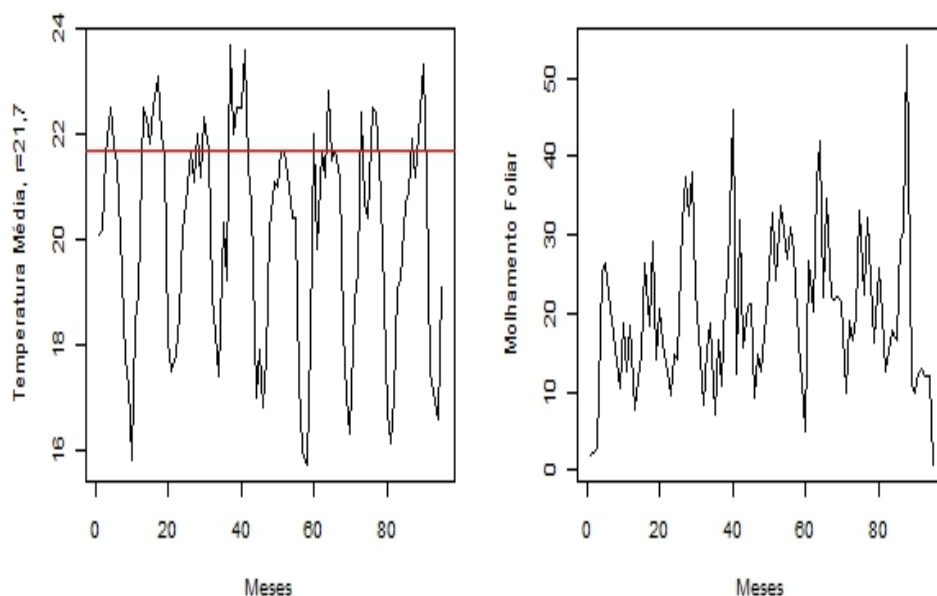


Figura 2 Representação gráfica da temperatura média mensal (em $^{\circ}C$) e respectivo ponto limiar ($r = 21,7^{\circ}C$), lado esquerdo, e do molhamento foliar (em %), no período de outubro de 1999 a agosto de 2007, na região de Varginha, MG

Os valores de μ , α_j e λ_j significativos, associados às suas componentes periódicas e aos seus respectivos erros padrão do modelo de regressão utilizado na estimação da sazonalidade (Equação (19)) são apresentados na Tabela 1.

A verificação da convergência foi feita pelo método de cadeias múltiplas, descrito por Gelman e Rubin (1992). A convergência é alcançada para $\sqrt{\hat{R}} \approx 1$. Para cada parâmetro, foram tomadas duas cadeias em paralelo com valores iniciais distintos. Foram consideradas 10000 iterações para cada cadeia sendo que as 1000 primeiras foram desprezadas e, das observações restantes, uma foi tomada a cada 2, de modo a obter uma amostra independente e identicamente distribuída. O retardo d foi tomado igual a 1 por ser também o que conduziu a um modelo

mais completo com mais variáveis após terem sido testados retardos de 2, 3, 4 e 5 tempos anteriores. Após a divisão feita nos dados pela variável *threshold*, o ajuste do modelo do primeiro regime foi feito com 67 observações e o ajuste do segundo com 27.

Tabela 1 Resultado da análise de regressão de mínimos quadrados para a sazonalidade presente na série de temperatura média.

Parâmetros	C P	E P	D P
μ	–	$2,0110 \times 10^1$	$8,6200 \times 10^{-2}$
α_1	$\cos \frac{2\pi t}{12}$	$-8,8020 \times 10^{-1}$	$1,2340 \times 10^{-1}$
λ_1	$\sin \frac{2\pi t}{12}$	2,3810	$1,2120 \times 10^{-1}$
α_2	$\cos \frac{4\pi t}{12}$	$9,3390 \times 10^{-1}$	$1,2310 \times 10^{-1}$
λ_5	$\sin \frac{10\pi t}{12}$	$-2,6810 \times 10^{-1}$	$1,2340 \times 10^{-1}$

Nota: CP-Componente Periódica; EP-Estimativa Paramétrica; DP-Desvio Padrão.

Na Tabela 2 são apresentadas as estimativas bayesianas média, mediana e moda dos parâmetros do modelo TARSO ajustado com o respectivo valor *threshold* utilizado, 0,411, correspondente estacionário de $r = 21,7^0C$. São também apresentados os erros padrão de cada parâmetro, que comprovam a significância dos mesmos, bem como o fator de convergência de Gelman e Rubin (1992), $\sqrt{\hat{R}}$, em torno de 1 indicando a convergência de cada cadeia de parâmetros. Nota-se que tanto a variável *threshold* temperatura média como a variável molhamento foliar se mostraram determinantes para a ferrugem, embora suas contribuições tenham sido relacionadas a valores diferentes de temperatura.

A equação do modelo ajustado é dada por

$$Y_t = \begin{cases} 0,3147Y_{t-1} - 6,35T_{t-1} + a_t^{(1)} & , \text{ se } T_{t-1} \leq 0,411 \\ 0,6631Y_{t-1} + 0,7412U_{t-1} + a_t^{(2)} & , \text{ se } T_{t-1} > 0,411 \end{cases} .$$

Observa-se que, para temperaturas menores que 0,411, o valor atual de ferrugem

é constituído por um acréscimo de 0,3147 do valor da ferrugem do período anterior e por um decréscimo de 6,35 da variável limiar temperatura média do mês passado. Para temperaturas superiores a 0,411, o índice de ferrugem é respondido pelo acréscimo de 0,6631 do valor da incidência do mês anterior e por 0,7412 do valor do molhamento foliar do mês anterior. A presença não simultânea de temperatura e molhamento foliar, nos regimes, talvez possa ser responsabilizada pela falta de coincidência de condições de temperatura e molhamento foliar, já que, quando essa coincidência existe, a ferrugem é desencadeada mais rapidamente.

Há que se considerar também que existem outras variáveis que atuam na ferrugem como: intensidade, duração e frequência de chuvas; orvalho; período de molhamento, dentre outras que não foram consideradas no ajuste. Mesmo com essa limitação, foi possível mostrar que, de fato, os fatores relacionados ao ambiente contribuem para a formação da ferrugem, ratificando os estudos já realizados. O ajuste obtido sugere cautela por parte do produtor para com sua lavoura. As epidemias ocorrem na presença de condições ambientais favoráveis.

Tabela 2 Estimativas bayesianas, erro padrão e fator de convergência, $\sqrt{\hat{R}}$, para cada parâmetro do modelo TARSO(2;1,1,0,1,0,1) ajustado com duas variáveis independentes.

Valor	Threshold (τ)	Parâmetros	Média	Mediana	Moda	Erro Padrão	$\sqrt{\hat{R}}$
		ϕ_{11}	0,3147	0,3160	0,3138	0,447	0,9990
		β_{11}	-6,3500	-6,3791	-6,6629	11,5530	0,9740
$r = 0,411$	(21, 7)	ϕ_{21}	0,6631	0,6633	0,6592	0,7034	1,0100
		θ_{21}	0,7412	0,7381	0,7631	0,8819	0,9990
		τ_1	0,0042	0,0041	0,0041	0,0029	0,9990
		τ_2	0,0059	0,0058	0,0055	0,0066	0,9990

4.1.2 Modelo TARSO com uma variável independente

Nesta segunda análise, as variáveis utilizadas foram dados atualizados de lapaf e temperatura média. O processo foi feito com 137 observações referentes ao período de outubro de 1999 a fevereiro de 2011. Assim como no ajuste da seção anterior, aqui também foram considerados vários valores de r no intervalo de 20^0C a 24^0C e retardos d variando de 1 a 5. No entanto, o melhor ajuste foi obtido considerando a defasagem $d = 1$ e o valor *threshold* $r = 22, 3^0C$ com correspondente estacionário igual a 0,9567. Esse valor também foi determinado subtraindo o valor da temperatura média em $t = 14$ (novembro de 2000) da sazonalidade estimada por meio da Equação (19) cujos coeficientes μ , α_j e λ_j significativos, associados às suas componentes periódicas e aos seus respectivos erros padrão estão apresentados na Tabela 3.

Tabela 3 Resultado da análise de regressão de mínimos quadrados para a sazonalidade presente na série de temperatura média.

Parâmetros	C P	E P	D P
μ	—	$2,0100 \times 10^1$	$7,1490 \times 10^{-2}$
α_1	$\cos \frac{2\Pi t}{12}$	$-7,9890 \times 10^{-1}$	$1,0140 \times 10^{-1}$
λ_1	$\sin \frac{2\Pi t}{12}$	2,4590	$1,0060 \times 10^{-1}$
α_2	$\cos \frac{4\Pi t}{12}$	$9,737 \times 10^{-1}$	$1,0220 \times 10^{-1}$

Nota: CP-Componente Periódica; EP-Estimativa Paramétrica; DP-Desvio Padrão.

A análise bayesiana seguiu da mesma forma, sendo também tomadas 10000 iterações em duas cadeias em paralelo para cada parâmetro. Foram eliminadas as 1000 primeiras; uma observação foi tomada a cada duas (2) de modo a obter uma amostra independente e identicamente distribuída. Após a divisão das observações, feita pela variável *threshold* temperatura média, ajustou-se um modelo para o primeiro regime com 116 observações e outro para o segundo com apenas 19

observações. Nota-se com isso que a maioria dos índices de temperatura média localizam-se abaixo de $22,3^{\circ}C$, o que condiciona grande parte da incidência da ferrugem pertencer ao primeiro regime.

Na Tabela 4 são apresentadas as estimativas bayesianas, dadas pela média, mediana e moda, dos parâmetros com $r = 0,9567$, o erro padrão, bem como o elemento do critério de Gelman e Rubin (1992) em torno de 1 confirmando a convergência de todas as estimativas de parâmetros.

O modelo ajustado pode ser melhor visualizado por sua equação

$$Y_t = \begin{cases} 0,4323Y_{t-1} - 4,3765T_{t-1} + a_t^{(1)} & , \text{ se } T_{t-1} \leq 0,9567 \\ 0,4927Y_{t-1} + a_t^{(2)} & , \text{ se } T_{t-1} > 0,9567 \end{cases} .$$

Observa-se neste caso que, quando a temperatura está abaixo de $0,9567$, a incidência atual da ferrugem é acrescida em $0,4323$ do seu valor passado e a temperatura média defasada ocasiona um decréscimo de $-4,3765$. Para temperatura acima de $0,9567$, a ferrugem é acrescida em $0,4927$ do seu valor passado. O maior número de observações, no regime 1, pode explicar em parte suas duas estimativas significativas.

A análise também foi realizada considerando $r = 17,1^{\circ}C$. Esse valor de temperatura ocorre em $t = 93$ (junho de 2007) com correspondente estacionário $r = 0,4327$. A obtenção desse valor foi feita subtraindo o valor da temperatura em $t = 93$, da sazonalidade estimada pela Equação (19) cujos parâmetros são os mesmos apresentados na Tabela 3.

Nesse ajuste foram utilizados os dados do período de outubro de 1999 a fevereiro de 2011 sendo que as estimativas bayesianas (média, mediana, moda), o erro padrão e o fator de convergência, $\sqrt{\hat{R}}$, para cada parâmetro, estão apresentados na Tabela 5.

Tabela 4 Estimativas bayesianas, erro padrão e fator de convergência, $\sqrt{\hat{R}}$, para cada parâmetro do modelo TARSO(2;1,1,1,0) com uma variável independente.

Valor <i>Threshold</i> (τ)	Parâmetros	Média	Mediana	Moda	Erro Padrão	$\sqrt{\hat{R}}$
	ϕ_{11}	0,4323	0,4319	0,4373	0,3357	1,02
	β_{11}	-4,3765	-4,3779	-4,3907	7,6866	0,99
$r = 0,9567$ (22,3)	ϕ_{21}	0,4927	0,4920	0,5024	0,7752	1,03
	τ_1	0,0039	0,0039	0,0039	0,002	0,98
	τ_2	0,0063	0,0061	0,0056	0,0083	1,01

Tabela 5 Estimativas bayesianas, erro padrão e fator de convergência, $\sqrt{\hat{R}}$, para cada parâmetro do modelo TARSO(2;1,0,1,0) com uma variável independente.

Valor <i>Threshold</i> (τ)	Parâmetros	Média	Mediana	Moda	Erro Padrão	$\sqrt{\hat{R}}$
$r = 0,4327$ (17, 1)	ϕ_{11}	0,3973	0,3972	0,3961	0,3597	0,9820
	ϕ_{21}	0,6431	0,6442	0,6436	0,6880	0,9900
	τ_1	0,0040	0,0040	0,0041	0,0022	0,9740
	τ_2	0,0047	0,0047	0,0047	0,0045	0,9900

A equação do modelo considerando $r = 17,1^{\circ}C$ é dada por

$$Y_t = \begin{cases} 0,3973Y_{t-1} + a_t^{(1)} & , \text{ se } T_{t-1} \leq 0,4327 \\ 0,6431Y_{t-1} + a_t^{(2)} & , \text{ se } T_{t-1} > 0,4327 \end{cases} .$$

Nota-se que a incidência da ferrugem do mês anterior contribui em 0,3973 para a formação da incidência atual em temperaturas menores que 0,4327 e com 0,6431 em temperaturas superiores a 0,4327. A variável *threshold* temperatura média não foi significativa em nenhum dos dois regimes. Este fato era esperado já que temperaturas inferiores a $18^{\circ}C$ são limitantes para a infecção, ou seja, não afetam a propagação da ferrugem. O mesmo ocorre para temperaturas superiores a $28^{\circ}C$, quando se espera que ela seja não significativa. Neste caso, porém, não foi possível comprovar a ausência da temperatura média na composição do modelo, pois o maior valor de temperatura média no conjunto de dados disponível é $23,7^{\circ}C$.

5 CONCLUSÕES

Foi possível comprovar a influência das variáveis climáticas, temperatura média e molhamento foliar, na formação dos índices de ferrugem. Verificou-se que, de fato, valores de temperatura na faixa de $20^{\circ}C$ a $24^{\circ}C$ são favoráveis ao progresso da ferrugem e que, para valores menores que $18^{\circ}C$, a ferrugem não sofre influência da variável temperatura média.

REFERÊNCIAS

BROEMELING, L. D. **Bayesian analysis of linear models**. New York: M. Dekker, 1985. 472 p.

CHALFOUN, S. M. **Doenças do cafeeiro**: importância, identificação e métodos de controle. Lavras: UFLA/FAEPE, 1998. 93 p.

FUNDAÇÃO PROCAFÉ. **Estação de avisos**: Sul de Minas. Disponível em: <<http://www.fundacaoprocafe.com.br> = *estacaoeavisos* = *suldeminas* = *boletimdeaviso*>. Acesso em: 23 maio 2011.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-472, Nov. 1992.

JAPIASSÚ, L. B. et al. Influência da carga pendente, do espaçamento e de fatores climáticos no desenvolvimento da ferrugem do cafeeiro. In: SIMPÓSIO DE PESQUISA DOS CAFÉS DO BRASIL, 5., 2007, Águas de Lindóia. **Anais...** Brasília: EMBRAPA, 2007. p. 5.

JEFFREYS, H. **Theory of probability**. 3rd ed. London: Oxford University, 1961. 470 p.

JIAN, W. B. et al. A nonlinear time series model for spring flow: an example from Shanxi Province, China. **Ground Water**, Dublin, v. 36, n. 1, p. 147-150, Jan. 1998.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, 2009. Disponível em: <<http://www.r-project.org>>. Acesso em: 1 fev. 2010.

SÁFADI, T.; MORETTIN, P. A. Análise Bayesiana do modelo Open Loop Threshold Autoregressive. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 62, n. 217, p. 91-105, 2001.

TONG, H. **Non-linear time series**: a dynamical system approach. Oxford: Oxford University, 1990. 564 p.

TONG, H. On a threshold model in pattern recognition and signal processing. In: ———. **Pattern recognition and signal processing**. Netherlands: Sijhoff & Noordhoff, 1978. p. 575-586. (NATO Asis Series E: Applied Science, 29).

———. **Threshold models in non-linear time series analysis**: lecture notes in statistics. New York: Springer-Verlag, 1983. 323 p.

TONG, H.; LIM, K. S. Threshold autoregression, limit cycles and cyclical data. **Journal of the Royal Statistical Society B**, London, v. 42, n. 3, p. 245-292, 1980.

TSAY, R. S. Testing and modeling multivariate threshold models. **Journal of the American Statistical Association**, New York, v. 93, n. 443, p. 1188-1202, 1998.

ANEXO

	Página
PROGRAMA A: Programa R de simulação de um modelo TARSO com duas variáveis independentes	80

PROGRAMA A: Programa R de simulação de um modelo TARSO com duas variáveis independentes

```

#Função para gerar a série independente $$Y$
simulaTARSO2=function(phi10,phi20,phi1,beta1,theta1,phi2,beta2,
theta2,r,d,n,dp1,dp2,Z,U){
  ##função para simular os dados de um tarso
  ## phi10 e phi20 – valor da constante do regime 1 e 2
  ## phi1 e phi2 – vetor contendo os parâmetros para a variável Y
  no regime 1 e 2
  ## beta1 e beta2 – vetor contendo os parâmetros para a variável U
  no regime 1 e 2,
  ##colocar 0 no primeiro valor caso não queira considerar Zt
  ## theta1 e theta2 – vetor contendo os parâmetros para a variável
  U no regime 1 e 2,
  ##colocar 0 no primeiro valor caso não queira considerar Ut
  ## r – valor do threshold
  ## d – valor da defasagem para considerar o threshold
  ## número de dados gerados
  ## dp1 e dp2 – erro padrão do erro dos regimes 1 e 2
  ## Z – série com valores de Z para simular os dados
  p1=length(phi1)
  m1=length(beta1)-1
  q1=length(theta1)-1
  p2=length(phi2)
  m2=length(beta2)-1
  q2=length(theta2)-1

  Y=rep(0,n)
  inicial=max(p1,m1,q1,p2,m2,q2)+1
  j=0
  i=inicial
  while(i<(n+1)){
    if(Z[i-d]<=r){
      Y[i]=phi10+sum(phi1*Y[(i-1):(i-p1)])+sum(beta1*Z[i:(i-m1)])+
      sum(theta1*U[i:(i-q1)])+rnorm(1,0,dp1)
    }
    if(Z[i-d]>r){
      Y[i]=phi20+sum(phi2*Y[(i-1):(i-p2)])+sum(beta2*Z[i:(i-m2)])+
      sum(theta2*U[i:(i-q2)])+rnorm(1,0,dp2)
    }
    if(i>(n/2) && j==0){
      Y[1:(inicial+1)]=Y[i:(i-inicial)]
      Z[1:(inicial+1)]=Z[i:(i-inicial)]
      U[1:(inicial+1)]=U[i:(i-inicial)]
      i=inicial
      j=1
    }
  }
}

```

```

    i=i+1
  }
  return(Y)
}

##Função para estimar os parâmetros do modelo Tarso##

TARSOBAY2=function(Y,Z,Z1,U,r,d,p1,m1,q1,p2,m2,q2,y11,y21,z11,z21,
u11,u21,INT){
## 1) Colocar na posição Y série estacionária dependente, Z série
estacionária independente, Z1 série estacionária independente
## 2) r valor do threshold
## 3) d – defasagem do threshold
## 4) p1 – número de defasagens da série dependente no primeiro
regime
## 5) m1 – número de defasagens da série independente no primeiro
regime
## 6) p2 – número de defasagens da série dependente no segundo
regime
## 7) m2 – número de defasagens da série independente no segundo
regime
## 8) y11 – se for incluir a constante da série dependente no
primeiro regime y11=1, caso contrário y11=0
## 9) y21 – se for incluir a constante da série dependente no
segundo regime y21=1, caso contrário y21=0
## 10) z11 – se for incluir a constante da série independente no
primeiro regime z11=1, caso contrário z11=0
## 11) z21 – se for incluir a constante da série independente no
segundo regime z21=1, caso contrário z21=0
## 12) INT – número de iterações no processo MCMC

n=length(Y)
h=max(d,p1,m1,q1,p2,m2,q2)
h=h+1
k1=0
k2=0
#Contar número de observações em cada regime
for(i in h:n){
  if(Z1[i-d]<=r) k1=k1+1
  if(Z1[i-d]>r) k2=k2+1
}
##Primeiro Regime##
Y1=matrix(0,k1,p1+1)
m11=m1
q11=q1
if(z11==1) m11=m11+1
if(u11==1) q11=q11+1
z1=matrix(0,k1,m11)

```

```

u1=matrix(0,k1,q11)

##Segundo Regime##
Y2=matrix(0,k2,p2+1)
m21=m2
q21=q2
if(z21==1) m21=m21+1
if(u21==1) q21=q21+1
z2=matrix(0,k2,m21)
u2=matrix(0,k2,q21)

##Guardar valores dos regimes##
k1=0
k2=0
for(i in h:n){
  ##Primeiro Regime
  if(Z1[i-d]<=r){
    k1=k1+1
    for(j in 1:(p1+1)){
      Y1[k1,j]=Y[i-j+1]
    }
    if(z11==1 && m11>0){
      for(j in 1:m11){
        z1[k1,j]=Z[i-j+1]
      }
    }
    if(z11!=1 && m11>0){
      for(j in 1:m11){
        z1[k1,j]=Z[i-j]
      }
    }
    if(u11==1 && q11>0){
      for(j in 1:q11){
        u1[k1,j]=U[i-j+1]
      }
    }
    if(u11!=1 && q11>0){
      for(j in 1:q11){
        u1[k1,j]=U[i-j]
      }
    }
  }
}
##Segundo Regime
if(Z1[i-d]>r){
  k2=k2+1
  for(j in 1:(p2+1)){
    Y2[k2,j]=Y[i-j+1]
  }
}

```



```

}
if (z21==1 && m21>0){
  for (j in 1:m21){
    z2[k2,j]=Z[i-j+1]
  }
}
if (z21!=1 && m21>0){
  for (j in 1:m21){
    z2[k2,j]=Z[i-j]
  }
}
if (u21==1 && q21>0){
  for (j in 1:q21){
    u2[k2,j]=U[i-j+1]
  }
}
if (u21!=1 && q21>0){
  for (j in 1:q21){
    u2[k2,j]=U[i-j]
  }
}
}
}
Y11=Y1[,2:(p1+1)]
Y21=Y2[,2:(p2+1)]
##Montando as matrizes A1 e B1
if (m11>0 && q11>0) X1=cbind(Y11,z1,u1) ##matriz de delineamento
do primeiro regime
if (m11>0 && q11==0) X1=cbind(Y11,z1)
if (m11==0 && q11>0) X1=cbind(Y11,u1)
if (m11==0 && q11==0) X1=as.matrix(Y11)
if (y11==1) X1=cbind(1,X1)
if (m21>0 && q21>0) X2=cbind(Y21,z2,u2) ##matriz de delineamento
do segundo regime
if (m21>0 && q21==0) X2=cbind(Y21,z2)
if (m21==0 && q21>0) X2=cbind(Y21,u2)
if (m21==0 && q21==0) X2=as.matrix(Y21)
if (y21==1) X2=cbind(1,X2)
r1=ncol(X1)
r2=ncol(X2)
B1=t(X1)%*%Y1[,1]
B2=t(X2)%*%Y2[,1]
A1=t(X1)%*%X1
A2=t(X2)%*%X2
int=INT
##Valores iniciais##
par1=matrix(0,int,r1)

```

```

par2=matrix(0,int,r2)
tau1=rep(0,int)
tau2=rep(0,int)
##calculo das posteriores##
par1[1,]=20
par2[1,]=20
tau1[1]=2
tau2[1]=2

require(MASS)

for(i in 2:int){

  ##posteriores dos gama's são normais N(mi, si)
  M1=solve(A1)%*%B1
  S1=solve(tau1[i-1]*A1)
  par1[i,]=mvrnorm(1,M1,S1)

  M2=solve(A2)%*%B2
  S2=solve(tau2[i-1]*A2)
  par2[i,]=mvrnorm(1,M2,S2)

  ##posterioris dos taus são gama(bi, ci)
  b1=(k1-(p1+m1+1))/2
  b2=(k2-(p2+m2+1))/2
  aux1=Y1[1]-par1[i-1,]%*%t(X1)
  c1=(aux1)%*%t(aux1)/2
  aux2=Y2[1]-par2[i-1,]%*%t(X2)
  c2=(aux2)%*%t(aux2)/2
  tau1[i]=rgamma(1,b1,c1)
  tau2[i]=rgamma(1,b2,c2)
}
PAR=cbind(par1,par2,tau1,tau2)
PHI1=c("phi10","phi11","phi12","phi13","phi14","phi15","phi16",
"phi17","phi18","phi19","phi110")
PHI2=c("phi20","phi21","phi22","phi23","phi24","phi25","phi26",
"phi27","phi28","phi29","phi210")
beta1=c("beta10","beta11","beta12","beta13","beta14","beta15",
"beta16","beta17","beta18","beta19","beta110")
beta2=c("beta20","beta21","beta22","beta23","beta24","beta25",
"beta26","beta27","beta28","beta29","beta210")
theta1=c("theta10","theta11","theta12","theta13","theta14",
"theta15","theta16","theta17","theta18","theta19","theta110")
theta2=c("theta20","theta21","theta22","theta23","theta24",
"theta25","theta26","theta27","theta28","theta29","theta210")

if(y11==1) PHI11=PHI1[1:(p1+1)]
if(y11!=1) PHI11=PHI1[2:(p1+1)]

```

```

if (y21==1) PHI21=PHI2[1:(p2+1)]
if (y21!=1) PHI21=PHI2[2:(p2+1)]
if (z11==1 && m11>0) beta11=beta1[1:(m1+1)]
if (z11!=1 && m11>0) beta11=beta1[2:(m1+1)]
if (z21==1 && m21>0) beta21=beta2[1:(m2+1)]
if (z21!=1 && m21>0) beta21=beta2[2:(m2+1)]
if (u11==1 && q11>0) theta11=theta1[1:(q1+1)]
if (u11!=1 && q11>0) theta11=theta1[2:(q1+1)]
if (u21==1 && q21>0) theta21=theta2[1:(q2+1)]
if (u21!=1 && q21>0) theta21=theta2[2:(q2+1)]
TAU=c("tau1","tau2")
par=PHI11
if (m11>0) par=c(par,beta11)
if (q11>0) par=c(par,theta11)
par=c(par,PHI21)
if (m21>0) par=c(par,beta21)
if (q21>0) par=c(par,theta21)
par=c(par,TAU)
CADEIAS=PAR
colnames(CADEIAS)=par
return(CADEIAS)
}#termina função

##Função para diagnóstico bayesiano
diagnostico=function(cadeias){
  require(coda)
  RafteryLewis=raftery.diag(cadeias, q=0.025, r=0.005, s=0.95,
converge.eps=0.001)
  Geweke=geweke.diag(cadeias, frac1=0.1, frac2=0.5)
  valorp=2*(1-pnorm(abs(Geweke$z)))
  Geweke=rbind(Geweke$z, valorp)
  rownames(Geweke)=c("Z-score","valorp")
  HeidelbergerWelch=heidel.diag(cadeias, eps=0.1, pvalue=0.05)
  n=nrow(cadeias)
  n1=n/2
  n2=n1+1
  x=cadeias[1:n1,]
  x1=cadeias[n2:n,]
  x=mcmc(x)
  x1=mcmc(x1)
  x3=mcmc.list(x1,x)
  GelmanRubin=gelman.diag(x3, confidence=0.95, transform=FALSE,
autoburnin=TRUE)
  resul=list(RafteryLewis=RafteryLewis,Geweke=Geweke,
HeidelbergerWelch=HeidelbergerWelch,GelmanRubin=GelmanRubin$psrf)
  return(resul)
}

```

```

##Função de saída dos parâmetros estimados e gráficos das cadeias
parametrosTARSO=function (CADEIAS, thin , burnin){
  require (boa)
  require (modest)
  nomes=colnames (CADEIAS)
  int=nrow (CADEIAS)
  n=(int-burnin)/thin
  n1=ncol (CADEIAS)
  par=matrix (0 , n , n1)
  j=burnin
  for (i in 1:n){
    for (k in 1:n1){
      par [ i , k]=CADEIAS [ j , k]
    }
    j=j+thin
  }

  CI=matrix (0 , n1 , 2)
  for (i in 1:n1){
    CI [ i , ]=boa . hpd ( par [ , i ] , 0.05)
  }
  parametros=CI

  ##Media ##
  MEDIA=matrix (0 , n1 , 1)
  for (i in 1:n1){
    MEDIA [ i , ]=mean ( par [ , i ])
  }
  parametros=cbind ( parametros , MEDIA)

  ##Mediana ##
  MEDIANA=matrix (0 , n1 , 1)
  for (i in 1:n1){
    MEDIANA [ i , ]=median ( par [ , i ])
  }
  parametros=cbind ( parametros , MEDIANA)

  require (modeest)
  ##Moda ##
  MODA=matrix (0 , n1 , 1)
  for (i in 1:n1){
    a=mlv ( par [ , i ])
    MODA [ i , ]=a$M
  }
  parametros=cbind ( parametros , MODA)

  rownames ( parametros)=nomes

```

```
colnames(parametros)=c("LI_inferior","LI_superior","media",
"mediana","moda")

CADEIASFINAL=par
colnames(CADEIASFINAL)=nomes
resul=list(parametros=parametros,CADEIASFINAL=CADEIASFINAL)
return(resul)
}

graficos=function(cadeias){
  nomes=colnames(cadeias)
  n=length(nomes)
  par(mfrow=c(2,n))
  for(i in 1:n){
    hist(cadeias[,i],main=nomes[i],xlab="")
  }
  for(i in 1:n){
    plot(cadeias[,i],main=nomes[i],xlab="interação",ylab="")
  }
}
```

CAPÍTULO 3

Sistema de alarme ótimo para processos TARSO

RESUMO

Os sistemas de alarme ótimos têm tido atualmente grande utilidade na detecção e advertência de ocorrência de catástrofes. Essa metodologia foi utilizada com o objetivo de construir um sistema de alarme ótimo para processos TARSO de ordem p sendo a ordem $p = 1$ utilizada para exemplificar seu funcionamento. Estudos de simulação foram feitos para criar mais um mecanismo de previsão para a doença ferrugem do café via modelo TARSO com ajuste bayesiano. Esse tipo de ajuste, por meio das probabilidades a posteriori, permitem que o sistema seja atualizado a cada ponto do tempo. A metodologia foi aplicada às séries de índice de ferrugem e à série de temperatura média sendo possível verificar que o alarme resultou em uma detecção de catástrofe alta em pontos nos quais a catástrofe ainda não ocorreu e em detecções baixas se o ponto já estava no estado de catástrofe.

Palavras-chave: Ferrugem. Inferência bayesiana. Modelo TARSO. Limiar.

ABSTRACT

Nowadays, alarm systems have been very useful in detecting and warning of catastrophe occurrence. This methodology was used with the aim of building an optimal alarm system for TARSO processes of order p , being the order $p = 1$ used to exemplify its operation. Simulation studies were done to create more a mechanism for forecasting of coffee rust disease, via TARSO model with Bayesian fit. This fit type through of a posteriori probability, allow the system to be updated each time point. The methodology was applied to rust rates and average temperature series. It was possible to verify that the alarm resulted in a detection of high catastrophe in points at which the catastrophe did not occur and in the low detections, if the point is already in a catastrophe state.

Keywords: Rust. Bayesian inference. TARSO Model. Threshold.

1 INTRODUÇÃO

Nos últimos anos, tem havido um interesse potencial em descobrir quando uma variedade de processos aleatórios cruzarão um certo nível crítico u . Na terminologia de Lindgren (1985), esse fato é chamado de "catástrofe" e é usado como terminologia mais atraente.

Exemplos de catástrofes para Monteiro, Pereira e Scotto (2008) e Svensson et al. (1996) abrangem inundações de planícies, falta de energia elétrica, itens faltosos em um estoque, predição do crescimento em média do nível do mar e frequências de inundações devidas ao aquecimento global decorrentes do efeito estufa, avaliação dos impactos na saúde em razão da poluição do ar, bem como diversas outras situações.

A maneira mais simples para construir um sistema de alarme, também chamado de preditor ótimo de catástrofe, é por meio da utilização de um preditor simples $\hat{Y}_{t+h} = E[Y_{t+h}|Y_s, -\infty < s \leq t]$, para $h > 0$, em que um alarme é dado toda vez que o preditor excede algum nível. Lindgren (1985), porém, argumenta que tais alarmes simples não funcionam bem se os princípios de predição são julgados não em quão \hat{Y} se aproxima de Y em média, mas na habilidade de detectar os eventos, localizá-los precisamente no tempo, e produzir tão poucos alarmes falsos quanto possível.

Dessa forma, levando em consideração tais ideias, Lindgren (1975, 1980, 1985) e Maré (1980) estabeleceram os princípios ótimos dos cruzamentos de níveis e correspondentemente de sistemas de alarme ótimos para processos aleatórios. A seguir são elencados alguns trabalhos desses autores bem como de outros.

Maré (1980) desenvolveu um sistema de alarme ótimo e aplicou o método a processos gaussianos. Para ele, um sistema de alarme tem uma alta habilidade de detectar catástrofes se a proporção de catástrofes detectadas dentre todas é próxima

de um. Se, ao mesmo tempo, a proporção de alarmes corretos dentre todos os alarmes é tão alta quanto possível, então o sistema de alarme é dito ser ótimo.

Lindgren (1985) utilizou um processo estacionário e gaussiano para prever o cruzamento futuro de um nível u investigando critérios para um preditor com bom nível de cruzamento. Em um exemplo foi mostrado que o alarme ótimo localizou o cruzamento superior mais precisamente e numa fase mais precoce que o alarme simples, que tem uma tendência a dar o alarme tardiamente.

Turkman e Turkman (1990) sugeriram uma versão preditiva bayesiana para processos autorregressivos de ordem 1, $AR(1)$, apesar de não terem feito nenhuma tentativa para solucionar o problema computacionalmente. O trabalho desses autores foi uma tentativa para superar a desvantagem dos sistemas de alarme introduzidos por Lindgren (1980) e Maré (1980) os quais ignoram as variações amostrais dos parâmetros do modelo, além das probabilidades a posteriori dos eventos não serem atualizadas a cada ponto do tempo.

Svensson et al. (1996) apresentaram um preditor ótimo de cruzamentos de níveis e catástrofes para processos autorregressivos e médias móveis e investigou o desempenho do preditor. Para eles, um preditor de catástrofe ótimo é o preditor que dá um número mínimo de alarmes falsos para uma probabilidade de detecção fixada. Utilizaram as características operacionais como uma ferramenta de avaliação para comparar e construir os preditores, isto é, usando a probabilidade de detecção de uma catástrofe para o preditor. Uma condição explícita para o preditor ótimo de catástrofe baseado na predição linear de valores futuros do processo é dado e comparado com um preditor simples de catástrofe, os quais alarmam quando os valores preditos do processo excedem um dado nível. Seus resultados indicaram que o preditor ótimo de catástrofe é melhor que o preditor simples. Svensson e Holst (1997) aplicaram essas ideias para desenvolver um sistema de

alarme ótimo para prever os níveis de altura das águas no mar Báltico.

Antunes, Amaral-Turkman e Turkman (2003) estenderam os resultados de Turkman e Turkman (1990) para processos autorregressivos de ordem p , $AR(p)$, mostrando como as características do alarme podem ser obtidas numericamente. Esses autores nomearam o sistema de alarme de Lindgren (1980) e Maré (1980) de previsão de eventos "off-line", em que o sistema é construído para eventos incondicionais, supostos terem a mesma probabilidade em todo tempo devido à suposição de estacionariedade. Em contrapartida estabeleceram os sistemas de alarme chamados previsão de eventos "on-line", para os quais as probabilidades a posteriori são atualizadas a cada ponto temporal.

Por exemplo, se o objetivo é construir um sistema de alarme para detecção do evento cruzamento superior $C_{t,j} = \{Y_{t+j-1} \leq u < Y_{t+j}\}$ para algum u , baseado nos resultados de variáveis aleatórias $\{Y_t, Y_{t-1}\}$, então as observações disponíveis até o tempo $t - 2$, $D_t = \{Y_{t-2}, Y_{t-3}, \dots\}$ devem ser utilizadas para atualizar a probabilidade do evento cruzamento superior. Neste caso, o sistema de alarme deve ser projetado para eventos condicionais $\{C_{t,j}|D_t\}$ em cada ponto de tempo t (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

Monteiro, Pereira e Scotto (2008) desenvolveram um sistema de alarme para prever se um processo de contagem ultrapassa superiormente um certo nível u sempre que esse fato é predito. Concluíram que a técnica leva a eventos preditores ótimos no sentido de dar o menor número de alarmes falsos para um determinado tamanho de alarme. Obtiveram um número alto de alarmes falsos e sugeriram, como solução futura para este problema, incluir no modelo informações externas adicionais, ou ainda, considerar o nível de catástrofe variando no tempo.

Costa, Scotto e Pereira (2010) desenvolveram um sistema de alarme para

predizer se uma série financeira temporal ajustada via modelo FIAPARCH cruza superiormente ou inferiormente algum nível particular u e dá um alarme sempre que o cruzamento é predito. Fizeram uma aplicação a um conjunto de dados referentes aos retornos diários da bolsa de valores de São Paulo. Concluíram que o sistema de predição "on-line" produz um número mínimo de alarmes falsos.

O objetivo, neste capítulo, será desenvolver um sistema de alarme ótimo para processos TARSO de ordem p e aplicar a metodologia a um TARSO de ordem 1. A importância desse método está na possibilidade de contribuir com mais um mecanismo de previsão para a doença que mais ocorre nas lavouras cafeeiras brasileiras, a ferrugem.

Para isto, na seção 2.1, é apresentada a metodologia do alarme sob a perspectiva bayesiana de acordo com Antunes, Amaral-Turkman e Turkman (2003). Em 2.2 é desenvolvido o sistema de alarme para processos TARSO de ordem p . Em 2.3 o alarme para um modelo TARSO de ordem $p = 1$ é especificado. Na seção 3.1, em resultados e discussão, é feito um estudo de simulação para comprovar o desempenho do alarme a partir de TARSO de ordem 1 e em 3.2 uma aplicação é realizada à série de índices de ferrugem na presença da variável climática temperatura média.

2 METODOLOGIA

Nesta seção será apresentado alguns conceitos básicos em eventos de predição sob a perspectiva bayesiana e o desenvolvimento do alarme para processos TARSO de ordem p e especificamente para a ordem $p = 1$.

2.1 Conceitos básicos em eventos de predição

Seja Y_t um processo estocástico discreto no espaço paramétrico $\theta \subset \mathbb{R}^k$ com k fixado. A sequência de tempo $\{t = p + 1, \dots, t - 1, t, t + 1, \dots\}$ é dividida em três seções: $\{p + 1, \dots, t - p\}$ - passado, $\{t - p + 1, \dots, t\}$ - presente e $\{t + 1, \dots\}$ - futuro. Essas seções, para algum $p > 0$, originam três conjuntos distintos:

- $D_t = \{Y_{p+1}, \dots, Y_{t-p}\}$ - dados ou experimento informativo;
- $\mathbf{Y}_2 = \{Y_{t-p+1}, \dots, Y_t\}$ - experimento presente;
- $\mathbf{Y}_3 = \{Y_{t+1}, \dots\}$ - experimento futuro.

Termos como catástrofe e região de alarme são comuns em eventos de predição e, por isso, são dados em forma de definição.

Definição 1 (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

O evento de interesse $C_{t,j}$, ou catástrofe, é qualquer evento na σ -álgebra gerada por \mathbf{Y}_3 . Um evento preditor ou alarme para o evento de interesse $C_{t,j}$ é qualquer evento na σ -álgebra gerada por \mathbf{Y}_2 .

$C_{t,j}$ é o evento cruzamento superior

$$C_{t,j} = \{Y_{t+j-1} \leq u < Y_{t+j}\},$$

para algum inteiro positivo j e para algum número real u .

Definição 2 (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

Um alarme é dado, em um instante t , para um evento futuro $C_{t,j}$, se o valor observado de \mathbf{Y}_2 pertence ao evento preditor. Se o evento de interesse ocorre quando um alarme é dado, então diz-se que o evento é detectado. Se o evento de interesse não ocorre quando um alarme é dado, então diz-se que um alarme falso foi dado.

As probabilidades preditivas para quaisquer eventos $C_{t,j}$ e $A_{t,j}$ gerados, respectivamente, por \mathbf{Y}_3 e \mathbf{Y}_2 são dadas por

$$\begin{aligned} P(A_{t,j}|D_t) &= \int_{\theta} P(A_{t,j}|\theta, D_t)p(\theta, D_t)d\theta \\ P(C_{t,j}|A_{t,j}, D_t) &= \frac{\int_{\theta} P(A_{t,j} \cap C_{t,j}|\theta, D_t)p(\theta|D_t)d\theta}{\int_{\theta} P(A_{t,j}|\theta, D_t)p(\theta|D_t)d\theta}. \end{aligned}$$

Definição 3 (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

Um evento $A_{t,j} \in \sigma_{\mathbf{Y}_2}$ é uma região de alarme ou evento preditor para o evento $C_{t,j}$ em \mathbf{Y}_3 se, sempre que $\mathbf{Y}_2 \in A_{t,j}$, é provável que o evento $C_{t,j}$ ocorrerá.

Definição 4 (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

A região de alarme $A_{t,j} \in \sigma_{\mathbf{Y}_2}$ tem tamanho $\alpha_{t,j}$ e probabilidade de detecção $\gamma_{t,j}$ se

$$\begin{aligned} \alpha_{t,j} &= P(A_{t,j}|D_t) \\ \gamma_{t,j} &= P(C_{t,j}|A_{t,j}, D_t). \end{aligned}$$

$\alpha_{t,j}$ pode ser visto como a proporção de tempo gasto no estado do alarme, ao passo que $\gamma_{t,j}$ é a probabilidade de prever corretamente um evento j - passos à frente.

Definição 5 (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

A região de alarme $A_{t,j}$ é ótima de tamanho $\alpha_{t,j}$ se

$$P(C_{t,j}|A_{t,j}, D_t) = \sup_{B \in \sigma_{Y_2}} P(C_{t,j}|B, D_t), \quad (1)$$

em que o supremo é tomado sobre todos os conjuntos $B \in \sigma_{Y_2}$ tal que $P(B|D_t) = \alpha_{t,j}$.

A região de alarme $A_{t,j}$ é ótima se, entre todas as regiões com o mesmo tamanho de alarme, ela tem a mais alta probabilidade de detecção. Ou ainda, a região de alarme é ótima se ela fornece o menor número de alarmes falsos para uma dada probabilidade de detecção (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

Definição 6 (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

Um sistema de alarme ótimo de tamanho $\alpha_{t,j}$ é uma família de regiões de alarme $A_{t,j}$ que satisfazem a Expressão (1).

Lema 1 (LINDGREN, 1985): Sejam $P(\mathbf{y}_2|D_t)$ e $P(\mathbf{y}_2|C_{t,j}, D_t)$ as densidades preditivas de \mathbf{Y}_2 e \mathbf{Y}_2 condicionado ao evento $C_{t,j}$, respectivamente. O sistema de alarme $A_{t,j}$ com

$$A_{t,j} = \{\mathbf{y}_2 \in \mathbb{R}^p : \frac{P(\mathbf{y}_2|C_{t,j}, D_t)}{P(\mathbf{y}_2|D_t)} \geq k_{t,j}\}$$

ou equivalentemente,

$$A_{t,j} = \{\mathbf{y}_2 \in \mathbb{R}^p : \frac{P(C_{t,j}|\mathbf{y}_2, D_t)}{P(C_{t,j}|D_t)} \geq k_{t,j}\}, \quad (2)$$

em que $k_{t,j}$ é tal que $P(\mathbf{Y}_2 \in A_{t,j}|D_t) = \alpha_{t,j}$, é ótima de tamanho $\alpha_{t,j}$.

As definições acima asseguram que, em qualquer tempo t , para um dado tamanho de alarme $\alpha_{t,j}$, $A_{t,j}$ fornece a mais alta probabilidade de detecção. A questão é como escolher o próprio tamanho do alarme. Uma escolha lógica é $\alpha_{t,j} = P(C_{t,j}|D_t)$, ou seja, a proporção de tempo gasto na região de alarme é igual à proporção de tempo na região de catástrofe $C_{t,j}$ (ANTUNES; AMARAL-TURKMAN; TURKMAN, 2003).

A região de alarme (2) pode ainda ser reescrita como

$$A_{t,j} = \{\mathbf{y}_2 \in \mathbb{R}^p : P(C_{t,j}|\mathbf{y}_2, D_t) \geq \bar{k}\} \quad (3)$$

sendo $\bar{k} = k_{t,j}P(C_{t,j}|D_t)$. Dessa forma, a região de alarme depende do valor de $k_{t,j}$. A seguir são listadas algumas sugestões para escolha de \bar{k} .

Antunes, Amaral-Turkman e Turkman (2003) disseram que o valor ótimo de \bar{k} , denominado aqui por \bar{k}^* , pode ser obtido fixando o tamanho do alarme $\alpha_{t,j} = P(C_{t,j}|D_t)$ ou, de maneira alternativa, calculando-se as características do alarme para diferentes valores de \bar{k} e escolhendo aqueles que resultam na melhor coleção de características operacionais.

Svensson et al. (1996) sugeriram que \bar{k} deveria ser escolhido de modo que a probabilidade de alarme correto e a probabilidade de detecção do evento fossem aproximadamente iguais, isto é, $P(C_{t,j}|A_{t,j}) \simeq P(A_{t,j}|C_{t,j})$.

Com a determinação da região de alarme (3) é possível calcular as características do sistema de alarme dadas por

1. Tamanho do alarme: $P(A_{t,j}|D_t)$.
2. Probabilidade de detecção do evento: $P(A_{t,j}|C_{t,j}, D_t)$.
3. Probabilidade de alarme correto: $P(C_{t,j}|A_{t,j}, D_t)$.
4. Probabilidade de alarme falso: $P(C_{t,j}^c|A_{t,j}, D_t) = 1 - P(C_{t,j}|A_{t,j}, D_t)$.

5. Probabilidade de não - detecção do evento:

$$P(A_{t,j}^c | C_{t,j}, D_t) = 1 - P(A_{t,j} | C_{t,j}, D_t)$$

2.2 Sistemas de alarme ótimos para processos TARSO com 2 regimes

Nesta seção será desenvolvida a metodologia dos sistemas de alarme para processos TARSO com 2 regimes, duas variáveis independentes e ordem p . Em seguida será desenvolvido o TARSO de ordem 1 para uma variável independente.

Considere o modelo TARSO condicionado em $p = \max\{p_1, p_2, m_1, m_2, q_1, q_2\}$,

$$Y_t = \begin{cases} \phi_{10} + \sum_{i=1}^{p_1} \phi_{1i} Y_{t-i} + \sum_{\nu=0}^{m_1} \beta_{1\nu} T_{t-\nu} + \sum_{u=0}^{q_1} \theta_{1u} U_{t-u} + a_t^{(1)} & , \text{ se } T_{t-d} \leq r \\ \phi_{20} + \sum_{i=1}^{p_2} \phi_{2i} Y_{t-i} + \sum_{\nu=0}^{m_2} \beta_{2\nu} T_{t-\nu} + \sum_{u=0}^{q_2} \theta_{2u} U_{t-u} + a_t^{(2)} & , \text{ se } T_{t-d} > r, \end{cases}$$

e suas características descritas na seção (2) do capítulo 2. A densidade preditiva desse modelo foi desenvolvida na subseção (2.2.3) do mesmo capítulo sendo dada por

$$P(y_{n+h} | \mathbf{y}, d, r) \propto \frac{1}{2} \left[\sum_{t=p+1}^{n+h} Y_t^2 - \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right)' \left(\sum_{t=p+1}^{n+h} X_{it} X_{it}' \right)^{-1} \left(\sum_{t=p+1}^{n+h} X_{it} Y_t \right) \right]^{-\frac{n_i+h+1}{2}}.$$

Foi demonstrado que $P(y_{n+h} | \mathbf{y}, d, r)$ segue uma distribuição t-Student não-central com $\nu = n_i + h$ graus de liberdade, média $E^{(h)}$ e variância $V^{(h)}$. Ou seja,

$$P(y_{n+h} | \mathbf{y}, d, r) \sim St(\nu, E^{(h)}, V^{(h)}),$$

em que

$$E^{(h)} = \frac{(\mathbf{B}_i^{(n+h-1)})' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}}{(1 - X_{in+h}' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h})},$$

$$V^{(h)} = \left\{ \frac{C^{(n+h-1)} - (\mathbf{B}_i^{(n+h-1)})' (\mathbf{A}_i^{(n+h)})^{-1} (\mathbf{B}_i^{(n+h-1)})}{1 - X_{in+h}' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}} - \left(\frac{(\mathbf{B}_i^{(n+h)})' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}}{1 - X_{in+h}' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}} \right)^2 \right\} \frac{1}{(n_i + h)}$$

e n_i é o número de elementos no regime i com

$$\mathbf{X}_{in+h} = (1, Y_{n+h-1}, \dots, Y_{n+h-p_i}, T_{n+h}, T_{n+h-1}, \dots, T_{n+h-m_i}, U_{n+h}, U_{n+h-1}, \dots, U_{n+h-q_i})'$$

e $\mathbf{A}_i^{(m)}$ e $\mathbf{B}_i^{(m)}$ matrizes desenvolvidas na seção (2) do capítulo 2.

Com a preditiva estabelecida, todas as probabilidades relacionadas ao sistema de alarme podem ser computadas. A região de alarme agora é da forma

$$A_{t,j} = \{y_{t-p+1}, \dots, y_t, \in \mathbb{R}^p : P(C_{t,j} | y_{t-p+1}, \dots, y_t, D_t) > \bar{k}\},$$

em que $\bar{k} = k_{t,j} P(C_{t,j} | D_t)$, isto é, ela é estimada sobre um grid em \mathbb{R}^p .

Com a finalidade de exemplificar o mecanismo do alarme para o processo TARSO, tomou-se um modelo de ordem $p = 1$ com uma variável independente Z_t que é também a variável *threshold*. O modelo encontra-se descrito a seguir.

2.3 Sistema de alarme ótimo para o TARSO(2;1,1,1,1) com 2 regimes

Seja $\{Y_t, t = 1, \dots, n\}$ o processo TARSO (p_1, p_2, m_1, m_2) com dois regimes condicionado em $p = \max(p_1, p_2, m_1, m_2) = 1$. As ordens individuais são $p_1 = 1, p_2 = 1, m_1 = 1, m_2 = 1$ de modo que a equação do modelo é dada por

$$Y_t = \begin{cases} \phi_{11}Y_{t-1} + \beta_{11}Z_{t-1} + a_t^{(1)} & , \text{ se } Z_{t-d} \leq r \\ \phi_{21}Y_{t-1} + \beta_{21}Z_{t-1} + a_t^{(2)} & , \text{ se } Z_{t-d} > r, \end{cases} \quad (4)$$

sendo $a_t^{(i)} \sim N(0, \tau_i^{-1})$, $\gamma_i = (\phi_{i1}, \beta_{i1})'$ e τ_i os parâmetros com $i = 1, 2$. Fazendo $X_{it} = (Y_{t-1}, Z_{t-1})'$, o Modelo (4) pode ser reescrito como

$$Y_t = \begin{cases} \gamma_1' X_{1t} + a_t^{(1)} & , \text{ se } Z_{t-d} \leq r; \\ \gamma_2' X_{2t} + a_t^{(2)} & , \text{ se } Z_{t-d} > r. \end{cases}$$

Tomando $j = 2$ em $C_{t,j} = \{Y_{t+j-1} \leq u < Y_{t+j}\}$, o evento de interesse a ser predito - a catástrofe - toma a forma

$$C_{t,2} = \{Y_{t+1} \leq u < Y_{t+2}\}$$

e a região de alarme $A_{t,2}$ desse evento torna-se

$$A_{t,2} = \{y_t : P(C_{t,2}|y_t, D_t) > k_{t,2}P(C_{t,2}|D_t)\},$$

ou de forma equivalente

$$A_{t,2} = \{y_t : P(C_{t,2}|y_t, D_t) > \bar{k}\}, \quad (5)$$

em que $\bar{k} = k_{t,2}P(C_{t,2}|D_t)$.

Pela Expressão (5), o primeiro passo na construção do sistema de alarme consiste em calcular a probabilidade $P(C_{t,2}|y_t, D_t)$, probabilidade de catástrofe condicional em D_t e y_t e $P(C_{t,2}|D_t)$, probabilidade de catástrofe condicional em D_t . Essas probabilidades são dadas respectivamente, por

$$\begin{aligned}
P(C_{t,2}|y_t, D_t) &= P(Y_{t+1} \leq u < Y_{t+2}|y_2, \dots, y_{t-1}, y_t) \\
&= \int_{-\infty}^u \int_u^{\infty} P(y_{t+2}, y_{t+1}|y_t, D_t) dy_{t+2} dy_{t+1} \\
&= \int_{-\infty}^u \left[\int_u^{\infty} P(y_{t+2}|y_{t+1}, y_t, D_t) dy_{t+2} \right] P(y_{t+1}|y_t, D_t) dy_{t+1} \\
&= \int_{-\infty}^{u_1} \left[\int_{u_2}^{\infty} P(w_{t+2}|w_{t+1}, w_t, D_t) dw_{t+2} \right] P(w_{t+1}|w_t, D_t) dw_{t+1} \\
&= \int_{-\infty}^{u_1} P(w_{t+1}|w_t, D_t) [1 - F_\nu(u_2)] dw_{t+1} \quad (6)
\end{aligned}$$

e

$$\begin{aligned}
P(C_{t,2}|D_t) &= P(Y_{t+1} \leq u < Y_{t+2}|y_2, \dots, y_{t-1}) \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^u \int_u^{\infty} P(y_{t+2}, y_{t+1}|D_t) P(y_t|D_t) dy_{t+2} dy_{t+1} dy_t \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^u \left[\int_u^{\infty} P(y_{t+2}|y_{t+1}, D_t) dy_{t+2} \right] P(y_{t+1}|D_t) dy_{t+1} P(y_t|D_t) dy_t \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{u_1} \left[\int_{u_2}^{\infty} P(w_{t+2}|w_{t+1}, D_t) dw_{t+2} \right] P(w_{t+1}|D_t) dw_{t+1} P(w_t|D_t) dw_t \\
&= \int_{-\infty}^{\infty} P(w_t|D_t) dw_t \int_{-\infty}^{u_1} P(w_{t+1}|D_t) [1 - F_\nu(u_2)] dw_{t+1}. \quad (7)
\end{aligned}$$

As variáveis w_t , w_{t+1} e w_{t+2} são variáveis transformadas obtidas no capítulo 2, seção (2.2.3) Equação (16), F_ν é a função de distribuição com $\nu = n_i + 2$ graus de liberdade.

As probabilidades que aparecem nas Expressões (6) e (7) são distribuições

preditivas e para qualquer h , a distribuição preditiva de y_{n+h} dado \mathbf{y}, d e r foi desenvolvida na seção (2.2.3) do capítulo 2 e é uma distribuição t-Student não-central com $\nu = n_i + h$ graus de liberdade, média $E^{(h)}$ e variância $V^{(h)}$. Ou seja,

$$P(y_{n+h}|\mathbf{y}, d, r) \sim St(\nu, E^{(h)}, V^{(h)}).$$

Na notação de matrizes,

$$E^{(h)} = \frac{(\mathbf{B}_i^{(n+h-1)})' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}}{(1 - X_{in+h}' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h})} \quad (8)$$

e

$$V^{(h)} = \left\{ \frac{C^{(n+h-1)} - (\mathbf{B}_i^{(n+h-1)})' (\mathbf{A}_i^{(n+h)})^{-1} (\mathbf{B}_i^{(n+h-1)})}{1 - X_{in+h}' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}} - \left(\frac{(\mathbf{B}_i^{(n+h)})' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}}{1 - X_{in+h}' (\mathbf{A}_i^{(n+h)})^{-1} X_{in+h}} \right)^2 \right\} \frac{1}{(n_i + h)}. \quad (9)$$

Dessa forma, fazendo $h = 1$ e $h = 2$ nas Expressões (8) e (9) tem-se

$$u_1 = [V^{(1)}]^{-\frac{1}{2}}(u - E^{(1)})$$

e

$$u_2 = [V^{(2)}]^{-\frac{1}{2}}(u - E^{(2)}).$$

Vale lembrar que no desenvolvimento preditivo do Modelo (4) tem-se

$$\mathbf{X}_{in+h} = \begin{pmatrix} Y_{n+h-1} \\ Z_{n+h-1} \end{pmatrix}, \mathbf{B}_i^{(m)} = \begin{pmatrix} \sum_{t=p+1}^m Y_{t-1} Y_t \\ \sum_{t=p+1}^m Z_{t-1} Y_t \end{pmatrix}, C^{(m)} = \left(\sum_{t=p+1}^m Y_t^2 \right)$$

e

$$\mathbf{A}_i^{(m)} = \begin{pmatrix} \sum_{t=p+1}^m Y_{t-1}^2 & \sum_{t=p+1}^m Y_{t-1} Z_{t-1} \\ \sum_{t=p+1}^m Z_{t-1} Y_{t-1} & \sum_{t=p+1}^m Z_{t-1} Z_{t-1} \end{pmatrix},$$

sendo o número m de elementos somados nessas expressões substituídos pelas somas que aparecem em $E^{(h)}$ e $V^{(h)}$, expressões (8) e (9) e que em $P(C_{t,2}|D_t)$ as distribuições são condicionadas somente ao passado e $P(C_{t,2}|D_t)$ é fixa para qualquer t . As expressões $E^{(h)}$ e $V^{(h)}$, neste caso, são dadas por

$$E^{(h)} = \frac{(\mathbf{B}_i^{(n+h-2)})' (\mathbf{A}_i^{(n+h-1)})^{-1} X_{in+h}}{(1 - X'_{in+h} (\mathbf{A}_i^{(n+h-1)})^{-1} X_{in+h})}$$

e

$$V^{(h)} = \left\{ \frac{C^{(n+h-2)} - (\mathbf{B}_i^{(n+h-2)})' (\mathbf{A}_i^{(n+h-1)})^{-1} (\mathbf{B}_i^{(n+h-2)})}{1 - X'_{in+h} (\mathbf{A}_i^{(n+h-1)})^{-1} X_{in+h}} - \left(\frac{(\mathbf{B}_i^{(n+h-1)})' (\mathbf{A}_i^{(n+h-1)})^{-1} X_{in+h}}{1 - X'_{in+h} (\mathbf{A}_i^{(n+h-1)})^{-1} X_{in+h}} \right)^2 \right\} \frac{1}{(n_i + h)}.$$

Com o cálculo das integrais (6) e (7) é possível determinar a região de alarme (5) e calcular as características do sistema de alarme, ou seja, tamanho do alarme, probabilidade de detecção do evento, probabilidade de alarme correto, probabilidade de alarme falso e probabilidade de não-deteção do evento. Segue o procedimento sugerido por Antunes, Amaral-Turkman e Turkman (2003) para determinação das características operacionais do alarme:

1. Forme um grid $G = \{y_t \in \mathbb{R}\}$ de maneira que

$$P[(Y_t) \in G|D_t] \simeq 1;$$

2. para cada $(y_t) \in G$, calcule a probabilidade

$$P(C_{t,2}|y_t, D_t) = \int_{-\infty}^{u_1} P(w_{t+1}|w_t, D_t)[1 - F_\nu(u_2)]dw_{t+1};$$

3. para vários valores adequadamente escolhidos de \bar{k} , forme os conjuntos

$$A_{t,2} = \{y_t \in G : \int_{-\infty}^{u_1} P(w_{t+1}|w_t, D_t)[1 - F_\nu(u_2)]dw_{t+1} \geq \bar{k}\},$$

em que $\bar{k} = k_{t,2}P(C_{t,2}|D_t)$;

4. para cada \bar{k} , calcule o tamanho $\alpha_{t,\bar{k}}$ de $A_{t,2}$ por

$$\begin{aligned} \alpha_{t,\bar{k}} &= P(Y_t \in A_{t,2}|D_t) \\ &= \int_{A_{t,2}} P(y_t|D_t)dy_t. \end{aligned}$$

As características operacionais podem então ser calculadas para cada valor de \bar{k} , ou simplesmente para o valor ótimo de \bar{k} , \bar{k}^* , e são dadas por

1. Probabilidade de alarme correto

$$P(C_{t,2}|A_{t,2}, D_t) = \frac{P(C_{t,2} \cap A_{t,2}|D_t)}{P(A_{t,2}|D_t)},$$

em que

$$\begin{aligned} P(C_{t,2} \cap A_{t,2}|D_t) &= P((Y_{t+1} \leq u < Y_{t+2}) \cap Y_t \in A_{t,2}|D_t) \\ &= \int_{A_{t,2}} \int_{-\infty}^u \int_u^\infty P(y_{t+2}, y_{t+1}|y_t, D_t)P(y_t|D_t)dy_{t+2}dy_{t+1}dy_t. \end{aligned}$$

Assim,

$$P(C_{t,2}|A_{t,2}, D_t) = \frac{\int_{A_{t,2}} \int_{-\infty}^u \int_u^{\infty} P(y_{t+2}, y_{t+1}|y_t, D_t)P(y_t|D_t)dy_{t+2}dy_{t+1}dy_t}{\int_{A_{t,2}} P(y_t|D_t)dy_t}.$$

2. Probabilidade de detecção do evento

$$P(A_{t,2}|C_{t,2}, D_t) = \frac{P(A_{t,2} \cap C_{t,2}|D_t)}{P(C_{t,2}|D_t)} = \frac{\int_{A_{t,2}} \int_{-\infty}^u \int_u^{\infty} P(y_{t+2}, y_{t+1}|y_t, D_t)P(y_t|D_t)dy_{t+2}dy_{t+1}dy_t}{\int_{-\infty}^{\infty} \int_{-\infty}^u \int_u^{\infty} P(y_{t+2}, y_{t+1}|D_t)P(y_t|D_t)dy_{t+2}dy_{t+1}dy_t}.$$

3. Probabilidade de alarme falso

$$P(C_{t,2}^c|A_{t,2}, D_t) = 1 - P(C_{t,2}|A_{t,2}, D_t).$$

4. Probabilidade de evento não-detectado

$$P(A_{t,2}^c|C_{t,2}, D_t) = 1 - P(A_{t,2}|C_{t,2}, D_t).$$

Como não existe um valor exato para \bar{k} , ele foi tomado seguindo as sugestões de Antunes, Amaral-Turkman e Turkman (2003), isto é, escolhido o que resultou em uma melhor coleção de características operacionais, ou seja, maior probabilidade de detecção do evento e maior probabilidade de alarme correto e o que fez com que $P(C_{t,2}|D_t) \cong P(A_{t,2}|D_t)$.

3 RESULTADOS E DISCUSSÃO

Nesta seção será feito um estudo de simulação para ilustrar o desempenho do sistema de alarme de um conjunto de dados gerado por um processo TARSO de ordem 1, isto é, TARSO(2;1,1,1,1). Um sistema de alarme aplicado a dados reais de ferrugem será realizado a partir de um TARSO(2;1,0,1,1).

3.1 Estudo de simulação

Foram simuladas 5000 observações cujos valores foram tomados o mais próximo possível do observado na prática para dados de ferrugem do café. Na Tabela 1 são apresentados os valores reais e simulados para cada parâmetro do modelo TARSO(2;1,1,1,1) especificado. Também foi simulada a variável *threshold* Z sazonal, similar à variável *threshold* temperatura média. O valor *threshold* r foi de $21,5^{\circ}C$ e a defasagem d foi igual a 1.

Tabela 1 Valores reais e simulados dos parâmetros do modelo TARSO(2;1,1,1,1).

Parâmetros	Valor Real	Valor Simulado
ϕ_{11}	0,7200	0,7150
β_{11}	0,5200	0,5100
ϕ_{21}	0,9800	1,0330
β_{21}	0,3100	0,3100
τ_1	0,0200	0,0250
τ_2	0,0500	0,0510

Na Figura 1 é apresentada a trajetória de uma amostra simulada do modelo TARSO(2;1,1,1,1).

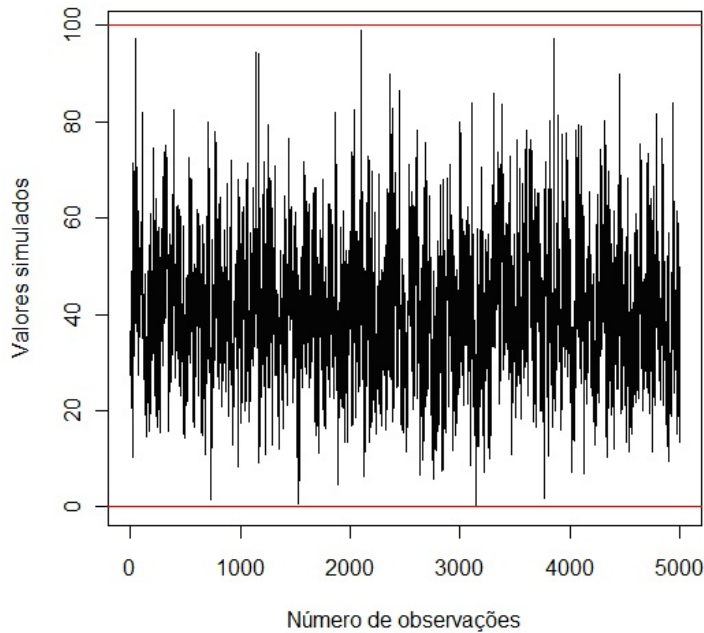


Figura 1 Valores simulados em % de um processo TARSO(2;1,1,1,1)

O evento de interesse é a catástrofe definida pelo cruzamento superior de um nível fixo u em $t + 2$ dado por

$$C_{t,2} = \{Y_{t+1} \leq u < Y_{t+2}\}. \quad (10)$$

Os valores utilizados do nível de cruzamento superior u foram 5%, 10%, 15% e 20%. Os tempos de análise t foram escolhidos de modo a abranger um número menor e um número maior de informações passadas. Na escolha desses pontos também foram levados em consideração pontos onde a Expressão (10) fosse satisfeita e um ponto onde ela não fosse satisfatória.

Como descrito na seção de construção do alarme, primeiramente foi ob-

tida a região de alarme construindo um grid variando de 0% a 100%, isto é, $(0\%, 100\%) \in \mathbb{R}$ para y_t de modo que

$$P[(Y_t) \in G|D_t] \approx 1.$$

Em cada ponto do grid foi calculada $P(C_{t,2}|y_t, D_t)$, isto é, a probabilidade de ocorrência do evento $C_{t,2} = \{Y_{t+1} \leq u < Y_{t+2}\}$ dado o valor presente do processo e seus valores passados. Para ilustrar, na Figura 2, é apresentado o gráfico dessas probabilidades em $t = 4752$ no qual pode-se notar um valor máximo de probabilidade igual a 0,5230.

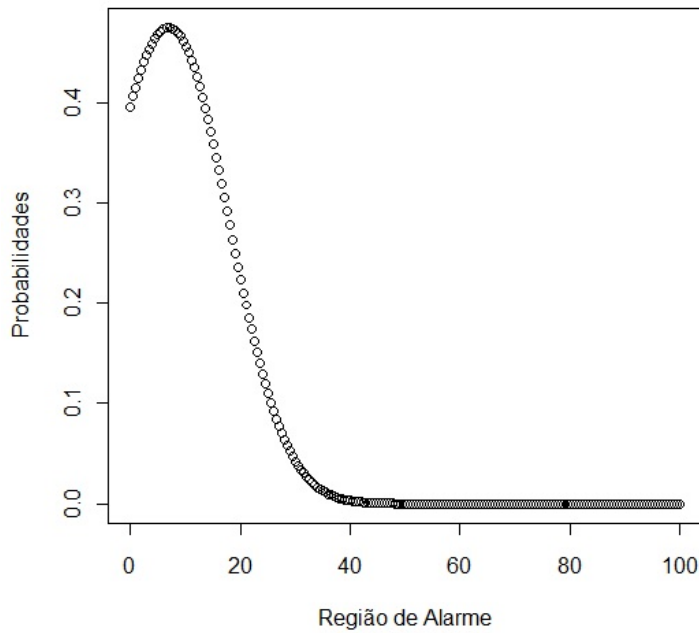


Figura 2 Valores das probabilidades $(y_t \in G)$ calculadas pela Expressão (6) em $t=4752$

O próximo passo na análise do sistema de alarme consiste em determinar

as regiões por meio da equação

$$A_{t,2} = \left\{ y_t : \frac{P(C_{t,2}|y_t, D_t)}{P(C_{t,2}|D_t)} > k_{t,2} \right\},$$

a qual indica que um valor y_t pertencerá à região se a razão $\frac{P(C_{t,2}|y_t, D_t)}{P(C_{t,2}|D_t)}$ superar $k_{t,2}$ variando de $P(C_{t,2}|D_t)$ a $P(C_{t,2}|D_t) + n \times 0,01$, com $n \in \mathbb{R}^+$. Na Figura 3 são mostradas as regiões de alarme também para o ponto $t = 4752$ com seus respectivos limites inferiores e superiores.

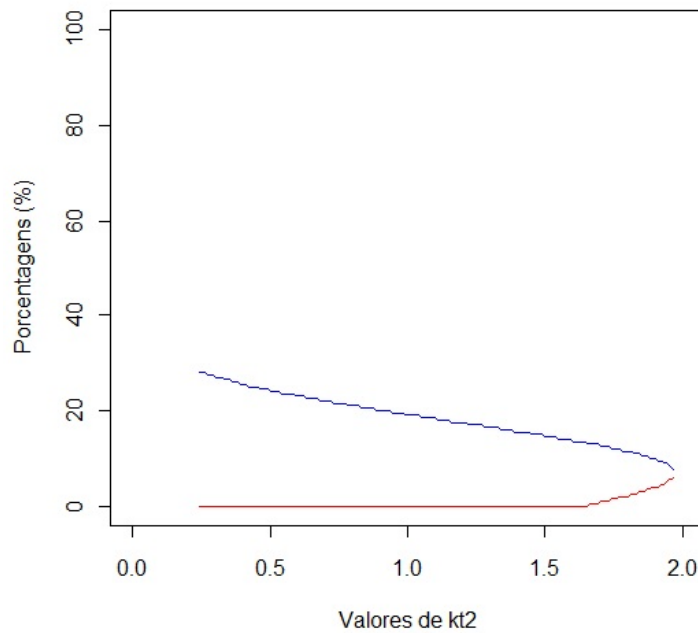


Figura 3 Limites inferiores e superiores das regiões de alarme para as observações do grid em $t=4752$

Na Figura 3 tem-se, por exemplo, que para um $k_{t,2}P(C_{t,2}|D_t) = \bar{k} = 1,3901$ o limite inferior da região de alarme é 0 e o limite superior é 15,5%

indicando que, para este valor de \bar{k} , o evento preditor é constituído pelos valores do grid variando de 0 a 15,5%. O tamanho do alarme, neste caso, é dado por $P(A_{t,2}|D_t) = \alpha_{t,2} = \int_0^{15,5} P(y_t|D_t)dy_t = 0,2326$. Esse tamanho é o que mais se aproxima de $P(C_{t,2}|D_t) = 0,2401$. Dessa forma, as características operacionais do alarme, probabilidade de alarme correto, $P(C_{t,2}|A_{t,2}, D_t)$, probabilidade de detecção do evento, $P(A_{t,2}|C_{t,2}, D_t)$, probabilidade de alarme falso, $P(C_{t,2}^c|A_{t,2}, D_t)$, e probabilidade de não-deteccção do evento, $P(A_{t,2}^c|C_{t,2}, D_t)$ foram calculadas para este tamanho. \bar{k} , neste caso, é \bar{k}^* , o valor ótimo de \bar{k} . Esse é o critério sugerido por Antunes, Amaral-Turkman e Turkman (2003).

Nas Tabelas 2 e 3 são apresentadas as características operacionais do alarme para $t=4752$ e para mais alguns pontos da série simulada. Os níveis de cruzamento utilizados em cada ponto, os valores de \bar{k}^* , e as regiões de alarme ótimas a partir das quais serão calculadas as características operacionais, são dados pela Tabela 2. Na Tabela 3 são apresentadas, além das características operacionais do processo TARSO considerado, a probabilidade de ocorrência da catástrofe dado D_t , $P(C_{t,2}|D_t)$. Essa probabilidade é útil para comparar com o tamanho do alarme, $P(A_{t,2}|D_t)$. Assim $P(C_{t,2}|D_t)$ deve ser o mais próximo possível de $P(A_{t,2}|D_t)$.

Tabela 2 Nível de cruzamento, \bar{k}^* e região de alarme dos pontos analisados no sistema de alarme.

t	u	\bar{k}^*	RA
738	10%	1,9246	[4,5;9]
4535	15%	0,9387	[0;19,5]
4752	20%	1,3901	[0;15,5]
4892	5%	1,4725	[0,5;12,5]

Nota: RA-Região de Alarme.

Os resultados apresentados na Tabela 3 mostram que $P(C_{t,2}|D_t) \approx P(A_{t,2}|D_t)$ para todos os pontos considerados. No ponto $t = 4752$, utilizando um nível de cruzamento superior $u = 20\%$, a probabilidade de alarme correto foi de 52,30% e a probabilidade de detecção do evento foi de 50,67%. As probabilidades de alarme falso e probabilidade de não-deteção do evento foram respectivamente iguais a 47,70% e 49,33%. Em $t = 738$ com um nível de cruzamento $u = 10\%$, a probabilidade de alarme correto ficou em torno de 46,68% e a probabilidade de detecção do evento foi de 42,22%. A probabilidade de alarme falso foi de 53,32% e a probabilidade de não-deteção do evento foi 57,78%. No ponto $t = 4892$ com um nível $u = 5\%$, a probabilidade de alarme correto foi de 41,22%, a probabilidade de detecção do evento foi 40,29%, a probabilidade de alarme falso foi 58,78% e a de não-deteção do evento, 59,71%. Em $t = 4535$, para um nível $u = 15\%$ o alarme teve seu pior desempenho. As probabilidades de alarme correto e de detecção do evento ficaram em torno de 0,08% ao passo que para alarme falso e não-deteção do evento foram aproximadamente 99,92%.

O melhor e o pior comportamento das características operacionais, respectivamente, nos pontos $t = 4752$ e $t = 4535$, pode ser justificado analisando os valores do processo no ponto e nos dois valores futuros que compõem o evento catástrofe. Na Tabela 4 são apresentados os valores do processo em $t, t + 1, t + 2$. Nota-se que em $t = 4752$, a expressão

$$C_{t,2} = \{Y_{t+1} \leq u < Y_{t+2}\}$$

é satisfeita para o nível de cruzamento utilizado. O mesmo não ocorre em $t = 4535$, para o nível utilizado $u = 15\%$. Neste caso, o processo já se encontra no estado de catástrofe.

Tabela 3 Características operacionais do alarme em diferentes pontos para o modelo TARSO(2;1,1,1,1).

t	u	$P(C_{t,2} D_t)$	$P(A_{t,2} D_t)$	$P(C_{t,2} A_{t,2}, D_t)$	$P(A_{t,2} C_{t,2}, D_t)$	$P(C_{t,2}^c A_{t,2}, D_t)$	$P(A_{t,2}^c C_{t,2}, D_t)$
738	10%	0,0483	0,0437	0,4668	0,4222	0,5332	0,5778
4535	15%	0,0477	0,0502	0,0008	0,0008	0,9992	0,9992
4752	20%	0,2401	0,2326	0,5230	0,5067	0,4770	0,4933
4892	5%	0,1805	0,1764	0,4122	0,4029	0,5878	0,5971

Tabela 4 Valores simulados do processo TARSO(2;1,1,1,1) em $t, t + 1, t + 2$.

t	Y_t	Y_{t+1}	Y_{t+2}
4535	27,3010	35,6444	40,5846
4752	20,1642	17,2734	46,5555

3.2 Aplicação aos dados de ferrugem do café

Nesta seção serão apresentados os resultados do sistema de alarme do modelo TARSO(2; 1, 0, 1, 1) para a série de incidência da ferrugem do cafeeiro em lavoura adensada com produção alta (lapaf) e variável *threshold* temperatura média.

Na Figura 4 são apresentados os gráficos das séries lapaf e temperatura média com o valor *threshold* $r = 21,5^{\circ}C$. Ambas as séries são sazonais de periodicidade igual a 12 meses. O valor r de temperatura tomado pertence à faixa considerada ótima no alastramento da doença ferrugem.

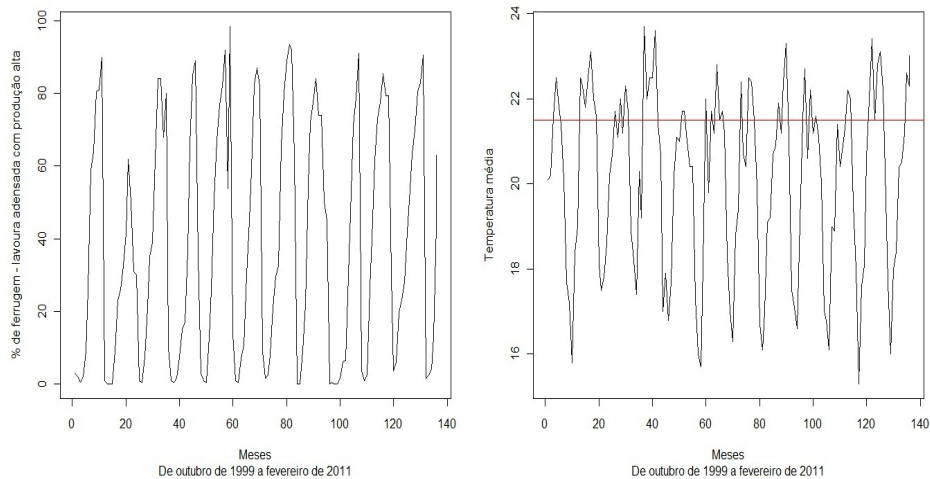


Figura 4 Índices de ferrugem e temperatura média com valor *threshold* $r = 21,5^{\circ}C$

As estimativas bayesianas, tomadas como a média da distribuição a posteriori, bem como os limites inferior e superior, respectivamente, do intervalo de credibilidade, e ainda o fator de convergência de Gelman e Rubin (1992) dado pela raiz quadrada do valor escalar \hat{R} , são apresentados na Tabela 5. Nota-se que houve convergência de todas as cadeias de parâmetros e que o índice de ferrugem atual é influenciado apenas pelo seu valor defasado para temperaturas abaixo do limiar $r = 21,5^{\circ}C$, já para valores de temperaturas acima de $r = 21,5^{\circ}C$, o índice atual é respondido pelo índice do período anterior e também pela temperatura do período anterior.

Tabela 5 Estimativas do modelo TARSO(2;1,0,1,1).

Regime	Parâmetros	Média	LI	LS	$\sqrt{\hat{R}}$
Regime 1	ϕ_{11}	0,8391	0,7436	0,9315	1,030
	τ_1	0,0014	0,0010	0,0018	0,9990
	ϕ_{21}	1,1115	0,9809	1,2583	1,0100
Regime 2	β_{21}	0,3267	0,1411	0,5253	0,9740
	τ_2	0,0101	0,0059	0,0148	1,0100

O evento de interesse é a catástrofe definida pelo cruzamento superior de um nível fixo u em $t + 2$ dado por

$$C_{t,2} = \{Y_{t+1} \leq u < Y_{t+2}\}, \quad (11)$$

sendo $u = 5\%$ e 10% . Nesses índices, o nível de ferrugem ainda é considerado baixo e, portanto, eficazes para a inserção de medidas de controle da doença. Os pontos de estudo das características operacionais do alarme são $t = 12$, $t = 39$ e $t = 133$. Foram escolhidos, dessa forma, de modo a permitir a análise do comportamento do alarme em pontos com poucas informações passadas e com muitas

informações passadas e também em pontos nos quais o evento de interesse, a catástrofe, dada pela Expressão (11), é satisfeita e onde não é.

Na Tabela 6 são apresentados os pontos de análise, o nível de cruzamento superior ($u\%$), o valor de \bar{k}^* e as regiões de alarme ótimas a partir das quais são obtidas as características do alarme. O valor de \bar{k}^* é o valor que faz com que $P(C_{t,2}|D_t) \approx P(A_{t,2}|D_t)$. Essa aproximação pode ser vista na Tabela 7.

Tabela 6 Nível de cruzamento, \bar{k}^* , e região de alarme dos pontos de cruzamento analisados na construção do alarme.

t	$u\%$	\bar{k}^*	RA
12	5	1,3301	[92;100]
39	10	1,4921	[1;5,5]
133	5	0,2845	[1;19,5]

Nota: RA-Região de Alarme.

Ainda, na Tabela 7 são apresentadas as características operacionais do alarme: probabilidade de alarme correto, probabilidade de detecção do evento, probabilidade de alarme falso e probabilidade de não-deteção do evento para os pontos $t = 12$, $t = 39$ e $t = 133$. O ponto $t = 12$ refere-se a setembro de 2000 no qual nota-se, para o nível de cruzamento $u = 5\%$ uma probabilidade de alarme correto e de detecção do evento em torno de 0%. Dessa forma seus complementares, probabilidade de alarme falso e não-deteção do evento, ficam muito próximos de 100%. O índice de infecção em $t = 39$ refere-se a dezembro de 2002 e considerando um cruzamento de $u = 10\%$, a probabilidade de alarme correto é de 52,13% e a probabilidade de detecção do evento é de 60,70%. Neste ponto, as probabilidades de alarme falso e de não-deteção do evento são, respectivamente, 47,87% e 39,30%. A observação $t = 133$ refere-se a outubro de 2010 e considerando um nível de cruzamento superior de 5%, a probabilidade de alarme correto é de

41,89% e a probabilidade de detecção do evento é de 41,04%. As probabilidades de alarme falso e de não-deteção do evento ficam em torno de 58%.

O desempenho do alarme pode ser melhor justificado analisando a Tabela 8 na qual são apresentados os índices de ferrugem nos pontos considerados e seus dois futuros mais próximos que são os índices que compõem o evento catástrofe dado pela Expressão (11).

Analisando a Tabela 8, o desempenho ruim do alarme em $t = 12$ justifica-se pelo fato do evento catástrofe, $C_{t,2} = \{Y_{t+1} \leq u < Y_{t+2}\}$, não ser satisfeito para o nível de cruzamento considerado. Porém, qualquer outro nível que se considerasse o alarme teria desempenho bom, já que, na prática, uma catástrofe aqui é rara pois em setembro o café está sem folhas e está fora do período infeccioso que se estende de dezembro a maio.

As probabilidades encontradas em outubro e dezembro já eram esperadas, isto é, em $t = 133$ e $t = 39$, já que inicia-se o período de infecção que coincide com o início do período chuvoso.

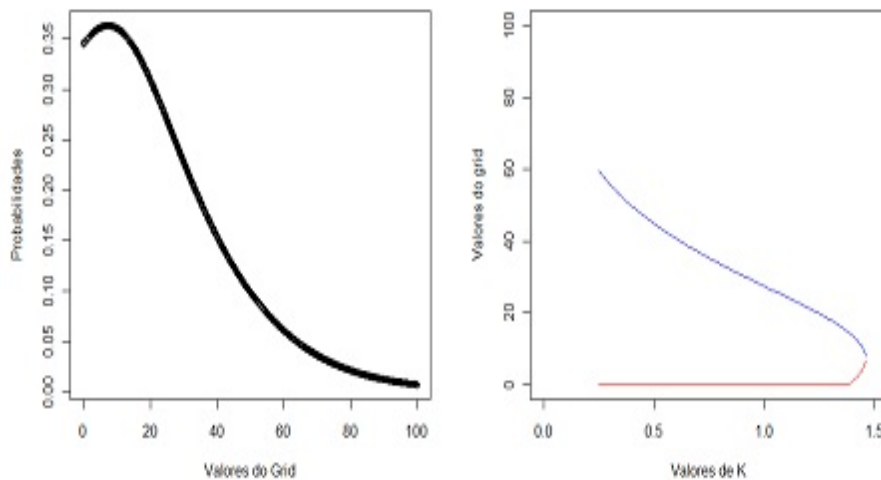
Como ilustração, na Figura 5 são apresentados os gráficos das probabilidades e das regiões de alarme para o ponto $t = 133$.

Tabela 7 Características operacionais do alarme em $t = 12$, $t = 39$ e $t = 133$ do processo TARSO(2;1,0,1,1).

u	t	$P(C_{t,2} D_t)$	$P(A_{t,2} D_t)$	$P(C_{t,2} A_{t,2}, D_t)$	$P(A_{t,2} C_{t,2}, D_t)$	$P(C_{t,2}^c A_{t,2}, D_t)$	$P(A_{t,2}^c C_{t,2}, D_t)$
5%	12	0,0521	0,0594	0,0009	0,0010	0,9991	0,9990
10%	39	0,2141	0,2493	0,5213	0,6070	0,4787	0,3930
5%	133	0,2495	0,2444	0,4189	0,4104	0,5811	0,5896

Tabela 8 Índices de ferrugem em t , $t + 1$ e $t + 2$.

t	Y_t	Y_{t+1}	Y_{t+2}
12	1%	0%	0%
39	1,5%	7,5%	15%
133	2,5%	4%	13%

Figura 5 Valores de probabilidade e regiões de alarme do ponto $t = 133$

4 CONCLUSÕES

O sistema de alarme se comporta razoavelmente bem em pontos nos quais a catástrofe ainda não ocorreu. Tanto na simulação como nos dados reais de ferrugem do café, o sistema fornece probabilidades mais altas de alarme correto e detecção em níveis de cruzamento mais altos. O maior número de observações passadas em um ponto parece não favorecer tanto as características do alarme, como ocorreu no estudo de simulação.

CONSIDERAÇÕES FINAIS

A técnica do sistema de alarme ótimo é de grande aplicação prática e pode ser adaptada a qualquer modelo estatístico que descreva doenças de qualquer outra cultura. Nota-se, porém, que na determinação de um conjunto de características operacionais de desempenho alto são muitas escolhas a serem feitas. No caso da ferrugem, por exemplo, depende-se de escolhas do valor *threshold*, defasagem, escolhas precisas \bar{k} ótimo, \bar{k}^* e, também, do nível u de cruzamento.

A ferrugem pode ser afetada por diversos outros fatores relacionados ao clima, motivo pelo qual pretende-se em um próximo trabalho melhorar o desempenho do alarme incorporando outras variáveis climáticas e, dessa forma, ampliar informações sobre esse problema de pesquisa.

REFERÊNCIAS

- ANTUNES, M.; AMARAL-TURKMAN, M. A.; TURKMAN, K. F. F. A bayesian approach to event prediction. **Journal of Time Series Analysis**, Clevedon, v. 24, n. 6, p. 631–646, Dec. 2003.
- COSTA, C.; SCOTTO, M. G.; PEREIRA, I. Optimal alarm systems for fiaparch processes. **Revstat - Statistical Journal**, Lisboa, v. 8, n. 1, p. 37–55, 2010.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-472, Nov. 1992.
- LINDGREN, G. Model processes in nonlinear prediction with application to detection and alarm. **The Annals of Probability**, Hayward, v. 8, n. 4, p. 775–792, 1980.
- . Optimal prediction of level crossing in gaussian processes and sequences. **The Annals of Probability**, Hayward, v. 13, n. 3, p. 804–824, 1985.
- . Prediction from a random time point. **The Annals of Probability**, Hayward, v. 3, n. 3, p. 412–423, 1975.
- MARÉ, J. de. Optimal prediction of catastrophes with applications to gaussian processes. **The Annals of probability**, Hayward, v. 8, n. 4, p. 841–850, 1980.
- MONTEIRO, M.; PEREIRA, I.; SCOTTO, M. G. Optimal alarm systems for count processes. **Communications in Statistics-Theory and Methods**, New York, v. 37, n. 19, p. 3054–3076, 2008.
- SVENSSON, A. et al. Optimal prediction of catastrophes in autoregressive moving-average processes. **Journal of time Series Analysis**, Clevedon, v. 17, n. 5, p. 511-531, Oct. 1996.
- SVENSSON, A.; HOLST, J. Prediction of high water levels in the Baltic. **Journal of the Turkish Statistical Association**, Ankara, v. 1, n. 1, p. 39-46, 1997.

TURKMAN, M. A. A.; TURKMAN, K. F. Optimal alarm systems for autoregressive processes: a bayesian approach. **Computational Statistics & Data Analysis**, New York, v. 10, n. 3, p. 307–314, Dec. 1990.