

UNIVERSIDADE FEDERAL DE ALFENAS

KAROLINE PEREIRA BRANCO

INFERÊNCIA PROBABILÍSTICA PARA SEGURO PARAMÉTRICO

Alfenas/MG

2023

KAROLINE PEREIRA BRANCO

INFERÊNCIA PROBABILÍSTICA PARA SEGURO PARAMÉTRICO

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria.

Orientador: Prof. Dr. Luiz Alberto Beijo
Coorientador: Prof. Dr. Reinaldo Antônio Gomes Marques

Alfenas/MG

2023

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Branco, Karoline Pereira.

Inferência probabilística para seguro paramétrico / Karoline Pereira

Branco. - Alfenas, MG, 2023.

81 f. : il. -

Orientador(a): Luiz Alberto Beijo.

Dissertação (Mestrado em Estatística Aplicada e Biometria) -

Universidade Federal de Alfenas, Alfenas, MG, 2023.

Bibliografia.

1. Inferência bayesiana. 2. Precificação atuarial. 3. Distribuição GEV. 4. Gerenciamento de risco climático. I. Beijo, Luiz Alberto, orient. II. Título.

KAROLINE PEREIRA BRANCO

INFERÊNCIA PROBABILÍSTICA PARA SEGURO PARAMÉTRICO

O Presidente da banca examinadora abaixo assina a aprovação da Dissertação apresentada como parte dos requisitos para a obtenção do título de Mestre em Estatística Aplicada e Biometria pela Universidade Federal de Alfenas. Área de concentração: Estatística Aplicada e Biometria

Aprovada em: 29 de junho de 2023.

Prof. Dr. Luiz Alberto Beijo - Universidade Federal de Alfenas - UNIFAL-MG

Presidente da Banca Examinadora

Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Reinaldo Antônio Gomes Marques

Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Prof. Dr. Gilberto Rodrigues Liska

Instituição: Universidade Federal de São Carlos - UFSCar

Profa. Dra. Thais Cristina de Oliveira da Fonseca

Instituição: Universidade Federal do Rio de Janeiro - UFRJ

Prof. Dr. Helton Graziadei de Carvalho

Instituição: Fundação Getúlio Vargas - FGV



Documento assinado eletronicamente por **Luiz Alberto Beijo, Professor do Magistério Superior**, em 29/06/2023, às 16:35, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código



verificador **1025769** e o código CRC **357523AE**.

AGRADECIMENTOS

Gostaria de prestar meus mais sinceros agradecimentos a todos que de alguma forma contribuíram para a realização desta pesquisa.

Agradeço ao Criador do Universo, porque sem Ele nada seria possível.

Agradeço aos meus pais, Alessandra Branco e Paulo Branco, por todo esforço empenhado na minha educação, por sua presença, apoio e amor incondicional na minha vida sempre.

Agradeço a todos os mestres que contribuíram com a minha formação acadêmica, em especial, agradeço aos meus orientadores Luiz Beijo e Reinaldo Antônio Marques pela atenção e pelas inúmeras horas dedicadas para sanar as minhas dúvidas, me ensinar e me colocar na direção correta.

Agradeço ao meu namorado, Caio Tiso, por todo apoio emocional, amor e incentivo durante a minha jornada.

Agradeço aos meus avós, Sebastião Batista Pereira e Lucy Aparecida Pereira, pelo carinho e incentivo que sempre me deram.

Agradeço à Universidade Federal de Alfenas pelo suporte e estrutura oferecidos no período de desenvolvimento deste trabalho.

Agradeço à Cooperativa Regional de Cafeicultores em Guaxupé - Cooxupé - pela concessão dos dados utilizados na pesquisa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Portanto, gostaria também de agradecer à CAPES pelo apoio financeiro.

Esta pesquisa foi vencedora do 1º Prêmio SUSEP de pesquisa em Seguros

(<https://www.premiosusep.susep.gov.br/>)

“Se enxerguei mais longe, foi porque me apoiei sobre os ombros de gigantes”.

(Isaac Newton, 1676)

RESUMO

O objetivo deste trabalho foi elaborar e aplicar um método de modelagem para contratos de seguros agrícolas paramétricos com cobertura para a ocorrência de eventos climáticos extremos, visando a determinação de gatilhos indenizatórios com risco base reduzido. O método proposto utilizou a distribuição generalizada de valores extremos (GEV) para modelar os eventos climáticos extremos e empregou as probabilidades de excedência encontradas para os quantis extremos como variável explicativa de um modelo logístico predictor de perdas na safra. A inferência bayesiana foi aplicada para a estimação dos parâmetros da distribuição GEV e dos coeficientes do modelo logístico. Posteriormente, verificou-se a acurácia e o custo para seguradora e segurado para todos os possíveis gatilhos contratuais, com a intenção de fornecer embasamento técnico para que o gestor do contrato identifique o gatilho mais seguro e com custos viáveis para o produto. Após a apresentação do método foi realizado um estudo de caso visando a sua aplicação para a elaboração de um contrato para a proteção de lavouras de café contra a ocorrência de veranicos extremos no período da florada em algumas cidades do Estado de Minas Gerais. Foram ajustados dois modelos, um com distribuição a *priori* informativa para a estimação dos parâmetros das distribuições GEV e outro com distribuição a priori não informativa. Os resultados encontrados foram promissores. O modelo proposto apresentou 87,5% de acurácia para as duas estruturas de distribuição a priori ao relacionar o evento climático com a ocorrência de perdas na safra, mesmo em um cenário de escassez de dados. Além disso, com a utilização a *priori* informativa foi possível encontrar um gatilho ótimo em relacionar o evento climático e a ocorrência de perdas e que trouxe um custo viável de comercialização para ambos os agentes envolvidos. A utilização da abordagem inferencial bayesiana possibilitou, por meio de intervalos de credibilidade, que a incerteza do processo fosse quantificada com precisão razoável em todas as etapas da modelagem, fornecendo maior grau de embasamento para os gestores do contrato tomarem decisões. Conclui-se que o método aqui proposto mostrou-se promissor e pode ser adaptado para contratos de diferentes culturas e eventos climáticos.

Palavras-chave: inferência bayesiana; precificação atuarial; distribuição GEV; gerenciamento de risco climático.

ABSTRACT

The objective of this work was to develop and apply a modeling method for parametric agricultural insurance contracts with coverage for the occurrence of extreme weather events, aiming at determining indemnity triggers with reduced base risk. The proposed method used the generalized distribution of extreme values (GEV) to model extreme weather events and used the exceedance probabilities found for the extreme quantiles as an explanatory variable for a logistic model that predicts crop losses. Bayesian inference was applied to estimate the parameters of the GEV distribution and the coefficients of the logistic model. Subsequently, the accuracy and cost for the insurer and insured for all possible contractual triggers were verified, with the intention of providing a technical basis for the contract manager to identify the safest trigger and with viable costs for the product. After presenting the method, a case study was carried out aiming at its application for the elaboration of a contract for the protection of coffee plantations against the occurrence of extreme dry spells during the flowering period in some cities of the State of Minas Gerais. Two models were fitted, one with an informative a priori distribution for estimating the parameters of the GEV distributions and the other with a non-informative a priori distribution. The results found were promising. The proposed model showed 87.5% accuracy for the two a priori distribution structures when relating the weather event to the occurrence of crop losses, even in a scenario of scarce data. In addition, with the informative a priori use, it was possible to find an optimal trigger to relate the climatic event and the occurrence of losses and that brought a viable cost of commercialization for both agents involved. The use of the Bayesian inferential approach made it possible, through credibility intervals, for the uncertainty of the process to be quantified with reasonable precision in all stages of the modeling, providing a greater degree of basis for the contract managers to make decisions. It is concluded that the method proposed here proved to be promising and can be adapted for contracts of different cultures and climatic events.

Keywords: bayesian inference; actuarial pricing; GEV distribution; climate risk management.

LISTA DE FIGURAS

Figura 1	- Exemplo da obtenção da série de máximos de ordem 1 para uma base de dados composta por cinquenta observações, considerando blocos de dez anos.....	25
Figura 2	- Location of the studied cities.....	52
Figura 3	- Summary of research steps.....	55
Figura 4	- Values of a series of second-order maximums of extreme dry-spells in the R1 period (Sep-Dec) in the cities of Guaxupé, Monte Santo de Minas (MSM) and Monte Carmelo (MC).....	56
Figura 5	- Sequences of the maximum consecutive days with less than 3mm of rain in the cities of Guaxupé, Monte Santo de Minas and Monte Carmelo in the years of 2014 to 2021.....	57
Figura 6	- HPD _{95%} interval for the exceedance probability of the observed dry spells in Guaxupé (G), Monte Santo de Minas (MSM), and Monte Carmelo (MC) in the years from 2014 to 2021.....	58
Figura 7	- Posterior distribution of the trigger at the 87.5% accuracy considering informative prior for GEV (a) and noninformative prior for GEV (b).....	59
Figura 8	- Posterior distribution of the trigger when cost to insurer is the lowest found - \$2,458.33, considering informative prior (a) and non-informative prior (b).....	60
Figura 9	- Median (a) and mean (b) cost and their HPD _{95%} interval to the insured to informative prior model and median (c) and mean (d) cost and their HPD _{95%} interval to the insured to non informative prior model.....	61
Figura 10	- Median total cost and their HPD _{95%} interval to the insured considering the triggers 21% to 27%, for informative (a) and non-informative (b) prior model.....	61
Figura 11	- Median total cost considering informative prior model to the insured (a) and insurer (b) and considering noninformative prior to the insures (c) and insurer (d) and their HPD _{95%} interval.....	62

LISTA DE QUADROS

Quadro 1	- Características dos contratos de seguro tradicional e de seguro paramétrico.....	20
Quadro 2	- Esquematização das seis fases fenológicas do cafeeiro arábica, durante 24 meses, nas condições climáticas tropicais do Brasil.....	39

LISTA DE TABELAS

Tabela 1	- Possible events predicted by the model	47
Tabela 2	- Cost structure for the insurer and insured	48
Tabela 3	- Jeffreys' interpretation scale for the Bayes Factor.....	50
Tabela 4	- Description of the prior information in fitting the GEV distribution.....	53
Tabela 5	- Hypothetical costs used for evaluating the proposed model.....	54
Tabela 6	- Descriptive statistics for extreme dry spells data for the cities of Guaxupé, Monte Santo de Minas and Monte Carmelo.....	55
Tabela 7	- Results of the $HPD_{95\%}$ intervals of the parameters of the GEV distribution according to the city and the results of the criteria for the analysis of convergence of the chains of the posterior distribution of the parameters, namely Raftery and Lewis (R&L), Geweke (GW) and Heidelberger-Welch (H&W).....	56
Tabela 8	- Credibility interval $HPD_{95\%}$ for the logistic model coefficients, the posterior median for results, the Bayes factor in favor of H1 and $\exp(\beta)$	57
Tabela 9	- Credibility interval $HPD_{95\%}$ for the posterior distribution of the trigger at the accuracy of 87.5%.....	59
Tabela 10	- Credibility interval $HPD_{95\%}$ for the posterior distribution of the trigger when cost to the insurer is the lowest found - \$2,458.33....	59

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DE LITERATURA	18
2.1	SEGURO AGRÍCOLA PARAMÉTRICO.....	18
2.2	ANÁLISE DE VALORES EXTREMOS.....	24
2.3	REGRESSÃO LOGÍSTICA.....	29
2.4	INFERÊNCIA BAYESIANA.....	31
2.4.1	Estimação Bayesiana por meio de algoritmo MCMC.....	34
2.4.2	Teste de hipótese Bayesiano.....	36
2.5	CULTURA DE CAFÉ.....	38
2.6	ARTIGO 1 - BAYESIAN MODELING TO AGRICULTURAL PARAMETRIC INSURANCE FOR EXTREME WEATHER EVENTS.....	42
5	CONSIDERAÇÕES FINAIS	66
	REFERÊNCIAS	67
	APÊNDICE A	73
	APÊNDICE B	79

1 INTRODUÇÃO

A contratação de seguros agrícolas é uma estratégia para a mitigação dos riscos incidentes sobre a produção agrícola, que é altamente dependente das condições climáticas (HEIMFARTH; MUSSHOF, 2011). Estes seguros são ferramentas por meio das quais os segurados são indenizados quando da ocorrência de eventos adversos que prejudiquem a produção e causem, conseqüentemente, perdas financeiras. Desse modo, o seguro agrícola é fundamental para o desenvolvimento do agronegócio, uma vez que possibilita a modernização e a competitividade do setor, tendo em vista que instituições de crédito terão maior segurança em financiar os produtores (FORNAZIER; SOUZA; PONCIANO, 2012).

Uma modalidade que pode ser aplicada em contratos visando à proteção agrícola é o seguro paramétrico (CARTER; CHIU, 2018). De modo geral, no processo de elaboração de um contrato de seguro paramétrico é estipulado um índice que deve ser monitorado como, por exemplo, volume de chuvas (mm), temperatura (°C) ou dias secos e quando este índice atingir um limiar pré-estabelecido, chamado de gatilho, o pagamento da indenização acordada é automaticamente realizado. Neste tipo de contrato existe a pressuposição de que o gatilho se relaciona com uma perda futura na safra (MILLER, 2020).

Desse modo, o contratante do seguro paramétrico é indenizado sem que haja a necessidade de esperar a produção ser realizada, para que então um perito quantifique o prejuízo e somente depois a indenização seja calculada e paga, como ocorre nos produtos tradicionais de seguro agrícola. As características deste tipo de seguro superam, então, em alguns aspectos as características dos seguros agrícolas tradicionais. O primeiro aspecto é a agilidade no pagamento de indenizações, que depende apenas da conferência do gatilho para ser executada.

O segundo aspecto é relacionado ao custo do produto que tende a ser mais barato para o segurado, já que a seguradora não precisa repassar para o segurado os custos com mão de obra qualificada para quantificar perdas. O terceiro aspecto é associado à solução de problemas relacionados à assimetria de informações, como ocorre nos contratos de seguro tradicionais. Isto porque a indenização do seguro paramétrico é paga com base em um gatilho acordado previamente e que é facilmente auditável por ambas as partes e que, portanto, é mais difícil de ser manipulado (PROKOPCHUK *et al.*, 2020; DALHAUS; MUSSHOF; FINGER, 2018).

A despeito destes benefícios, um entrave à aceitação e disseminação deste tipo de seguro é a ocorrência do risco base, que gera insegurança à seguradora e ao segurado sobre o produto (SUN *et al.*, 2020). Este risco se materializa quando ocorre o pagamento das indenizações mesmo que não haja perda real na produção (risco de base positivo) ou quando não há pagamento de indenização quando há perda real na produção (risco base negativo). Ou seja, este risco pode ser caracterizado como a probabilidade de que o gatilho indenizatório definido no contrato não seja capaz de correlacionar de forma correta a ocorrência de perdas futuras com as variações do índice monitorado (TURNER, 2020).

A materialização do risco base é indesejada, uma vez que causa enriquecimento ou empobrecimento do segurado, o que não é o objetivo do contrato de seguro. Entretanto, é possível que contratos bem elaborados consigam reduzir o risco base e ampliar a confiança das seguradoras em sua comercialização e dos segurados em sua aquisição (CARTER; CHIU, 2018). Desse modo, a elaboração de estratégias que consigam relacionar a produção agrícola com as variações de um índice climático de forma confiável, possibilitaria a maior difusão deste tipo de contrato.

O seguro paramétrico pode ser bastante útil para agricultores como forma de adaptação e gestão de riscos nas lavouras, advindos de eventos climáticos extremos que, apesar de menos frequentes, ocasionam grandes prejuízos (AFSHAR *et al.*, 2021). Para que seja possível a elaboração de um contrato paramétrico visando a proteção de lavouras contra determinado evento extremo, primeiro é necessário que o comportamento deste evento seja conhecido e modelado.

Para modelar eventos extremos, são empregadas as metodologias específicas da Teoria de Valores Extremos (COLES, 2001). Os métodos dessa teoria permitem aos pesquisadores avaliar a probabilidade de ocorrência e a magnitude dos eventos extremos de forma probabilística (AHMAD *et al.*, 2022). Uma distribuição de probabilidade amplamente utilizada na modelagem de eventos extremos é a distribuição generalizada de valores extremos, do inglês Generalized Extreme Value (GEV), proposta por Jenkinson (1955) (AGUIRRE; NOGUEIRA; BEIJO, 2020). Esta distribuição é usada para modelar séries de valores extremos que são obtidos pelo método de blocos máximos (COLES, 2001). Pesquisas como Martins *et al.* (2015), Carvalho *et al.* (2013) e Butturi-Gomes, Beijo e Avelar (2019) também utilizaram a distribuição GEV para modelar eventos climáticos extremos, obtendo resultados satisfatórios.

Após a determinação do comportamento do evento extremo, é preciso verificar a relação entre este e a ocorrência de perdas na produção resultante. Uma estratégia que pode ser utilizada para este tipo de ação é o emprego de métodos de regressão, que realizam tarefas de classificação e predição (Draper e Smith 1998). Para prever a ocorrência de eventos binários, como "perda" ou "sem perda" em decorrência de um evento climático, o método de regressão logística é bastante apropriado (FERNANDES *et al.*, 2021).

Conforme aplicado na pesquisa de Figueiredo *et al.* (2018), esse método de regressão fornece a probabilidade de uma perda ser observada e, após definir um limiar de classificação, os eventos são então definidos como "perda" ou "sem perda" de acordo com ele. Assim, o limiar de classificação mais preciso é aquele que consegue prever de forma mais assertiva o evento de interesse e, portanto, é aquele que reduz o risco base.

É importante ressaltar que o gatilho deve apresentar não só segurança, mas também um custo viável para o segurado e para a seguradora. Figueiredo *et al.* (2018) propuseram uma estratégia para avaliar e selecionar gatilhos de indenização para contratos de seguro paramétricos para inundações em algumas regiões da Jamaica. Na estratégia proposta, foram utilizadas métricas estatísticas para avaliar a segurança na indicação de perda de cada possível gatilho, bem como os custos para a seguradora e para o segurado. Após avaliar a segurança e o custo-benefício do contrato, pode-se selecionar o melhor gatilho de indenização, ou seja, o gatilho com menor risco base e custo.

Para realizar a estimação dos parâmetros da distribuição GEV e dos coeficientes da regressão logística optou-se aqui pela abordagem inferencial Bayesiana. Esta abordagem apresenta boa performance em processos de estimação quando os dados se referem a valores extremos, que são escassos por natureza. Isso porque essa abordagem permite a inclusão de evidências complementares sobre o processo por meio da inserção de conhecimento prévio (AHMAD *et al.*, 2022). A lógica Bayesiana também facilita o entendimento das estimativas para o senso comum, pois a especificação de uma distribuição para os parâmetros torna sua interpretação mais direta (GELMAN *et al.*, 2013). Além disso, esta abordagem inferencial possibilita a quantificação das incertezas do processo de estimação por meio da especificação de intervalos de credibilidade para as quantidades desconhecidas estimadas (PAULINO; TURKMAN; MURTEIRA, 2003).

No contexto brasileiro a adoção de contratos de seguro agrícola é bastante relevante, uma vez que a produção nacional alimenta 10% da população mundial e é responsável por 40% das exportações, cerca de um quarto de toda a riqueza e um quinto de todos os empregos gerados no país (Centro de Estudos Avançados em Economia Aplicada, 2022; Embrapa, 2020). Dentre outras culturas, uma na qual o Brasil se destaca é na produção de café. O país é o maior produtor e exportador desse grão no mundo, tendo em 2022 o país produziu mais de 39 milhões de sacas que chegaram a mais de 120 países e geraram uma receita cambial de mais de US\$ 9 bilhões.

Apesar destes números expressivos e da importância econômica e social que essa cultura apresenta para o país, uma vez que também gera milhares de empregos direta e indiretamente, a maior parte das lavouras contam com a sorte em relação às variações climáticas (Centro de Estudos Avançados em Economia Aplicada, 2021). Isto porque, segundo a Confederação da Agricultura e Pecuária do Brasil, apenas 130 mil dos 1,8 milhão de hectares produtores de café possuem algum tipo de seguro. Isto demonstra que o mercado de seguros agrícolas para lavouras de café ainda apresenta grande potencial de expansão no país.

O cafeeiro é uma planta que tem características que podem ser enquadradas na estruturação de um contrato de seguro paramétrico. A principal característica associada é a diferença de necessidade hídrica a depender da fase fenológica da planta. Uma das fases com maior necessidade de precipitação pluvial é a primeira fase reprodutiva do cafeeiro, na qual ocorrem a florada, a formação dos chumbinhos e a expansão dos frutos (CAMARGO; CAMARGO, 2001). Estresse hídrico nesta fase pode prejudicar o crescimento e desenvolvimento dos grãos. Portanto, há coerência na elaboração de um contrato de seguro paramétrico que enseje a proteção da lavoura em relação à períodos extremos de estiagem nesta fase crítica.

Apesar disso, não existe na literatura uma metodologia focada na modelagem deste tipo de produto. Isto torna explícita a necessidade de inovações na forma de modelar contratos de seguros paramétricos. Como sugerido nesta pesquisa, um processo de modelagem focado no impacto que a ocorrência de eventos extremos traz para as lavouras por um viés probabilístico, permite que gestores dos contratos consigam quantificar os riscos do produto durante todo o processo decisório relacionado, o que pode facilitar a gerência de contratos de seguro paramétrico, tornando-os mais confiáveis e aceitos.

Desse modo, o objetivo geral deste trabalho foi o desenvolvimento de um método para a estruturação de contratos de seguros agrícolas paramétricos que visem a proteção de lavouras contra a ocorrência de eventos climáticos extremos e que tenham o menor risco base possível, resultando em contratos seguros para os agentes envolvidos. Para atingir esse objetivo, a teoria de valores extremos via distribuição generalizada extrema (GEV) foi aplicada para modelar eventos climáticos extremos, regressão logística foi usada para criar o modelo preditor de perdas e a determinação dos gatilhos mais adequados foi realizada conforme proposto por Figueiredo *et al.* (2018).

Além disso, foi utilizada a abordagem inferencial Bayesiana, permitindo que a incerteza em relação a todos os resultados fosse descrita como intervalos de credibilidade. O objetivo específico da pesquisa foi a elaboração de um contrato paramétrico contra períodos extremos de estiagem para lavouras de café em algumas cidades do sul de Minas Gerais, por meio de um estudo de caso que visou à aplicação do método proposto. A principal contribuição desta dissertação foi a aplicação dos pressupostos da teoria dos valores extremos e da abordagem Bayesiana na determinação de gatilhos seguros para contratos paramétricos agrícolas.

2 REVISÃO DE LITERATURA

2.1 SEGURO AGRÍCOLA PARAMÉTRICO

O seguro agrícola é uma ferramenta que possibilita que os produtores sejam indenizados quando da ocorrência de eventos adversos que prejudiquem a produção e causem, conseqüentemente, perdas financeiras. Para Fornazier, Souza e Ponciano (2012), o seguro agrícola é fundamental para o desenvolvimento do agronegócio. Isto porque esta ferramenta possibilita a modernização e competitividade do setor, uma vez que instituições de crédito terão maior segurança em financiar os produtores.

Nos últimos anos, observou-se no mercado segurador a emergência de um novo produto chamado “seguro paramétrico”, que se apresentou como um tipo de contrato que, dentre outras possibilidades, tem grande potencial de aplicação em seguros agrícolas (CARTER; CHIU, 2018). Nesta modalidade, o pagamento da indenização é realizado quando surge uma expectativa de perda na produção, não sendo necessário aguardar a constatação de prejuízo físico antes do pagamento, como ocorre nos contratos de seguros tradicionais (MILLER, 2020).

De acordo com Zurich (2021) e World Bank (2017), como na modalidade tradicional de seguro, a indenização depende da percepção física do prejuízo, ou seja, depende da constatação e estimativa da perda por profissionais qualificados, os pagamentos das indenizações ocorrem de modo moroso. Ademais, nesses contratos há o favorecimento da ocorrência de problemas ligados à assimetria de informação, como o risco moral e a seleção adversa (JONHSON, 2021; HEIMFARTH; MUSSHOF, 2011). Isto ocorre porque, em geral, os segurados possuem mais informações sobre seus comportamentos reais e exposição ao risco do que as seguradoras. Além disso, muitas vezes, o seguro tradicional incorre em custos de aquisição que não são economicamente viáveis para os pequenos produtores (CARTER; CHIU, 2018).

O seguro paramétrico é apresentado com a expectativa de auxiliar na superação dessas lacunas características dos produtos tradicionais (PROKOPCHUK et al., 2020; DALHAUS; MUSSHOF; FINGER, 2018). De acordo com o World Bank (2017) um produto de seguro paramétrico pode ser definido como um contrato no qual há o “pagamento de benefícios com base em um índice pré-determinado para a perda

de ativos e investimentos que ocorram como resultado do clima ou outros eventos catastróficos”.

No contrato de seguro paramétrico considera-se que o índice para o pagamento das indenizações foi acionado quando algum limite preestabelecido em contrato, chamado de gatilho da indenização, é alcançado ou superado (ZURICH, 2021). No caso de seguros agrícolas, por exemplo, o índice indenizatório pode ser relacionado a excesso de chuvas (mm), a períodos de secas (dias), a temperaturas extremas (°C), entre outros eventos que possam prejudicar o desenvolvimento das plantas (SHIRSATH *et al.*, 2019).

Considere que x é o índice indenizatório de um contrato de seguro paramétrico. Conforme adaptado de Lema *et al.* (2019), o mecanismo de pagamento de indenizações no contrato de seguro paramétrico pode ser descrito como:

$$\text{indicador de indenização} = \begin{cases} 0; & x < \text{gatilho} \\ 1; & x \geq \text{gatilho} \end{cases}$$

em que ou o gatilho indenizatório não é alcançado e não há pagamentos de indenizações, ou o gatilho é superado e a indenização acordada no contrato (I) é paga ao segurado.

O contrato de seguro paramétrico relacionado a índices climáticos pode ser de grande valia na gestão de riscos ligados à agricultura. Isto porque, caso ocorra um fator climático adverso que acione o gatilho durante alguma fase fenológica da safra, a indenização é paga instantaneamente sem que seja necessário esperar a colheita e o trabalho do perito para realizar a constatação do prejuízo. O quadro 1 apresenta algumas características dos contratos de seguro tradicional e de seguro paramétrico.

Quadro 1 - Características dos contratos de seguro tradicional e de seguro paramétrico

Seguro Tradicional	Seguro Paramétrico
O pagamento é acionado por uma perda real	O pagamento é acionado por um evento real
O reembolso é feito em acordo com a avaliação realizada por perito das perdas sofridas	O pagamento é feito de acordo com o valor do parâmetro conforme acordado previamente na apólice
A apólice está sujeita à exclusões e retenções auto seguradas	A apólice está sujeita ao índice ou parâmetro
A indenização somente é paga após avaliação por peritos e pelas seguradoras – processo demorado	A indenização é paga em até 30 dias, sem a necessidade de perícia - processo rápido
O período é anual	O período é personalizado
A redação da apólice é o padrão da empresa ou uma forma modificada do padrão da empresa - contrato de adesão	A apólice é altamente personalizável, visando atender as demandas de cada cliente

Fonte: adaptado de Turner (2020)

De acordo com Turner (2020) a estruturação de um seguro paramétrico ocorre com a determinação de três fatores: 1) O índice ou métrica que se relaciona diretamente com os custos e despesas do segurado; 2) O gatilho relacionado a este índice que causa o acionamento da indenização; e 3) O valor da indenização. Destarte, de acordo com Johnson (2021, p. 122), “Os contratos paramétricos especificam os dados e os algoritmos a serem usados, as áreas geográficas e os intervalos de tempo para avaliação e os resultados que irão ou não acionar um pagamento, tornando a tomada de decisão mais eficiente e auditável”.

De acordo com Johnson (2021) e Heimfarth e Musshoff (2011), a mudança na forma de estruturação de um contrato paramétrico em relação aos seguros tradicionais, bem como a mudança nos cálculos que determinam os pagamentos das indenizações nesta modalidade, resolveria os problemas de informação assimétrica nos contratos, como seleção adversa e risco moral, já que os eventos segurados são passíveis de controle por todos os agentes envolvidos.

Além disso, de acordo com Miller (2020), Turner (2020) e Radu e Alexandru (2022) os contratos de seguro paramétrico também trazem vantagens econômicas, uma vez que a seguradora dispensa mão de obra especializada para quantificar perdas e o segurado paga somente pela cobertura que deseja. Também há vantagens de tempo e de flexibilidade, uma vez que as indenizações são pagas automaticamente quando o gatilho é acionado e o produto é totalmente customizável às necessidades do cliente.

No Brasil, o seguro paramétrico ainda é um produto inovador que está iniciando seu processo de introdução no mercado. O primeiro contrato no país foi realizado no início de 2017, sendo um produto que visava à proteção de lavouras de milho, algodão e soja para uma grande produtora nos estados da Bahia, Minas Gerais e Mato Grosso (MIQUELLUTI; OZAKI, 2021). Após isto, somente em 2021 foi celebrado o segundo contrato de seguro paramétrico no Brasil, sendo este para a proteção de uma lavoura de cacau contra a ocorrência de seca no período da florada (WALENDORFF, 2021). Foi também somente em 2021 que esta modalidade entrou no catálogo do Governo Federal como produto que participa dos programas públicos de incentivo à contratação de seguro agrícola (NEWE SEGUROS, 2020).

Apesar das qualidades e potencialidades do seguro paramétrico, um dos entraves à sua aceitação por parte dos segurados e das seguradoras é a gestão de riscos dos contratos, uma vez que este tipo de seguro é suscetível ao chamado “risco base” (AFSHAR, 2021, SUN *et al.*, 2020). O risco base é descrito como incompatibilidade entre o gatilho escolhido e a real ocorrência de perdas na produção, que pode resultar em pagamentos de indenização quando não há a perda real (risco base positivo) ou então na omissão dos pagamentos de indenização quando a perda existe (risco base negativo) (FIGUEIREDO *et al.*, 2018; TURNER, 2020).

Segundo Sun *et al.* (2020) e Johnson (2021), o risco base é um problema bastante preocupante quando o gatilho do seguro paramétrico é o clima, como no caso de produtos paramétricos agrícolas. Isso acontece porque, para este caso, existem três riscos que devem ser controlados visando a minimização do risco base. São eles o risco espacial, o risco de projeto e o risco temporal. De acordo com os autores, o risco espacial acontece quando existe grande distância entre a área da plantação e a estação meteorológica, fazendo com que as condições meteorológicas da área segurada possam não ser condizentes com os dados obtidos pela estação.

O risco de projeto ocorre quando é feita uma relação equivocada entre o clima e o rendimento das culturas ou quando o gatilho climático escolhido para se relacionar com a produtividade não é adequado. Já o risco temporal surge quando são considerados períodos de risco equivocados para a cobertura do contrato, uma vez que o índice dos contratos é, em geral, definido para períodos específicos do calendário (SUN *et al.* 2020; JONHSON, 2021).

A definição equívoca do real índice causador de perdas nos contratos de seguros paramétricos gera grande preocupação para as seguradoras. Isto porque, quando o risco base se materializa, ocorre o enriquecimento ou empobrecimento indevido do segurado, fazendo com que a seguradora, muitas vezes, aja fora dos termos contratuais para atenuar as falhas nos seus dispositivos de definição de gatilhos (JONHSON, 2021).

É desejável que tanto o enriquecimento como o empobrecimento do segurado sejam evitados, uma vez que a razão de existência do seguro é tão somente o restabelecimento do equilíbrio econômico do segurado, sem que a indenização o deixe em situação financeira melhor ou pior (SOUZA, 2007). Destarte, tem-se que a correta compatibilização entre os eventos de perda e de indenização não são tarefa fácil, demonstrando que “embora seja verdade que eles [produtos paramétricos] são simples uma vez executados e após o recebimento do pagamento, isso não é verdade para sua criação” (TURNER, 2020, p. 2).

Entretanto, Carter e Chiu (2018) colocam que a minimização do risco base é possível em contratos bem projetados nos quais as estimativas paramétricas se correlacionam de fato com as perdas dos segurados. Dessa forma, alguns autores realizaram pesquisas buscando projetar contratos de seguros paramétricos que conseguissem relacionar de forma eficiente a ocorrência de perdas, ou variações na produção, com o pagamento das indenizações, visando o estabelecimento de contratos de maior segurança para os envolvidos.

Afshar *et al.* (2021) buscaram a redução do risco base em produtos paramétricos para a proteção de lavouras de arroz na Índia contra os riscos da variabilidade das chuvas de monções. O volume de chuvas (mm) foi especificado como o índice a ser monitorado no contrato proposto. Os autores utilizaram na pesquisa modelos de simulação de cultivos e monitoramento da fenologia dos cultivos via satélite. Como resultado, foi obtido um modelo capaz de explicar 54% da variabilidade da produção de arroz, considerando a variação de chuvas em diferentes

fases fenológicas do grão. Assim, o conhecimento das diferentes necessidades hídricas de cada estágio fenológico da planta permite a verificação da relação entre a ocorrência de chuvas e a variação na produção resultante, o que pode auxiliar na elaboração do contrato de seguro paramétrico agrícola com risco base reduzido.

Prokopchuk *et al.* (2020) desenvolveram um produto de seguro paramétrico para lavouras de milho na Ucrânia. Para isso, os autores trabalharam com um gatilho de indenização que combinou fatores de precipitação, temperatura e umidade relativa do ar nas fases de floração e enchimento dos grãos, que são as fases mais sensíveis a estas questões climáticas, para verificar o impacto desses fatores na produção resultante. Como resultado, os autores chegaram em um gatilho combinado capaz de calcular a variação da produção a cada combinação de fatores meteorológicos que houvesse nas fases fenológicas críticas, auxiliando a definição de gatilhos de contratos mais eficientes.

Dalhaus, Musshoff e Finger (2018) realizaram a estruturação de três modelos paramétricos para proteger lavouras de trigo alemãs durante a fase de floração. Como resultado, os autores afirmaram que os modelos ajustados foram capazes de reduzir a exposição ao risco dos segurados em relação a não ter seguro, bem como dois dos modelos também foram capazes de aumentar a utilidade esperada do produto de seguro para os envolvidos. Ademais, os pesquisadores enfatizaram que a consideração da fase fenológica das culturas na estruturação do produto paramétrico reduz o risco base temporal.

No Brasil, Miquelutti, Ozaki e Miquelluti (2022) desenvolveram um produto de seguro paramétrico climático para soja, visando proteção contra a seca no Estado do Paraná. Por meio de dados de precipitações diárias captadas por 41 estações meteorológicas do estado, os autores criaram um índice de precipitação padronizado. Posteriormente, desenvolveram um modelo de regressão quantílica geograficamente ponderada (GWQlasso) no qual a produção de soja era variável dependente e o índice de precipitação padronizado era a variável independente. Como resultado, os autores pontuaram que o modelo proposto teve desempenho satisfatório, inclusive superando a modalidade de seguro agrícola tradicional, e, portanto, seu emprego foi indicado para modelar contratos paramétricos mesmo em locais que apresentem quantidade limitada de dados.

Tendo em vista o importante papel dos eventos climáticos no desenvolvimento da agricultura em geral, Afshar *et al.* (2021) pontua que o seguro paramétrico pode

ser bastante útil para agricultores como forma de adaptação e gestão de riscos nas lavouras, advindos de eventos climáticos extremos. A menção a eventos extremos como um foco da utilização destes produtos, segundo os autores, advém do fato de que esses eventos, a despeito de serem menos frequentes, causam prejuízos consideráveis aos agricultores.

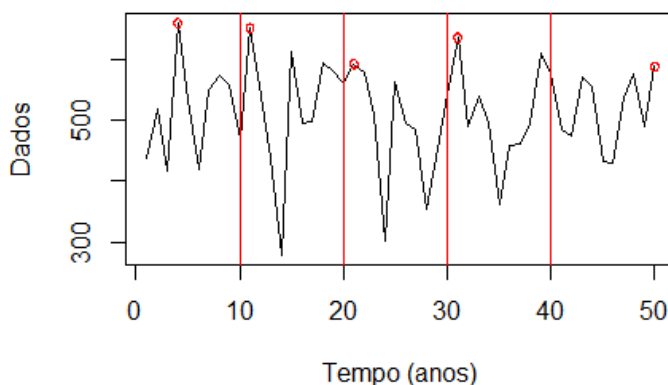
Assim, para caracterizar os eventos climáticos extremos de forma a possibilitar a compreensão de seu comportamento, bem como realizar previsões sobre sua ocorrência, a análise de valores extremos via teoria de valores extremos se mostra bastante relevante (BUTTURI-GOMES; BEIJO; AVELAR, 2019; CARVALHO *et al.*, 2013).

2.2 ANÁLISE DE VALORES EXTREMOS

Uma das formas para se obter uma série de valores extremos é por meio da extração de dados via blocos de máximos como realizado por Correia Filho, Alves e dos Santos (2017), Tavares e Santos (2022). Segundo Coles (2001), para a obtenção de uma série via blocos de máximos, primeiramente, os dados são divididos em blocos, que em geral são períodos de tempo, e, posteriormente, são extraídos os valores máximos de cada bloco, sendo estes utilizados para compor a série de extremos.

A série de máximos via blocos pode ser obtida pela seleção apenas do maior valor de cada variável aleatória (ou bloco), ou também pela seleção dos k maiores valores de cada bloco, sendo k a ordem da série (DA SILVA; DO NASCIMENTO, 2019). Então uma série de extremos pode ser descrita por $x_i^k, i = 1, \dots, B$ em que B é a quantidade de blocos analisados. A Figura 1 exemplifica a forma de extração de dados via blocos de máximos para uma série de dados qualquer composta por cinquenta observações, em que cada bloco refere-se a um período de 10 anos e, portanto, $B = 5$. Neste cenário hipotético, a série de máximos foi construída apenas pela extração do maior valor de cada bloco ($k=1$), sendo estes representados pelos círculos vermelhos.

Figura 1 - Exemplo da obtenção da série de máximos de ordem 1 para uma base de dados composta por cinquenta observações, considerando blocos de dez anos



Fonte: Elaboração própria

Obtida a série de máximos x_i^k , ou série de extremos, é possível encontrar a sua função de distribuição $G(x)$. O interesse em encontrar a distribuição dos eventos extremos é que seu conhecimento possibilita que sejam estimadas as suas probabilidades de ocorrência, o que torna possível a quantificação e gerenciamento dos riscos em atividades suscetíveis aos seus impactos (DE OLIVEIRA; CARVALHO, 2020). Estes eventos extremos merecem atenção nos processos de gestão de riscos uma vez que, conforme Rocco (2014), a despeito de apresentarem baixa probabilidade de ocorrência (baixa frequência), incorrem em grandes impactos (alta severidade).

Assim, para pesquisas onde a intenção é modelar valores extremos utiliza-se os preceitos da teoria de valores extremos (AGUIRRE; NOGUEIRA; BEIJO, 2021; BAUTISTA; ZOCCHI; ANGELOCCI, 2004; ALMEIDA, 2018; BEIJO, 2006; COLES, 2001). Dentro da Teoria de Valores Extremos (TVE), uma das distribuições mais utilizadas neste tipo de modelagem é a distribuição Generalizada de Valores Extremos (GEV), introduzida por Jenkinson em 1955. A introdução da GEV simplifica a análise de valores extremos, uma vez que garante que qualquer série de observações obtida via blocos de máximos converge para a distribuição GEV.

Esta distribuição é especificada por três parâmetros: μ que é o parâmetro de posição, σ que é o parâmetro de escala e ε que é o parâmetro de forma da distribuição. Na equação (1) está descrita a função de distribuição GEV, conforme Jenkinson (1955):

$$G(x|\mu, \sigma, \varepsilon) = \exp\left\{-\left[1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right\}, \quad (1)$$

em que $\mu \in \mathbb{R}$, $\sigma > 0$ e $\varepsilon \in \mathbb{R}$ e é definida em $\{x: 1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right) > 0\}$.

A função de verossimilhança da distribuição GEV é descrita pela equação:

$$L(\mu, \sigma, \varepsilon | x) = \frac{1}{\sigma^I} \prod_{i=1}^I \left\{ \left[1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\left(\frac{1+\varepsilon}{\varepsilon}\right)} \right\} \exp\left\{ \sum_{i=1}^I \left\{ -\left[1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}} \right\} \right\}, \quad (8)$$

em que x representa os valores da série de máximos, I é o tamanho da amostra e $\varepsilon \neq 0$.

A partir de $G(x)$ torna-se possível a obtenção da probabilidade de excedência de eventos máximos superiores a determinado quantil x da variável extrema estudada, sendo que para valores de $\varepsilon \neq 0$ esta probabilidade é obtida por meio da equação (2):

$$P(X > x) = 1 - G(x) = 1 - \exp\left\{-\left[1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right)\right]^{-\frac{1}{\varepsilon}}\right\}, \quad (2)$$

definida em $1 + \varepsilon\left(\frac{x-\mu}{\sigma}\right) > 0$ para $\varepsilon \neq 0$.

Quando $\varepsilon = 0$, a probabilidade de excedência de determinados eventos extremos é obtida por meio da equação (3):

$$P(X > x) = 1 - \exp\left\{-\exp\left(-\frac{x-\mu}{\sigma}\right)\right\}, \quad (3)$$

em que $-\infty < x < \infty$.

Ademais, ressalta-se que antes de empregar a distribuição GEV na modelagem de uma série de extremos é necessário verificar algumas características da base de dados. A primeira característica a ser investigada é a independência dos valores na

série de máximos, uma vez que a formulação da convergência da série para a distribuição GEV depende deste pressuposto (JENKINSON, 1955). A segunda característica a ser testada é a presença de tendência na série, uma vez que, caso sejam detectadas evidências de tendência, esta deve ser incorporada na modelagem (AGUIRRE, NOGUEIRA; BEIJO, 2020).

O teste de Ljung-Box, desenvolvido por Ljung e Box em 1978, é comumente empregado para avaliar a independência de uma série de dados. Trata-se de um teste de hipóteses no qual a hipótese nula (H_0) afirma que os dados da série são independentes e a hipótese alternativa (H_1) afirma que a série de dados não é independente.

Para a avaliação de tendência em uma série de dados, um teste empregado é o de Mann-Kendall. Este teste foi desenvolvido por Mann (1945) como um teste não paramétrico usado na detecção de tendência em séries de dados, sendo complementado por Kendall (1975), que forneceu a distribuição estatística do teste, possibilitando a verificação de existência de tendências não lineares e pontos de mudança. Por este teste, a hipótese nula (H_0) afirma que a série de dados é estacionária e a hipótese alternativa (H_1) afirma que a série é não estacionária.

Segundo Coles (2001) existem inúmeras possibilidades de aplicações para a teoria de valores extremos, que variam desde seu uso como ferramenta de gestão de riscos na indústria seguradora ou no mercado financeiro, como também na previsão de fenômenos naturais extremos a exemplo de precipitações, veranicos, temperaturas, terremotos, entre outras.

Destarte, diversos autores desenvolveram pesquisas para modelar fenômenos naturais extremos via Teoria de Valores Extremos. Aguirre, Nogueira e Beijo (2021), por exemplo, empregaram a TVE para modelar as temperaturas máximas mensais na cidade de Piracicaba-SP. Para isso, os autores utilizaram a distribuição GEV, que teve seus parâmetros estimados via inferência Bayesiana. Como resultado os autores especificaram os modelos mais adequados para modelar a temperatura máxima de cada mês na cidade, apresentando achados que podem auxiliar a realização do planejamento e gestão de atividades que são sensíveis às temperaturas máximas extremas na cidade.

Martins *et al.* (2018) atuaram na previsão de precipitações extremas, buscando modelar as precipitações máximas da cidade de Silvianópolis-MG. Os autores empregaram a distribuição GEV para modelar o evento extremo e empregaram a

inferência Bayesiana para estimar os parâmetros desta distribuição, testando diferentes estruturas de distribuição *à priori*, e também o método da máxima verossimilhança. Os resultados apontaram que as estimativas realizadas via inferência Bayesiana foram melhores do que os resultados encontrados pelo estimador de máxima verossimilhança. O bom desempenho do modelo fez os autores sugerirem seu uso para o planejamento de construções hidráulicas e atividades agrícolas em Silvianópolis-MG.

Carvalho *et al.* (2013) empregaram e compararam as distribuições GEV, lognormal, Pareto e Gama, com parâmetros estimados via método da máxima verossimilhança, para modelar veranicos extremos nas regiões Centro-Oeste, sudeste e Sul do Brasil. O objetivo era gerar informações sobre os tempos de retorno em que a ocorrência deste fenômeno tende a ser mais severa nestas regiões. Como resultado os autores verificaram que a distribuição GEV foi a que apresentou melhores resultados na modelagem dos veranicos extremos, encontrando que a região Sudeste é a mais vulnerável à ocorrência de veranicos extremos em diferentes tempos de retorno.

Ademais, o trabalho de Butturi-Gomes, Beijo e Avelar (2018) intentou avaliar o desempenho preditivo da distribuição GEV, estimada via inferência Bayesiana, em relação aos veranicos extremos. Para isto, os pesquisadores realizaram um estudo de simulação dos veranicos extremos no qual avaliaram o desempenho da GEV para cenários variados de tamanho amostral, intensidade de veranicos e densidades *a priori*. Como resultado, os autores pontuaram que a distribuição GEV apresentou desempenho satisfatório na previsão de veranicos máximos, para diferentes tamanhos de amostra e cenários avaliados. Ou seja, nesta pesquisa, os autores afirmaram a adequação desta distribuição para a predição de eventos climáticos extremos, mesmo em casos de eventos de natureza discreta, como são os veranicos.

Assim, dentre as diversas possibilidades de aplicações que a modelagem de eventos extremos proporciona, uma delas é ser empregada como instrumento de auxílio à previsão de eventos que possam impactar determinada atividade ou ramo de interesse. Destarte, nesta pesquisa este instrumento foi utilizado em conjunto a um modelo de regressão logística.

2.3 REGRESSÃO LOGÍSTICA

De acordo com Barreto (2011) modelos de regressão consistem em técnicas empregadas tanto para a realização de previsões quanto para a classificação de dados, sendo um método que avalia a variação observada em uma variável, chamada de variável dependente ou variável resposta, quando outras variáveis, chamadas de variáveis independentes ou explicativas, se alteram.

Em modelos de regressão, conforme Dobson (2002), considerando que $\mathbf{x} = (x_1, \dots, x_j)$ é o vetor formado pelas variáveis explicativas do modelo que são relacionadas à variável resposta Y , assume-se que a média da variável resposta $E(Y)$ é relacionada a \mathbf{x} por meio de uma função de ligação $g(\cdot)$. Esta relação é definida por:

$$g(E[Y]) = \eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j = \mathbf{x}^T \boldsymbol{\beta},$$

em que η é a função de ligação do modelo de regressão, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_j) \in \mathbb{R}$ é um vetor de coeficientes regressores de dimensão j .

Dentre os diversos modelos de regressão existentes, a regressão logística é empregada quando a variável resposta (Y) é binária, ou seja, a variável assume apenas dois valores, sendo “0” para eventos de “fracasso” e “1” para eventos de “sucesso”. Assim, considera-se que:

$$Y \sim \text{Bernoulli}(p),$$

em que $\mathbf{Y} = Y_n$ é o vetor formado pelas variáveis aleatórias independentes que representam a variável resposta do modelo, sendo $n = 1, \dots, j$ e $j \rightarrow \infty$, e “ p ” representa a probabilidade desconhecida de sucesso do evento de interesse. Assim, $p(Y = 1) = p$ e $p(Y = 0) = 1 - p$. A distribuição *Bernoulli*(p_i) é dada por:

$$p(Y_i = y) = p_i^y (1 - p_i)^{1-y},$$

em que y representa o valor observado da variável aleatória Y , sendo $y = 0$ ou $y = 1$.

Variáveis aleatórias independentes com distribuição *Bernoulli*(p_i) tem como função de verossimilhança o produto de suas funções de probabilidade. A função de verossimilhança da distribuição *Binomial*(n, p) é descrita por:

$$L(\beta_0, \beta_1 | \omega) = \prod_{i=1}^n \left[\frac{1}{1 + \exp^{-(\beta_0 + \beta_1 x_i)}} \right]^{\omega_i} \cdot \left[\frac{\exp^{-(\beta_0 + \beta_1 x_i)}}{1 + \exp^{-(\beta_0 + \beta_1 x_i)}} \right]^{1 - \omega_i},$$

em que n é o tamanho da série de dados.

Como a variável resposta do modelo logístico está limitada em 0 ou 1, o modelo logístico ajustado prevê apenas valores que estejam neste intervalo, fazendo com que a regressão logística seja bastante útil na previsão de probabilidades de eventos (FERNANDES *et al.*, 2020; PAMPEL, 2000). Conforme Domínguez-Almendros, Benítez-Parejo e Gonzalez-Ramirez (2011), no modelo de regressão logística a função de ligação η é obtida por meio da função *logit*, sendo $\eta = \log\left(\frac{p}{1-p}\right) = \mathbf{x}^T \boldsymbol{\beta}$. O modelo logístico simples, o qual contém apenas uma variável explicativa, está representado na equação (4):

$$\text{logit}(E[Y]) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x, \quad (4)$$

em que x é a variável explicativa, β_0 e β_1 são os coeficientes do modelo ajustado, sendo $\beta_0 \in \mathbb{R}$, $\beta_1 \in \mathbb{R}$ e p é a probabilidade $p(Y = 1 | x)$, em que $0 \leq p \leq 1$.

No modelo logístico, a interpretação dos parâmetros β_0 e β_1 não é direta, representando, respectivamente, o valor do *logit*(Y) quando $x = 0$ e o valor do *logit*(Y) quando x varia em uma unidade. Para que a interpretação do parâmetro β_0 seja mais inteligível deve-se interpretar o valor de $\exp\{\beta_0\}$. Este resultado indica qual o valor esperado da variável resposta quando a variável regressora é zero, sendo que no modelo logístico simples $1 - \exp\{\beta_0\}$ indicará, por complementaridade, o quanto da variação da variável resposta é explicada pela variável regressora (FERNANDES *et al.*, 2021).

Conforme Cheng, Varshney e Arora (2006), o modelo de regressão logística fornece as probabilidades de “sucesso” de um evento com base em determinadas características de entrada x e, a partir destas probabilidades, é possível classificar

eventos nas classes de interesse. A partir da equação (4), tem-se que a obtenção de $p(Y = 1 | x)$ por meio do modelo de regressão logística é realizada, conforme Pampel (2000), por meio da equação (5):

$$p(Y = 1 | x) = p = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 x)}} \cdot \quad (5)$$

Para realizar tarefas de classificação através das probabilidades $p(Y = 1 | x)$ fornecidas pelo modelo ajustado, basta que se estabeleça um limiar de corte (q) que define a partir de qual valor de probabilidade o evento será classificado como sucesso “1” ou fracasso “0”. Ou seja, a partir do limiar q , que é definido no intervalo de Y , sendo $0 \leq q \leq 1$, define-se a classificação de quaisquer eventos Y como:

$$Y = \begin{cases} 0 & ; p(Y = 1|x) < q \\ 1 & ; p(Y = 1|x) \geq q \end{cases}$$

O limiar mais eficiente é aquele que consegue classificar com maior assertividade as classes dos eventos 0 e 1. Diferentes métricas como acurácia, viés, sensibilidade e especificidade podem ser empregadas no processo de tomada de decisão, visando a especificação do limiar mais adequado para cada problema (Figueiredo *et al.*, 2018).

Para estimar os parâmetros do modelo de regressão logística (β_0, β_1) , bem como para estimar os parâmetros da distribuição GEV $(\mu, \sigma, \varepsilon)$ descrita na seção 2.2, é necessário o uso de alguma estratégia de inferência estatística. Em decorrência de suas características e vantagens, a abordagem inferencial bayesiana é bastante empregada no processo de inferência.

2.4 INFERÊNCIA BAYESIANA

De acordo com Gelman *et al.* (2003, p. 1) “A inferência Bayesiana é o processo de ajustar um modelo de probabilidade a um conjunto de dados e descrever o resultado por uma distribuição de probabilidade nos parâmetros do modelo e em quantidades não observadas como previsões para novas observações”. Conforme Wakefield (2013), na abordagem inferencial bayesiana os parâmetros desconhecidos

são tratados como variáveis aleatórias e, dessa forma, podem ser descritos por uma distribuição.

A inferência Bayesiana tem como alicerce o teorema de Bayes. Considere que o vetor θ representa todos os parâmetros desconhecidos de uma distribuição e que \mathbf{y} representa o vetor de dados observados. O teorema de Bayes fornece a distribuição a *posteriori* de θ , denotada por $p(\theta | \mathbf{y})$, após a observação de \mathbf{y} , de forma que:

$$p(\theta | \mathbf{y}) = \frac{L(\theta | \mathbf{y})p(\theta)d\theta}{\int L(\theta | \mathbf{y})p(\theta)d\theta},$$

em que $L(\theta | \mathbf{y})$ é a função de verossimilhança que agrega todas as informações disponibilizadas pelos dados observados, $p(\theta)$ é a distribuição a *priori* conjunta dos parâmetros que agrega toda a informação prévia que o pesquisador dispõe a respeito de θ antes da observação de \mathbf{y} e $\int L(\theta | \mathbf{y})p(\theta)d\theta$ é uma constante normalizadora empregada no teorema para garantir que a integral da razão tenha soma igual a um no espaço paramétrico.

Como o denominador da equação da *posteriori* é uma constante que não depende de θ que é a variável para a qual se intenta especificar a distribuição, conforme Wakefield (2013), é possível reescrever a equação da *posteriori*, analogamente ao Teorema de Bayes, como:

$$p(\theta | \mathbf{y}) \propto L(\theta | \mathbf{y})p(\theta)$$

em que α representa “proporcionalidade”, sendo a distribuição a *posteriori* dos parâmetros resultado da distribuição a *priori* atualizada pela informação advinda dos dados, que é fornecida pela função de verossimilhança.

Conforme Paulino, Turkman e Murteira (2003), a função de verossimilhança contém toda a informação presente na amostra ou experimento. Assim, considerada uma amostra aleatória $y = y_1, \dots, y_n$, $n \in \mathbb{N}$, $n \rightarrow \infty$, a função de verossimilhança para θ , conforme Roussas (2003) é expressa por:

$$L(\theta | \mathbf{y}) = \prod_{j=1}^n f(y_j, \theta),$$

em que $\theta \in \Theta$, sendo Θ o espaço paramétrico que representa o conjunto de todos os valores possíveis para θ , dada a distribuição assumida para a variável.

Em relação à distribuição a *priori* considerada para θ , Wakefield (2013) coloca que a sua escolha é crucial para o bom desempenho do processo de estimação bayesiano. De acordo com Box e Tiao (1973) a distribuição a *priori* representa todo o conhecimento que se tem a respeito de θ antes da obtenção de y e este conhecimento pode indicar a existência informações consideráveis para o processo de estimação ou mesmo ser expressa como uma ignorância relativa a respeito do comportamento de θ .

De acordo com Paulino, Turkman e Murteira (2003) a distribuição a *priori* informativa é aquela fornecida por especialistas da área de interesse e que traz informações relevantes para a especificação da distribuição do parâmetro. Box e Tiao (1973) ressaltam que a incorporação de informações a *priori* pode trazer estimativas mais assertivas para os parâmetros, mas adverte que informações equivocadas podem distorcer os resultados, aconselhando cautela no processo de definição, principalmente nos casos em que o tamanho da amostra é pequena, pois nestes a distribuição a *priori* tem maior influência sobre os resultados.

Quando há ignorância sobre θ e se conhece pouco ou nada a respeito do comportamento da distribuição a *priori* ela pode ser considerada como não informativa (PAULINO; TURKMAN; MURTEIRA, 2003). Wakefield (2014) define que distribuição a *priori* não informativa é aquela na qual a informação disponível tem o mínimo impacto sobre a distribuição a *posteriori*, prevalecendo o resultado obtido com a função de verossimilhança.

Entretanto, para que seja viabilizada a utilização da inferência Bayesiana, nos casos de falta de informação prévia dos parâmetros, é possível, de acordo com Box e Tiao (1973), ao considerar cautelosamente uma métrica apropriada, representar um estado de ignorância sobre θ por meio de uma aproximação adequada, considerando que sobre a região onde a verossimilhança é apreciável a distribuição *priori* seja constante.

Outra forma de definir uma distribuição a *priori* para os parâmetros nos casos em que pouca informação está disponível é especificar uma distribuição a *priori* vaga para θ . Neste caso, define-se a família da distribuição a *priori*, mas por falta de informações suficientes para especificá-la define-se distribuições a *priori* com variância elevada, tendendo a infinito, o que faz a ignorância sobre os parâmetros ser

representada na densidade da distribuição a *priori* (JEFFREYS, 1961 & GIANOLA; FERNANDO, 1986). Ressalta-se que os parâmetros das distribuições a *priori* são nomeados de hiperparâmetros, como forma de diferenciação em relação aos parâmetros estimados pela distribuição a *posteriori* (GELMAN *et al.*, 2003).

Nesse contexto, a opção pelo uso da inferência Bayesiana mesmo nos casos em que há escassez de informações a *priori* advém das vantagens oriundas desta abordagem. Coles (2001) destaca três vantagens do uso desta abordagem inferencial na análise de valores extremos. A primeira vantagem é que esta abordagem inferencial considera informações a *priori* para os parâmetros, o que diminui a incerteza dos resultados em casos de amostras menores, oriundas de dados escassos. A segunda vantagem colocada pelo autor é que a distribuição a *posteriori* dos parâmetros tem uma completude maior dos procedimentos de inferência do que os métodos clássicos. A terceira vantagem é a não dependência do procedimento a determinadas condições de regularidade como a unicidade, consistência e a normalidade assintótica dos estimadores, como acontece em métodos inferenciais clássicos.

Gelman *et al.* (2003) pontua que a lógica bayesiana facilita a compreensão dos resultados para o senso comum, uma vez que ao especificar uma distribuição para os parâmetros, torna possível que se interprete facilmente seus intervalos de probabilidade de ocorrência, também chamados de intervalos de credibilidade.

2.4.1 Estimação Bayesiana por meio do algoritmo MCMC

Entretanto, a despeito das suas vantagens, a inferência bayesiana que teve seus precursores na década de 1930 ficou por décadas estagnada. A estagnação ocorreu uma vez que o processo para a especificação das distribuições a *posteriori* para os parâmetros bayesianos não é, na maioria das vezes, uma tarefa fácil. Isto porque quando a distribuição a *posteriori* conjunta para os parâmetros ($p(\theta | x)$) é encontrada, para definir-se a distribuição marginal para cada um deles é necessário integrar $p(\theta | x)$, o que nem sempre é possível de ser realizado de forma analítica (GELMAN *et al.*, 2003).

Seja θ um vetor de parâmetros, $\mathbf{y} = [y_j]$, $j \in \mathbb{N}$ e $j \rightarrow \infty$, o vetor dos dados observados e $p(\theta | \mathbf{y})$ a distribuição a *posteriori* conjunta para θ . Conforme Paulino, Turkman e Murteira (2003), para encontrar a distribuição marginal para qualquer

elemento de θ , por exemplo, θ_1 , basta integrar $p(\theta|\mathbf{y})$ em relação a todos os outros elementos, de modo que:

$$p(\theta_1|\mathbf{y}) = \int_{\theta \neq \theta_1} p(\theta|\mathbf{y}) d\theta_{\theta \neq \theta_1} .$$

O trabalho de Gelfand *et al.* (1990) trouxe uma solução alternativa para o problema dos cálculos das integrais por meio do uso do algoritmo do Amostrador de Gibbs, elaborado por Geman e Geman (1984), que se apoia na teoria das Cadeias de Markov. Após o encontro de bons resultados este algoritmo passou a ter seu uso disseminado em análises bayesianas, uma vez que permitia a definição das marginais para os mais diversos tipos de parâmetros nas situações em que a forma algébrica não era viável (NOGUEIRA; SÁFADI; FERREIRA, 2004). Além disso, o desenvolvimento tecnológico computacional nas últimas décadas possibilitou o aumento significativo de sua implementação.

Gelman *et al.* (2003) pontua que a simulação das Cadeias de Markov, também chamada de Cadeias de Markov Monte Carlo (MCMC), é uma ferramenta na qual realiza-se a simulação de valores de θ para distribuições aproximadas de θ e, posteriormente, a partir da aceitação ou rejeição dos valores simulados, a distribuição aproximada vai sendo corrigida para que represente da melhor forma possível a distribuição a *posteriori* de interesse. Neste método, a amostragem é realizada de forma sequencial com a distribuição dos sorteios amostrados sendo definida em função do último valor amostrado em que os sorteios originam uma cadeia de Markov.

De acordo com teoria da probabilidade “uma cadeia de Markov é uma sequência de variáveis aleatórias $\theta = [\theta_1, \dots, \theta_q]$, para as quais, para qualquer t , $t = 1, \dots, q$, a distribuição de θ^t dados todos os θ anteriores dependa apenas do valor mais recente, θ^{t-1} ” (GELMAN *et al.*, p. 275, 2003). No MCMC a ideia é que as distribuições aproximadas melhoram a cada amostragem, tendendo a convergir para a distribuição a *posteriori* almejada.

Após a obtenção das cadeias a *posteriori* dos parâmetros via MCMC, é necessário avaliar se as cadeias geradas convergiram para as distribuições marginais de interesse. Segundo Dos Reis (2020), as cadeias a *posteriori* para os parâmetros convergem para a distribuição de interesse quando se tornam estacionárias, pois isso indica que a distribuição a *posteriori* completa está suficientemente próxima das

distribuições marginais encontradas. Assim, existem alguns critérios que podem ser utilizados na avaliação da convergência das cadeias, dentre eles: 1) Critério de Geweke (1992); 2) Critério de Raftery e Lewis (1992); e 3) Critério de Heidelberger e Welch (1983).

Segundo Nogueira, Sáfyadi e Ferreira (2004), o critério de Geweke avalia a convergência das cadeias *a posteriori*, considerando que a cadeia converge quando a diferença padronizada entre as suas médias não for grande. Este critério é avaliado através do módulo da estatística de Geweke ($|Gw|$). O critério de Raftery & Lewis estima quantas iterações se fazem necessárias para que o amostrador de Gibbs apresente convergência para a distribuição estacionária. Este critério é avaliado por meio do fator de dependência Raftery & Lewis. O critério de Heidelberger & Welch avalia a estacionariedade da cadeia dos parâmetros e também se o tamanho da amostra é suficiente para gerar estimativas sobre a média dos parâmetros com acurácia, sendo este critério avaliado por meio do valor-p de Heidelberger & Welch.

Verificada a convergência das cadeias se tem confiança de que existe uma amostra realizada da distribuição marginal de interesse. Para avaliar o comportamento do parâmetro pode-se realizar a estimação pontual ou intervalar e assim caracterizar o comportamento paramétrico. A estimação pontual pode ser uma estatística descritiva da distribuição dos parâmetros como a média ou mediana, por exemplo. Já a estimação intervalar fornece a informação sobre o intervalo de credibilidade dos parâmetros para um nível $1-\alpha$ de credibilidade. Neste trabalho considerou-se o intervalo de credibilidade denominado de HPD (*Highest Posterior Density*) que, dentre todos os possíveis intervalos para o mesmo $1-\alpha$, apresenta probabilidade de ocorrência máxima para a distribuição *a posteriori*, representando a região com menor amplitude possível para uma massa de probabilidade fixa (PAULINO; TURKMAN; MURTEIRA, 2003).

2.4.2 Teste de hipóteses Bayesiano

Pode ocorrer que o interesse na definição das distribuições *a posteriori* para os parâmetros de uma distribuição seja verificar determinadas suposições a respeito destes. Isto pode ser realizado por meio de testes de hipóteses. De acordo com Paulino, Turkman e Murteira (2003), o teste de hipóteses bayesiano é simples, uma vez que o conhecimento da distribuição de θ faz com que seja realizada a

interpretação probabilística direta das hipóteses $H_0: \theta \in \theta_0$ e $H_1: \theta \in \theta - \theta_0$, em que H_0 é denominada hipótese nula, H_1 é denominada hipótese alternativa e θ_0 é o espaço paramétrico sob o qual é estruturada H_0 e θ corresponde ao espaço paramétrico que o conjunto de todos os valores possíveis para θ .

Desse modo, uma das formas de se realizar o teste de hipóteses via estatística bayesiana é por meio do Fator de Bayes (WAKEFIELD, 2003). Conforme Gelman *et al.* (2003), o fator de Bayes pode ser descrito pela razão da densidade de probabilidade marginal em um modelo ($P(\theta \in \theta_0)$) com a densidade marginal em um segundo modelo ($P(\theta \in \theta - \theta_0)$).

Desenvolvido por Jeffreys (1961) o fator de bayes contrário à hipótese nula, ou a favor da hipótese alternativa, pode ser descrito como:

$$FB_{1,0} = \frac{O(1,0 | x)}{O(1,0)},$$

em que $O(1,0 | x) = \frac{P(H_1 | x)}{P(H_0 | x)}$ representa a chance a *posteriori* a favor de H_1 e $O(1,0) = \frac{P(H_1)}{P(H_0)}$ representa a chance a *priori* a favor de H_1 , sendo que basta realizar $FB_{1,0} = |FB_{0,1}|^{-1}$, caso o interesse seja testar as evidências em favor de H_0 .

Desse modo, conforme Paulino, Turkman e Murteira (p. 149, 2003), verifica-se que “o fator de Bayes representa um peso relativo da evidência contida nos dados a favor de uma ou outra hipóteses em confronto, sendo visualizável como o peso da evidência a posteriori, descontado do componente do peso da evidência a *priori*”.

O valor encontrado para FB é denominado por Good (1950) como “peso de evidência” e a interpretação deste peso levará à decisão de aceitar ou rejeitar a hipótese de interesse. Jeffreys (1998) propõe critérios de interpretação para os pesos de evidências em favor da hipótese testada de modo que se:

- $FB < 1$, existem evidências contrárias sobre a hipótese testada;
- $1 \leq FB < 3$, existem fracas evidências sobre a hipótese testada;
- $3 \leq FB < 10$, existem evidências substanciais sobre a hipótese testada;
- $10 \leq FB < 30$, existem evidências fortes sobre a hipótese testada;
- $30 \leq FB < 100$, existem evidências muito fortes sobre a hipótese testada; e
- $FB \geq 100$, existem evidências decisivas sobre a hipótese testada.

Apresentada a abordagem inferencial a ser empregada nesta pesquisa tanto para a o ajuste da distribuição GEV, como para o ajuste dos modelos logísticos, a próxima seção explana sobre as características da planta de café, que é uma planta que tem características que podem ser enquadradas na estruturação de um contrato seguro paramétrico.

2.5 CULTURA DE CAFÉ

No que tange a produção cafeeira mundial, o Brasil tem papel de destaque como o país que mais produz e exporta esse grão no mundo (CONAB, 2021). Isso faz com que a cultura cafeeira se configure como relevante alicerce econômico para o país, bem como fator de geração de renda para milhares de brasileiros em centenas de municípios, que atuam de forma direta ou indireta nesta atividade (CEPEA, 2021).

Em decorrência dessa relevância existem preocupações associadas às variações climáticas que não são controláveis pelos produtores e que podem afetar o bom desempenho da produção cafeeira. Segundo Ozaki (2007), para o desenvolvimento pleno de uma cultura são necessárias combinações de fatores pluviométricos e de temperatura que, caso sejam adversos, podem levar a perdas na produção e, conseqüentemente, a grandes prejuízos econômicos que podem extrapolar as cercas das fazendas e se multiplicar por toda a economia.

A combinação adequada dos fatores meteorológicos para que a safra alcance os níveis de produção desejados, varia a cada fase fenológica do cafeeiro (CAMARGO; CAMARGO, 2001; CAMARGO, 2010). Desse modo, torna-se necessário o conhecimento sobre a fenologia do cafeeiro para que se possa quantificar o impacto das condições climáticas observadas em cada fase fenológica. De acordo com Camargo e Camargo (2001), o ciclo fenológico completo do café arábica tem duração de dois anos e se divide em seis ciclos fenológicos, apresentados no Quadro 2.

Quadro 2 - Esquematização das seis fases fenológicas do cafeeiro arábica, durante 24 meses, nas condições climáticas tropicais do Brasil

	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago
	1ª Fase (V1)						2ª Fase (V2)					
1º Ano												
fenológico	Vegetação e Formação das Gemas Foliaves						Indução e Maturação das Gemas Foliaves					
	3ª Fase (R1)				4ª Fase (R2)		5ª Fase (R3)		6ª Fase (R4)			
2º Ano												
fenológico	Florada, Chumbinho e Expansão dos Frutos				Granação dos Frutos		Maturação dos Frutos		Repouso e Senescência dos Ramos			
	Fases com necessidade hídrica						Fases sem necessidade hídrica					

Fonte: Adaptado de Camargo e Camargo (2001)

A esquematização do ciclo fenológico, segundo os autores, permite que, dentre outras questões, seja possível identificar quais fases fenológicas são mais dependentes de recursos hídricos e em quais o estresse hídrico causa o favorecimento da produção. Segundo Arruda *et al.* (2000, p.782) “a identificação de fatores limitantes à produção e de épocas críticas de ocorrência de deficiência hídrica durante o ciclo fenológico do cafeeiro são metas importantes no planejamento e manejo da cultura”. Apesar de a disponibilidade hídrica durante o desenvolvimento da safra não ser o único fator de influência na produção, esta se configura como principal fator que afeta a produtividade do cafeeiro (PICINI *et al.*, 1999; MARTINS *et al.*, 2015; APARECIDO; ROLIM; SILVA, 2015).

Como apresentado no Quadro 1, no primeiro ano fenológico do cafeeiro (A1) ocorrem duas fases vegetativas. A primeira fase vegetativa (V1) acontece entre os meses de setembro e março. Nesta fase ocorre a vegetação e a formação das gemas foliaves da planta. A seca neste período prejudica a formação das gemas e a produção do ano seguinte sendo, portanto, um período que exige maior disponibilidade hídrica. A segunda fase vegetativa (V2) ocorre entre os meses de abril e agosto. Nesta fase, ocorre a indução e a maturação das gemas foliaves. Aqui a planta não apresenta grande necessidade hídrica, sendo que a ocorrência do estresse hídrico nesta fase pode inclusive trazer benefícios para a produção resultante (EMBRAPA, 2009; RUIZ-CÁRDENAS *et al.*, 2015).

O segundo ano fenológico possui quatro fases reprodutivas. Na primeira fase reprodutiva (R1), que acontece entre os meses de setembro e outubro, ocorre a florada do café, formação dos chumbinhos e expansão/crescimento dos frutos. Esta fase depende expressivamente de um aumento do potencial hídrico da planta, uma vez que estresse hídrico neste momento prejudica o crescimento dos frutos. A segunda fase reprodutiva (R2) que ocorre entre janeiro e março, é a fase de granação dos frutos. Esta fase também tem alta exigência hídrica, sendo que nela estiagens severas prejudicam a qualidade dos grãos, incorrendo em chochamento. (EMBRAPA, 2009; CAMARGO; CAMARGO, 2001; RUIZ-CÁRDENAS *et al.*, 2015).

A terceira fase reprodutiva do cafeeiro (R3) ocorre nos meses de abril a junho, fase em que ocorre a maturação dos frutos e, conseqüentemente, a finalização do período de produção. Nesta fase fenológica, a deficiência hídrica é benéfica e favorece a qualidade do produto final. Por fim, a quarta e última fase reprodutiva (R4) ocorre nos meses de julho e agosto e refere-se ao período de repouso e senescência dos ramos, conhecido como período de autopoda, no qual os ramos produtivos não-primários secam e morrem. Excesso de umidade nesta fase também é prejudicial ao cafeeiro (CAMARGO; CAMARGO, 2001; RUIZ-CÁRDENAS *et al.*, 2015).

Desse modo, verifica-se que a necessidade hídrica para o favorecimento da produção de uma safra de café ocorre nos ciclos V1, R1 e R2. Portanto, é coerente que produtos de seguro paramétrico que visem proteção das lavouras contra veranicos, principalmente veranicos extremos, sejam criados visando a cobertura para estes períodos. Segundo Ceballos, Martínez-Fernández e Luengo-Ugidos (2004), considera-se como veranico uma sequência de n dias consecutivos sem a observação de uma quantidade relevante de precipitação. Assim, “veranicos extremos” correspondem às maiores sequências de n dias sem precipitações relevantes em um determinado período.

Alguns autores pesquisaram a relação entre a frequência de chuvas em diferentes ciclos fenológicos e a produção da safra resultante. Arruda *et al.* (2000), por exemplo, por meio da análise de correlação entre alguns fatores meteorológicos e a produção de café resultante, chegou a resultados significativos ao nível de 5%, atestando que a frequência de chuvas é o fator com maior relevância para explicar a produção de café na cidade de Pindorama-SP. Os autores verificaram que existia uma correlação positiva entre a frequência de chuvas nas fases V1 e R2 e a produção observada, demonstrando que a produção mais alta coincidiu com chuvas mais

frequentes nesses períodos. Ademais, contrariando o proposto por EMBRAPA (2009) e Ruiz-Cárdenas *et al.* (2015) os autores também verificaram uma correlação negativa entre a frequência de chuvas e o período de florada e formação de chumbinhos (R1), demonstrando que naquela cidade chuvas mais frequentes em R1 coincidiram com a diminuição da produção observada.

Soares *et al.* (2021) também avaliaram o impacto do clima na produção resultante de café. Na pesquisa, os autores utilizaram de modelos lineares generalizados para verificar a influência de diversos fatores climáticos na produção resultante de café em algumas cidades do Sul e do Alto Paranaíba de Minas Gerais. Como resultado, em relação ao déficit hídrico foi constatado que o volume de chuvas na fase R2 se relaciona positivamente com a produção de uma safra, sendo discutido que a baixa intensidade de chuvas nessa fase tende a prejudicar a produtividade e a qualidade do café. Os autores não encontraram relação entre chuvas e produção nas outras fases fenológicas.

Júlio *et al.* (2016) criaram um índice padronizado de precipitação para identificar o histórico de períodos secos e chuvosos em algumas cidades das Matas de Minas nas fases de maior exigência hídrica para o café, a saber os ciclos V1, R1 e R2, e, posteriormente, confrontou-se este índice com a produção de café obtida nas cidades, entre os anos de 2000 e 2012. Como resultado os autores encontraram que a ocorrência de períodos de seca e períodos chuvosos nas fases críticas do café apresentaram estreita relação com a produção resultante, sendo que para anos em que foi observada seca nestas fases observou-se também redução da produção final.

Tendo em vista o impacto que os veranicos extremos em determinadas fases fenológicas podem ocasionar na produção resultante de uma safra de café, tem-se que uma estratégia de gestão de riscos em relação à sua ocorrência é a contratação de contratos de seguro paramétrico pelo produtor. Desse modo, o artigo a seguir explicitou uma estratégia metodológica para a elaboração de contratos de seguros paramétricos agrícolas para a proteção das lavouras contra a ocorrência de eventos climáticos extremos. Além disso, foi realizado um estudo de caso com a aplicação do método proposto para na estruturação de um contrato de seguro paramétrico contra veranicos extremos na primeira fase reprodutiva do cafeeiro nas cidades de Guaxupé, Monte Santo de Minas e Monte Carmelo, sendo todas elas produtoras de café no Estado de Minas Gerais.

ARTICLE TYPE

Bayesian modeling for agricultural parametric insurance for extreme weather events

Karoline Pereira Branco,^{*†} Luiz Alberto Beijo,[‡] and Reinaldo Antônio Gomes Marques[†][†]Actuarial Risk Lab, Federal University of Alfenas, Minas Gerais, Brazil[‡]Statistical Department, Federal University of Alfenas, Minas Gerais, Brazil^{*}Corresponding author. Email: karoline.branco@sou.unifal-mg.edu.br

Abstract

This research aimed to build an actuarial method for parametric agricultural insurance against the occurrence of extreme climate events, targeting contracts with reduced basis risk. To achieve the proposed objective, it was adjusted a loss predictor model for crops when an extreme climate event is observed. The probability of loss in a crop was used as the index of the parametric insurance contract. The indemnity trigger was defined as the probability of loss that presented the lowest cost for the insurer and the insured, as well as the one that best reconciled the occurrence of losses and indemnity payments. The proposed method applied the generalized distribution of extreme events (GEV) to model extreme climate events and the logistic model to predict losses. Both models were fitted via Bayesian inference. This inferential approach enabled the identification of uncertainties at all levels of modeling through credibility intervals. A case study for modelling a parametric contract to protect coffee trees against dry spells in the flowering period was performed. The case study demonstrated that the model performs with reasonable accuracy even in cases when data are scarce. It is concluded that the satisfactory results suggest that the proposed method can be used as a basis for modeling modelling several parametric insurance contracts related to agriculture.

Keywords: actuarial pricing, bayesian analysis, climate risk, extreme value theory, weather index insurance

1. Introduction

Agricultural development is one of the most powerful tools to end extreme poverty around the world. This is not only because, according to World Bank (2023) in 2050 agricultural production is expected to feed around 9.7 billion people, but also because growth in the agricultural sector is two to four times more effective in raising the incomes of the poorest compared to other sectors. In addition, agriculture is also central to economic growth: accounting for 4% of global gross domestic product (GDP) and can account for more than a quarter of GDP in developing countries (World Bank 2023).

Due to this great importance, there is a concern about the risks involved in agricultural activity. Climate Risk is one of the major challenges faced by farmers. It is because agricultural productivity is highly dependent on climatic conditions and these conditions are not under producers' control (Heimfarth and Musshoff 2011). Thus, there is a need to adopt strategies that provide security to investors in the sector, minimizing the impact of possible production losses caused by unfavorable weather conditions, especially for small producers who depend on this activity for their livelihood. One of the strategies for mitigating the climate risk on agricultural production is the contracting of agricultural insurance. Agricultural insurance is fundamental for the development of agribusiness, as it provides financial security to producers and, consequently, enables investments, modernization, and competitiveness of the sector (Fornazier, Souza, and Ponciano 2012), allowing the growth in agricultural productivity at a lower price.

Based on this, a new type of insurance called “parametric insurance” emerges and shows great potential for application in agricultural insurance contracts (Carter and Chiu 2018). The World Bank (2017) defines parametric insurance as a contract in which the indemnity payouts are based on a predetermined index that pointed to a loss of assets and investments that occur as a result of weather or other catastrophic events. In this type of insurance, when a predetermined trigger of the index in the contract is reached or surpassed, the payments are automatically executed (Miller 2020; Zurich 2021). Some examples of indemnity index in agricultural parametric insurance are excess of rainfall, droughts, and extreme temperatures, among other events which can harm the development of the crops (Shirsath *et al.* 2019).

The payment mechanism of parametric insurance makes it overcome traditional insurance in some aspects. This is because the parametric product has advantages such as greater agility in indemnity payments, the solution of moral hazard, and reduction of adverse selection (Skees *et al.* 2012; World Bank 2005). In addition, contracting this product tends to be viable for small producers. It is because this contract tends to have a lower acquisition cost since it does not require the insured to pay a specialist to quantify losses before paying the indemnity (Heimfarth and Musshoff 2011; Radu and Alexandru 2022).

However, despite the potential and benefits, parametric insurance contracts still have low demand and face difficulties in acceptance by insurers and insureds (World Bank 2017; Lichtenberg and Iglesias 2022). It is because parametric insurance contracts are very sensitive to the called “basis risk” (Afshar *et al.* 2021; Sum *et al.* 2020; Teh and Woolnough 2019; Lichtenberg and Iglesias 2022). The basis risk corresponds to the mismatch between the trigger and the conditions experienced on the field by the insured. This risk can cause undue payments of indemnities - positive basis risk - or the omission of indemnities when a claim occurs - negative basis risk. Both occurrences must be avoided (Afshar *et al.* 2021).

It is assumed that the development of control mechanisms for the reduction of basis risk on parametric insurance would result in better acceptance of it (Teh and Woolnough 2019). Carter and Chiu (2018) and Sengupta and Kousky (2020) state that the reduction of the basis risk is possible in well-designed contracts in which the indemnity trigger actually correlates with the losses of insureds. However, the correct compatibility between loss and indemnity payouts is not an easy task, demonstrating that “although it is true that they [parametric insurance contracts] are the simplicity personified once executed and after payment is received, this is not true for their creation” (Turner 2020, p.2).

Some authors developed different methodologies to reduce the basis risk by reconciling claims and indemnity payments. Teh and Woolnough (2019) analyzed the basis risk in contracts with binary and multivariate indexes for parametric insurance, creating a methodology capable of suggesting the best contract trigger based on utility theory assumptions. Shirsath *et al.* (2019) proposed the elaboration of a contract against climate risks in Indian crops, with a focus on Soybean and Pearl Millets. The researchers combined meteorological statistical analysis and crop growth modeling to define parametric insurance contract triggers. The results were promising in defining triggers that effectively capture the relationship between climate risk at important stages of crop growth.

Pai *et al.* (2022) proposed a Bayesian model of quantile regression combined with spatial correlation techniques for the elaboration of a parametric insurance contract against earthquakes in China. Contracts with a reasonable loss ratio and reduced base risk were obtained. Afshar *et al.* (2021) proposed the use of crop simulation models and phenology monitoring to reduce the base risk in the parametric insurance contract against the occurrence of extreme weather events on rice crops in India. The authors employed non-parametric statistical methods in the modeling and obtained satisfactory results. Dalhaus, Musshoff, and Finger (2018) proposed that it is possible to reduce the base risk of the parametric insurance contract in agricultural contracts by incorporating information about the phenology of the plants in the contract modeling.

The difficulty in designing agricultural parametric insurance contracts is due to the many variables to be considered in the process. First, it is necessary to understand the climate variable. For this, it is necessary to choose the appropriate modeling strategy that reliably describes it. Afterward, it is necessary to create a model that adequately reconciles the climatic variable with the occurrence of losses in the crop as well as proposed in as Afshar et al. (2021), Prokopchuk et al. (2020) and Dalhaus, Musshoff, and Finger (2018). Finally, it is also necessary to find the best indemnity contract trigger in define the relationship between climate and loss.

To understand the climate variable, it is first necessary to define what type of variable will be analyzed. Considering that in agriculture extreme weather events, despite having a low probability of occurrence have a great negative impact on crops, parametric contracts aimed at protecting crops against these events are very useful for farmers in risk management (Afshar et al. 2021). In the task to model extreme events, the precepts of the Theory of Extreme Values are used (Coles 2001). The methods proposed in this theory allow researchers to assess the probability of occurrence and the magnitude of extreme events (Ahmad et al. 2022). A probability distribution widely used in modeling extreme events is the generalized distribution of extreme values (*GEV*), proposed by Jenkinson (1955)(Aguirre, Nogueira, and Beijo 2020). This distribution is used to model a series of extreme values that are obtained by block maxima method (Coles 2001). Researchers like Martins et al. (2015), Carvalho et al. (2013) and Butturi-Gomes, Beijo, and Avelar (2019) also used the *GEV* distribution to model extreme weather events, obtaining satisfactory results.

To reconcile weather events with the occurrence of agricultural production losses, several methods can be applied. A simple strategy that can be used in this context is the regression method. This method makes it possible to carry out classification and prediction tasks (Draper and Smith 1998). To predict the occurrence of binary events, such as "loss" or "non-loss", as a result of a weather event, the logistic regression method is quite appropriate (Fernandes et al. 2021). As applied in Figueiredo et al. (2018), the logistic regression method provides the probability of observing a loss. So, after the definition of a classification threshold, events were labeled as "loss" or "non-loss" according to that probability. Following this, the authors defined it as the trigger that reduces the basis risk, the most accurate threshold in predicting the event of interest.

It is important to emphasize that the trigger must present not only security but also a viable cost for the insured and the insurer. Figueiredo et al. (2018) proposed a strategy for assessing and selecting indemnity triggers for parametric insurance contracts for floods in Jamaica. In the proposed strategy, statistical metrics were used to assess the safety in the indication of loss of each possible trigger, as well as the costs for the insurer and the insured that each one presents. After evaluating the security and cost-effectiveness of the contract, the best indemnification trigger could be selected.

To design the parametric contract model by the methods mentioned above, it is necessary to estimate the *GEV* distribution parameters and logistic regression coefficients. There are several ways to perform it, here we opted for the Bayesian inference approach. This approach is appropriate for dealing with extreme data that are scarce by nature because allows supplementary evidence via prior knowledge (Ahmad et al. 2022). Bayesian logic also facilitates the understanding of estimates for common sense, because the interpretation of the parameters is done directly on their distribution (Gelman et al. 2013).

Bayesian inference becomes possible to quantify the uncertainties about the parameter estimates through the credibility interval. Besides that, even in cases where there is a lack of prior knowledge, the Bayesian inferential approach has other advantages. It is the non-dependence of the procedure on certain regularity assumptions, as happens in classic inferential methods, and that the posterior distribution of the parameters has greater completeness of the inference procedures than the classical methods (Coles 2001).

This study aims to propose a Bayesian modeling strategy for agricultural parametric insurance for extreme weather events. The focus was on creating a model that helps insurers to define assertive

indemnity triggers by reconciling the occurrence of claims with the payouts at a lower cost for those involved. To achieve this goal, the theory of extreme values via extremely generalized distribution (*GEV*) was applied to model extreme weather events, logistic regression was used to create the loss predictor model and possible triggers were evaluated as proposed by Figueiredo *et al.* (2018). The Bayesian approach was used, allowing the uncertainty in relation to all results to be described as credible intervals. Thus, the main contribution of this article was the application of the assumptions of the theory of extreme values and the Bayesian approach in the definition of triggers for agricultural parametric insurance contracts.

After the presentation of the proposed method, a case study was performed. A parametric agricultural insurance contract against extreme dry spells during the critical period of coffee flowering was modeled. This contract was carried out for cities in the state of Minas Gerais, in Brazil. The relevance of this kind of contract comes from that country is the largest producer and exporter of grain in the world. Consequently, millions of Brazilians are financial dependent directly or indirectly on its full development to get their livelihood (CEPEA 2022).

The paper is organized as follows. Section 2 it is presented the modeling strategy for extreme events, the loss prediction model framework, and the statistical and financial evaluation methodology for indemnity triggers in the contract. In Section 3 an application example is performed. In Section 4 the results are discussed. Finally, conclusions are given in Section 5.

2. Methodology

2.1 Modeling series of extreme values

According to Coles (2001), a series of extremes can be obtained by the block maxima method. As defined by Ferreira and De Haan (2015) recall that $\tilde{X}_1, \tilde{X}_2, \dots$ are a sequence of *i.i.d* random variables with a common distribution function F . Define the block maxima of the sample $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ as:

$$X_i = \max_{(i-1)k < j \leq ik} \tilde{X}_j,$$

for $i = 1, \dots, B$, $n = kB$. Let $X_{1,B} \leq X_{2,B} \leq \dots \leq X_{B,B}$ be the order statistics of the block maxima X_1, X_2, \dots, X_k .

To model a series of independent extreme values obtained by blocks one of the most used distributions is the Generalized Extreme Value (*GEV*), introduced by Jenkinson (1955). Its cumulative density is given by:

$$G(x|\mu, \sigma, \varepsilon) = \exp\left[-\left\{1 + \varepsilon \left(\frac{x - \mu}{\sigma}\right)\right\}^{-\frac{1}{\varepsilon}}\right],$$

defined on the set $\{x : 1 + \varepsilon \left(\frac{x - \mu}{\sigma}\right) > 0\}$ where the parameter satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $\varepsilon \neq 0$. This density has three parameters: a location parameter, μ ; a scale parameter, σ ; and a shape parameter ε . The likelihood function of *GEV* distribution is useful in some methods of parameter estimation like maximum likelihood estimation and the Bayesian inference. Under the assumption of independent data, the likelihood function of the *GEV* distribution is given by:

$$L(\mu, \sigma, \varepsilon|\mathbf{x}) = \frac{1}{\sigma^I} \prod_{i=1}^I \left\{ \left[1 + \varepsilon \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\varepsilon}{\varepsilon}\right)} \right\} \exp \left\{ \sum_{i=1}^I \left\{ - \left[1 + \varepsilon \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1}{\varepsilon}\right)} \right\} \right\}, \quad (1)$$

where \mathbf{x} is the vector of extreme values, I is its size and $\varepsilon \neq 0$.

More likely risk events are more likely to incur claims and indemnity payments than less likely events. Thus, the probabilities associated with a possible loss-causing event are quite relevant in

quantifying risks. From the fitted *GEV* distribution is workable to estimate the exceedance probability of a quantile x . If $\varepsilon \neq 0$ it is defined by:

$$p_x = p(X > x) = 1 - F(x|\mu, \sigma, \varepsilon) = 1 - \exp \left\{ - \left[1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) \right] \right\}^{-\frac{1}{\varepsilon}},$$

defined for $1 + \varepsilon \left(\frac{x - \mu}{\sigma} \right) > 0$.

The knowledge of the probability of occurrence of risk events is useful in structuring a model that aims to predict losses related to their occurrence.

2.2 Logistic model as a loss predictor

To design a parametric insurance contract with reduced basis risk, it is important for insurers to understand the involved risk. One way to establish the risk is to research the history of loss, or claims, and their relationship with specific risk events (Figueiredo et al. 2018). So, let ω be the annual history of loss occurrence such as:

$$\omega \sim \text{Bernoulli}(p_\omega),$$

and so:

$$\omega = \begin{cases} 1 & , \quad O_y \leq uE_y \\ 0 & , \quad \text{otherwise,} \end{cases} \quad (2)$$

in which p_ω is represents $p(\omega = 1)$, where ω is a binary variable in which $\omega = 1$ indicates that loss occurs in year y , $\omega = 0$ indicates that loss don't occurs in year y , O_y is the observed event, E_y is the expected event and u is the chosen coverage level to the designed product.

The probability of occurrence of loss can be estimated by a simple logistic regression method, where the binary variable ω is the dependent variable. Here, the proposal uses the estimated exceedance probability of the climate risk event (p_x) as an independent variable. Thus, the proposed loss predictor model can be described as:

$$\text{logit}(E[\omega]) = \ln \left(\frac{p_\omega}{1 - p_\omega} \right) = \beta_0 + \beta_1 p_x, \quad (3)$$

where $p_\omega = p(\omega = 1|p_x)$ is the estimated probability of loss occurrence and $0 \leq p_\omega \leq 1$, $\beta_0 \in \mathbb{R}$ and $\beta_1 \in \mathbb{R}$ are the logistic regression coefficients and $0 \leq p_x \leq 1$ is the estimated exceedance probability of the climate risk event.

From equation 3, $p_\omega = p(\omega = 1|p_x)$ can be specified as:

$$p_\omega = p(\omega = 1|p_x) = \frac{1}{1 + \exp - \{ \beta_0 + \beta_1 p_x \}}.$$

To estimate the coefficients β_0 and β_1 of a logistic model as proposed in this paper, the likelihood function of the dependent variable ω is useful. So that, the likelihood function of the binomial distribution $L(\beta_0, \beta_1|\omega)$ is given by:

$$L(\beta_0, \beta_1|\omega, \mathbf{p}_x) = \prod_{i=1}^n \left[\frac{1}{1 + \exp - \{ \beta_0 + \beta_1 (p_x)_i \}} \right]^{\omega_i} \cdot \left[\frac{\exp - \{ \beta_0 + \beta_1 (p_x)_i \}}{1 + \exp - \{ \beta_0 + \beta_1 (p_x)_i \}} \right]^{1 - \omega_i}. \quad (4)$$

After fitting the logistic model, the probability of loss for each possible occurrence of the extreme climate risk event becomes known. In the proposed design of the parametric insurance contract, the probability of loss was defined as the indemnity index. Therefore, it is now necessary to identify the threshold that will trigger the indemnities at the lowest basis risk and cost.

2.3 Insurance trigger threshold analysis

The indemnity trigger that reduces the basis risk is the one that better conforms to losses and indemnity payments in the parametric contract (Figueiredo *et al.* 2018; Turner 2020). So, after it was stipulated that the estimated probability of loss (p_ω) would be the contract index, it was needed to define the loss payment mechanism that minimizes the basis risk. That is, it was necessary to define what is the probability of loss – the threshold – when indicated by the model should trigger the payments. The threshold chosen as an indemnity trigger must be accurate in predicting loss ($\omega = 1$) and non-loss ($\omega = 0$) occurrences.

For it, all possible activation triggers that indicate from 1% to 99% probability of loss were tested. Events with 0% or 100% probability of loss were disregarded because there is no uncertainty for these cases and so no insurance coverage. For simplicity the tested triggers were called q , where $q = 1\%, \dots, 99\%$. So, the estimated occurrence of loss for each q can be described as:

$$\omega_q = \begin{cases} 1 & , \quad p_\omega \geq q \\ 0 & , \quad \text{otherwise.} \end{cases}$$

Thus, as well as in Figueiredo *et al.* (2018), it was compared the actual loss occurrence (ω) with the expected occurrence of loss (ω_q), as present in Table 1.

Table 1. Possible events predicted by the model.

Condition	Event	Legend
$\omega = \omega_q = 1$	A	Loss correctly predicted
$\omega = 0$ and $\omega_q = 1$	B	False alarm
$\omega = 1$ and $\omega_q = 0$	C	Misses
$\omega = \omega_q = 0$	D	Non-loss correctly predicted

Then, the amount of the events of loss correctly predicted (A), false alarm (B), misses (C) and non-loss correctly predicted (D) for each trigger q was verified. These amounts are represented by $a(q)$, $b(q)$, $c(q)$ and $d(q)$, respectively. Then, the accuracy of each trigger in predicting loss and non-loss events was calculated as:

$$Accuracy_q = \frac{a(q) + d(q)}{a(q) + b(q) + c(q) + d(q)},$$

where the greater the accuracy of a trigger, the lower the basis risk and the greater the security associated with it in the parametric insurance contract.

Furthermore, it is not enough to define a trigger just for its accuracy. It must also be economically feasible for the insured and for the insurer. To economically evaluate the proposed contract different costs were attributed to the events A , B , C , D . The total cost in each trigger q can be described by the equation proposed by Figueiredo *et al.* (2018), in which

$$Cost_q = \frac{a(q)}{n} C_A + \frac{b(q)}{n} C_B + \frac{c(q)}{n} C_C + \frac{d(q)}{n} C_D,$$

where C_A , C_B , C_C , C_D are the costs of each events A , B , C , D and n is the amount of predicted events by the model.

Considering that the costs for each possible event vary for both the insurer and the insured, these costs were considered as in Table 2:

where E_O is the operational cost to be taken by the insurer whenever the model triggers a payout; E_{FA} is the cost of reputation loss and model re-calibration for insurer whenever the model triggers a

Table 2. Cost structure for the insurer and insured.

Agent	ω	$\omega_q = 1$	$\omega_q = 0$
Insured	1	$C_A = E_p(q)$	$C_C = E_p(q) + I$
	0	$C_B = E_p(q) - I$	$C_D = E_p(q)$
Insurer	1	$C_A = E_O$	$C_C = E_M$
	0	$C_B = E_O + E_{FA}$	$C_D = 0$

Adapted from Figueiredo et al. (2018).

payout that does not match to a real loss event; E_M is the cost of reputation loss and potential loss of client when the model fails to trigger a payout when an actual loss event occurs; I is the amount of indemnity contracted for when the model triggers a payout; and $E_p(q)$ is the insurance premium, obtained by:

$$E_p(q) = \frac{a(q) + b(q)}{n} .I.m,$$

where $m = 1 + \%loads$.

To define the trigger that reduces the basis risk in the proposed model it is necessary that the *GEV* distribution and the logistic regression model are completely specified. For this, Bayesian inference was applied to estimate the unknown quantities.

2.4 Bayesian Inference

"Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and unobserved quantities such as predictions for further observations" (Gelman et al. 2013, p.1). In the Bayesian inferential approach, the unknown parameters are treated as random variables and are described through a distribution (Wakefield 2013).

The main advantages of using Bayesian inference are: 1) the possibility of including prior information about the parameters in the inferential process. It reduces the uncertainties about the results, especially when dealing with scarce data; 2) The performed posterior density is more complete than those obtained by classic methods; 3) This inferential approach is not dependent on regularity assumptions required by the asymptotic maximum likelihood theory (Coles 2001).

The Bayesian inference approach is based on Bayes' theorem. It combines the likelihood function and the prior knowledge about the parameters. The posterior density of a parameter is described using Bayes' theorem, as follows:

$$p(\theta|x_1, \dots, x_n) = \frac{L(\theta|x_1, \dots, x_n) p(\theta)}{\int_{\Theta} L(\theta|x_1, \dots, x_n) p(\theta) d\theta} \propto L(\theta|x_1, \dots, x_n) p(\theta),$$

where x_1, \dots, x_n are the sampled values, θ is the parameter vector, $p(\theta)$ is the prior density for θ , $L(\theta|x)$ indicates the likelihood function of the adjusted distribution, $P(\theta|x)$ is the posterior distribution for θ and the integral is set over the parametric space Θ .

The choice of the prior density is crucial for the good performance of the Bayesian estimation process (Wakefield 2013). Prior information represents all the knowledge one has about θ before obtaining the sample x_1, \dots, x_n (Box and Tiao 2011). A prior is considered informative when it brings relevant information for specifying the distribution of the parameter and non-informative when there is ignorance about θ and little or nothing is known about the prior behavior (Paulino, Murteira, and Turkman 2003). The parameters of the prior distribution are called hyperparameters. When the

priori density has a very high variance, its density represents a complete ignorance about θ and is called a vague prior (Jeffreys 1998 & Gianola and Fernando 1986).

To estimate the joint posterior density for *GEV* parameters, it is needed the likelihood function of the *GEV* distribution presented in equation 1 and the joint prior density for its parameters, so that:

$$p(\mu, \sigma, \varepsilon | x_1, \dots, x_n) = L(\mu, \sigma, \varepsilon | x_1, \dots, x_n) p(\mu, \sigma, \varepsilon).$$

As well in Coles and Pericchi (2003) and Ahmad *et al.* (2022) was adopted as prior distribution for each *GEV* parameters:

$$\mu \sim N(m_\mu, S_\mu^2), \log(\sigma) \sim N(m_\sigma, S_\sigma^2), \varepsilon \sim N(m_\varepsilon, S_\varepsilon^2),$$

where was applied the logarithm in σ to secure the positivity of scale parameter (Ahmad *et al.* 2022 & Coles 2001). So, following Coles (2001), the proposed prior joint density for *GEV*'s parameters $P(\mu, \sigma, \varepsilon)$ can be described as:

$$p(\mu, \sigma, \varepsilon) = \exp\left(-\frac{1}{2} \left(\frac{\mu - m_\mu}{S_\mu}\right)^2\right) \cdot \exp\left(-\frac{1}{2} \left(\frac{\log(\sigma) - m_\sigma}{S_\sigma}\right)^2\right) \cdot \exp\left(-\frac{1}{2} \left(\frac{\varepsilon - m_\varepsilon}{S_\varepsilon}\right)^2\right),$$

where $m_\mu \in \mathbb{R}$, $m_\sigma > 0$, $m_\varepsilon \in \mathbb{R}$, $S_\mu > 0$, $S_\sigma > 0$ and $S_\varepsilon > 0$.

After specifying the posterior joint density for *GEV*'s parameters, was applied Markov Chain Monte Carlo (MCMC) algorithm to specify the margin distribution for each parameter. After removing the values generated in the adjustment period, the procedure leads to a sample of length s ($\theta_1, \dots, \theta_s$) that can be considered as observations of the stationary distribution for each parameter (Coles 2001). To check the convergence of the posterior densities generated by the chains there are tests such as the criteria of 1) Geweke *et al.* (1991), 2) Raftery *et al.* (1992) and 3) Heidelberger and Welch (1983).

Geweke's criterion evaluates the convergence of posterior chains, considering that the chain converges when the standardized difference between its means is not large. This criterion is evaluated through the Geweke absolute value ($|G_w|$). The Raftery & Lewis criterion estimates how many iterations are necessary for the Gibbs sampler to converge to the stationary distribution. This criterion is evaluated using the Raftery & Lewis dependency factor. The Heidelberger & Welch criterion assesses the stationarity of the parameter chain and also whether the sample size is sufficient to accurately generate estimates of the mean of the parameters. This criterion is evaluated using the p-value. Instead of using a summary statistic to describe the predicted value for each parameter, it can be used all s generated values in the MCMC process.

In the field of extreme value analysis, it is common to estimate probabilities that future events will reach extreme levels through predictive posterior distributions (Coles 2001). The exceedance probability of a quantile x_q of the random variable $X \sim \text{GEV}(\mu^r, \sigma^r, \varepsilon^r)$, supposing that the historical data x_1, \dots, x_n is available, is given by:

$$\hat{p}_{x^r} = \hat{p}(X \geq x_q | x_1, \dots, x_n)^r \propto 1 - \int_{\Theta} p(X \approx x_q | \mu^r, \sigma^r, \varepsilon^r) p(\mu^r, \sigma^r, \varepsilon^r | x_1, \dots, x_n) d(\mu, \sigma, \varepsilon), \quad (5)$$

where Θ is the parametric space, and $P(X \geq x | \mu^r, \sigma^r, \varepsilon^r)$ is the probability of exceedance for a quantile x of the maximum future random variable X , conditioned to the value of the parameters μ^r , σ^r and ε^r , and $p(\mu^r, \sigma^r, \varepsilon^r | x_1, \dots, x_n)^r$ is the posterior joint density for *GEV*'s parameters.

Although analytically the solution of the equation 5 is not feasible, it is easily approximated when posterior density was obtained by simulation (Coles 2001). So as:

$$\hat{p}_{x^r} = \hat{p}(X \geq x_q | x_1, \dots, x_n)^r = \frac{1}{s} \sum_{r=1}^s P(X \geq x_q | \mu^r, \sigma^r, \epsilon^r). \tag{6}$$

With the same idea, to estimate the logistic model coefficients by the Bayesian approach it was needed the likelihood function and the prior densities. It was used in the likelihood function of the binomial distribution in equation 4 and as prior density for coefficients, as well in Swaminathan and Gifford (1985), O'Brien and Dunson (2004) and Discacciati, Orsini, and Greenland (2015) was considered normal distribution for both, such as:

$$\beta_0 \sim N(m_{\beta_0}, S_{\beta_0}^2), \beta_1 \sim N(m_{\beta_1}, S_{\beta_1}^2).$$

Then, the joint prior distribution of the logistic model coefficients is given by:

$$p(\beta_0, \beta_1) \propto \exp \left[-\frac{1}{2} \left(\frac{\beta_0 - m_{\beta_0}}{S_{\beta_0}} \right)^2 \right] \cdot \left[-\frac{1}{2} \left(\frac{\beta_1 - m_{\beta_1}}{S_{\beta_1}} \right)^2 \right],$$

where $m_{\beta_0} \in \mathbb{R}$, $m_{\beta_1} \in \mathbb{R}$, $S_{\beta_0} > 0$, $S_{\beta_1} > 0$.

So, the joint posterior density for the coefficients is given by:

$$p(\beta_0, \beta_1 | \omega) = L(\beta_0, \beta_1 | \omega) p(\beta_0, \beta_1),$$

where to determine the marginal density for each coefficient, it was again used MCMC algorithm.

The knowledge about the posterior density for β_1 makes it possible to perform hypothesis tests related to it. That is, it is possible to verify whether the independent variable in the logistic model (\hat{p}_x) has a positive or negative impact on the dependent variable (ω). One way to perform the Bayesian hypothesis test is through Bayes Factor. Bayes factor is used to compare the data prediction strength between two different hypotheses and presents how many times the occurrence of one hypothesis is more likely than another (Ragulina and Reitan 2017). Developed by Jeffreys (1998) the Bayes factor against the null hypothesis H_0 , or in favor of the alternative hypothesis H_1 , can be described as:

$$FB_{1,0} = \frac{O(1, 0|x)}{O(1, 0)},$$

where $O(1, 0|x) = \frac{p(H_1|x)}{p(H_0|x)}$ is the posterior odds in favor of H_1 and $O(1, 0) = \frac{p(H_1)}{p(H_0)}$ represents the prior odds in favor of H_1 . The Bayes factor interpretation scale was proposed by Jeffreys (1998) which establishes the strength in favor of the tested hypothesis as:

Table 3. Jeffreys' interpretation scale for the Bayes Factor.

Bayes Factor	Strength of evidence
<1	Negative
1-3	Weak
3-10	Substantial
10-30	Strong
30-100	Very Strong
100	Decisive

The intervals of occurrence of the variables estimated by the Bayesian approach are called credibility intervals. The used credibility interval here was the Highest Posterior Density interval,

with a γ credibility level (HPD_γ). Among all possible intervals with a $\gamma = 1 - \alpha$ level of credibility, interval HPD presents the maximum probability of occurrence for the posterior density, representing the posterior region with the lowest amplitude and highest density (Paulino, Murteira, and Turkman 2003).

In order to evaluate the proposed design model of a parametric agricultural insurance contract against extreme weather events, an application was developed through a case study.

3. Application Example

3.1 Case study

Brazilian agricultural production feeds about 10% of the world, directly and through processed foods (Embrapa 2021). In the country, the agricultural sector is one of the sectors that most contribute to the growth of the national GDP, representing 21% of the sum of all wealth produced, one-fifth of all jobs, and more than 40% of Brazilian exports, reaching US\$ 96.7 billion in 2019 (Embrapa 2020).

Among all the crops produced in the country, Brazil stands out as the largest producer and exporter of coffee in the world (CONAB 2021). Just in 2022, the country produced more than 39 million bags that reached more than 120 countries, generating foreign exchange revenue of US\$ 9.2 billion. Thus, coffee growing is highly relevant to the country's economy, acting as an income generator for thousands of Brazilians who work directly or indirectly in this activity (CEPEA 2022). Despite this, only 130,000 of the 1.8 million producing hectares have some type of insurance. It let producers be highly subject to bad weather that jeopardizes harvests and affects the production and quality of grains, generating instability and often economic losses (CNA 2021).

As with other cultures, for coffee activity to develop fully, it is necessary that favorable climatic conditions coexist. Although water availability during coffee development is not the only factor influencing production, it is the main factor that affects coffee productivity (Picini et al. 1999; Martinez et al. 2015; Aparecido, Rolim, and Souza 2015). Arruda et al. 2000 and Soares et al. 2021 developed researches that related the occurrence of rainfall to coffee productivity. Therefore, it is consistent that parametric insurance products aimed at protecting coffee crops against dry spells, especially extreme dry spells, should be created.

Here, it was adopted the definition of Ceballos, Martinez-Fernández, and Luengo-Ugidos (2004) for dry spells. According to the authors, a sequence of n consecutive days without a relevant amount of pluvial precipitation is considered a dry spell. Thus, an "extreme dry spell" corresponds to the longest sequence of n days without relevant rain in a given period. By the definition of Castro and Vilela (1986), a day can be considered as without relevant rain if the observed rainfall is less than 3 mm.

The appropriate combination of meteorological factors to crop reaches the desired production levels varies at each phenological stage of the coffee tree (Camargo and Camargo 2001). The coffee tree has a phenological cycle that lasts two years. This period is divided into six phenological phases, the first two vegetative phases, and the last four reproductive phases. One phase of the coffee tree that presents greater sensitivity related to extreme dry spells is the first reproductive phase (R1) (Embrapa 2009; Camargo and Camargo 2001; Ruiz-Cardenas 2015). This stage occurs from September to December of the second phenological year of the plant. In this phase happens the flowering and expansion of fruits (Camargo and Camargo 2001). Therefore, in this work an application example of the proposed method for modeling a parametric insurance contract for coffee plantations against occurrences of extreme droughts in the phenological phase R1 was carried out.

3.2 Data

This study is based on a data set containing daily precipitation data and estimated and observed coffee annual production in bags per hectare for crops from 2014 to 2021. The data was provided by the Regional Cooperative of Coffee Growers in Guaxupé (COOXUPÉ) corresponding to three

cities: Guaxupé (G)¹, Monte Santo de Minas (MSM)² and Monte Carmelo (MC)³. All data was aggregated by the municipality. Besides that, it was used daily precipitation data of Muzambinho⁴ city, provided by the National Water Agency (ANA) in Brazil to be applied as prior information in the *GEV* distribution fitting. The data about rain in Muzambinho was from 1974 to 2021. All cities are in the State of Minas Gerais in Brazil, which is the largest coffee producer in the country (CONAB 2021). The location of the evaluated cities can be seen in Figure 1.

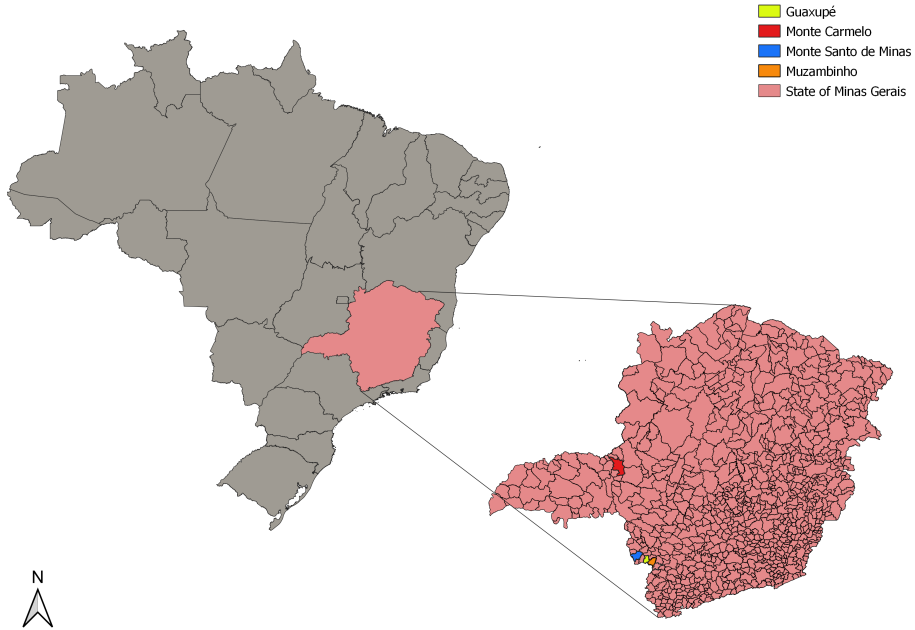


Figure 1. Location of the studied cities.

According to Reboita et al. (2015), all the cities in the present study have the same Köppen–Geiger climate classification, characterized by a tropical savannah climate with a dry season in winter (Aw).

3.3 Modelling framework

3.3.1 Defining the independent variable of the loss predictor model

The first step was to block the daily rain data of each city in flowering period for the eight years that composed the sample. Subsequently, sequences of consecutive days without relevant rain were identified. Due to the limited amount of data, it was decided to extract a series of second-order maximums ($k=2$). That is, for each city and year sampled, the two highest observed dry spells in R1 were selected to compose the series.

For each location, the independence of the data series was verified as also whether there was any tendency to be modeled. To verify independence, the Ljung–Box test was used and to verify trends existence the Mann–Kendall test was used. Both tests were made for a 5% level of significance. Each city, Guaxupé, Monte Santo de Minas and Monte Carmelo, had its own fitted *GEV* distribution. In

1. UTM coordinates E: 322818 - N: 7643426 - altitude: 870 m
2. UTM coordinates E: 295062 - N: 7655267 - altitude: 910 m
3. UTM coordinates: E: 237642 - N: 7927660 - altitude: 912 m
4. UTM coordinates: E: 342226 - N: 7636178 - altitude: 1059 m

order to characterize extreme dry spells to enable the understanding of their behavior, as well as to make predictions about their occurrence, the analysis of extreme values via *GEV* distribution proves to be quite relevant (Butturi-Gomes, Beijo, and Avelar 2019; Carvalho *et al.* 2013).

In order to obtain the prior information for the hyperparameters, it was fitted a *GEV* distribution for the extreme dry spells in Muzambinho city using the likelihood-based approach. Our prior specification for the hyperparameters is given as follows. For the location hyperparameters, we set the classical estimation for the mean and for the scale, we reduce the precision by three times less than what we found in our estimation procedure.

In addition, fitted *GEV* distributions consider non-informative priors. The purpose is to verify the influence of the prior sensitivity on the results. Hence, we specific hyperparameters density are described in Table 4:

Table 4. Description of the prior information used in fitting the *GEV* distributions.

Prior	μ	σ	ϵ
Informative	$N(16.06, 3.31)$	$N(1.73, 0.58)$	$N(-0.06, 0.07)$
Non-informative	$N(0, 10000)$	$N(-4.60, 9.21)$	$N(0, 100)$

To define the posterior density for each *GEV*'s parameters for each city it was applied the MCMC algorithm with 3 chains, 8,000 iterations, and 3, 000 burnings. As result was a sample of 15,000 values for each parameter. To verify the convergence of the chain, it was expected that the absolute value of the Geweke statistic would be less than 1.96 ($|GW| \leq 1.96$), the Raftery & Lewis dependency factor would be less than 5, and the Heidelberger & Welch *p-value* < 0.05. Therefore, observing the expected conditions for the three criteria, it was considered that there was no evidence of a lack of convergence for the parameter chains. After the MCMC process, it was possible to analyze 15,000 possible *GEV*'s to describe the data.

After that, it was selected only the quantile that represents the maximum observed dry spell for each city and year (x_{cy}). Considering equation 6, it was estimated the exceedance probability for each quantile ($\widehat{p}_{x_{cy}}^r$). Then, the vectors of exceedance probability were used as independent variables in the proposed loss predictor model, so that:

$$\widehat{p}_x^r = \left\{ \widehat{p}_{x_{cy}}^r \right\}, \tag{7}$$

for $c = G, MSM, MC, \gamma = 2014, \dots, 2021$ and $r = 1, \dots, 15, 000$.

3.3.2 Defining the dependent variable of the loss predictor model

In order to predict the occurrence of losses in final coffee production, the logistic model was adjusted as proposed in section 2.2. We define the ω variable of the model, according to equation 2, and a contractual coverage level of $u=85\%$ was used in our case study. As attested in Amaral *et al.* (2017), the tolerance of 15% in the definition of loss is close to the lowest level of profit obtained in agricultural activities. This would mean that, even in the most unfavorable profit scenario the producer would not be harmed as he would be fully reimbursed by the indemnity. So the used dependent variable can be described as:

$$\omega_{cy} = \begin{cases} 1 & , O_{cy} \leq 0.85E_{cy} \\ 0 & , otherwise, \end{cases}$$

where c represents the city, γ represents the year O_{cy} is the observed production in city c in year γ , and E_{cy} is the expected production in city c in year γ . For Guaxupé and Monte Santo de Minas was observed loss in 2014 and for Monte Carmelo was observed loss in 2015.

So, considering that all cities have the same climatic classification and due to the small amount of data, the dependent variable of this example was defined considering the three cities together, where:

$$\omega = \{ \omega_{cy} \},$$

for $c = G, MSM, MC$ and $y = 2014, \dots, 2021$.

Once the explanatory and the response variable are defined, it is possible to fit the logistic model which will predict the coffee production losses.

3.3.3 The loss predictor model

To estimate the probability of loss in final coffee production due to extreme dry spells observed in the R1 period, two logistic regression models were fitted. One model was fitted with informative priors for *GEV*'s distributions and the other was fitted with non-informative priors. Then, due to the lack of reliable prior information, estimate the coefficients β_0 and β_1 of the logistic models employed vague priors. The vague priors was described as $\beta_0 \sim N(m_{\beta_0} = 0; S_{\beta_0}^2 = 10^4)$ and $\beta_1 \sim N(m_{\beta_1} = 0; S_{\beta_1}^2 = 10^4)$.

As well as made to specify the *GEV*'s parameters distribution, to define the density of each logistic model coefficient three chains were run to check the convergence of the algorithm. The posterior inference was made with 5000 iterations, after the 3000 burning iterations. To identify whether the occurrence of extreme dry spells in a phenological period has an effect on the occurrence of losses, we started from the initial hypothesis that the greater the probability of dry spells, the greater the probability of observing a loss as a result of it. According to this hypothesis, contracts for periods with more likely dry spells tend to be more risky and result in greater probabilities of payment of compensation by the insurer.

Thus, the following hypotheses were considered for the test: $H_0 : \beta_1 \leq 0$ and $H_1 : \beta_1 > 0$. H_0 indicates that the increase in the probability of observed extreme dry spell does not increase the probability of loss. H_1 indicates that the increase in the probability of extreme dry spell causes an increase in the probability of loss. To test the hypotheses H_0 and H_1 by Bayes Factor, the interpretation was carried out according to the evidence scale proposed by Jeffreys (1998), presented in section 2.4. Thus, if the model has strong or decisive evidence in favor of H_1 , this hypothesis is accepted and the evaluation of the contractual triggers of the parametric insurance follows as explained in the section 2.3.

Finally, in the case study proposed it employed a similar costs as proposed by Figueiredo et al. (2018). The hypothetical adopted costs are shown in Table 5.

Table 5. Hypothetical costs used for evaluating the proposed model.

Description	Value
Indemnity (I)	\$100,000.00
Administration charges, profit, and other expenses (m)	15%
Operational cost related with administrative actions that need to be taken whenever the model triggers a payout (E_o)	5% of the indemnity (\$5,000)
Cost of insurer reputation loss and model calibration when events of false alarm occur (E_{FA})	2% of the indemnity (\$2000)
Cost of loss insurer reputation and model calibration when failure events occur (E_M)	35% of the indemnity (\$35000)

Briefly, the procedures followed in the research are shown in Figure 2.

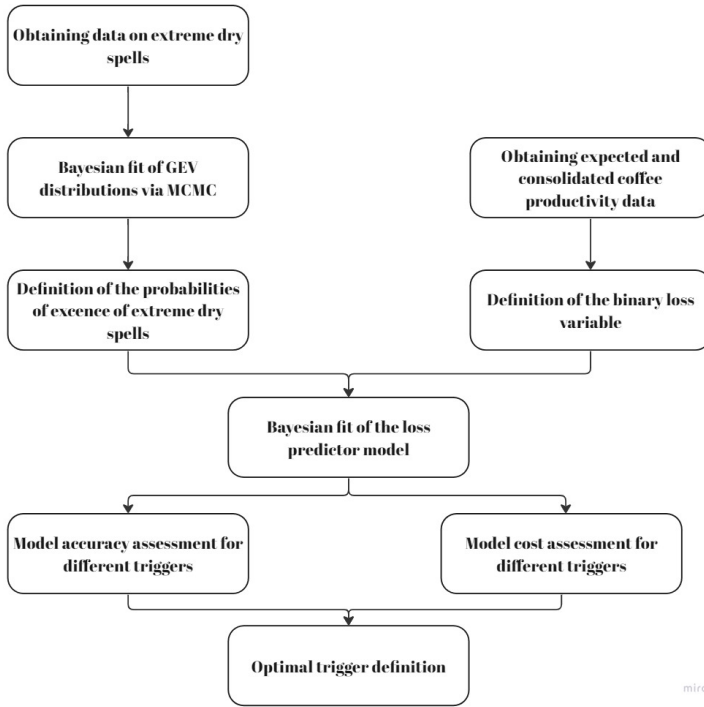


Figure 2. Summary of research steps.

4. Results and Remarks

4.1 Extreme dry spells analysis

The second-order maximum dry spells obtained for the R1 phenological phase of the coffee plant for each city are shown in Figure 3. Although all cities are located in the same region in Brazil, it is possible to verify a notable variation among the observations. Table 6 also highlights the difference among the cities when we compute the descriptive statistics.

Table 6. Descriptive statistics for extreme dry spells data for the cities of Guaxupé, Monte Santo de Minas and Monte Carmelo.

City	Min	Max	Median	Standard deviation
Guaxupé	7	32	16	7.22
Monte Santo de Minas	8	28	16.5	5.48
Monte Carmelo	9	44	16.50	9.40

Before carrying out the fit of the *GEV* distribution extreme dry spells of the data of each city, the assumption of independence of the series was verified and also was verified if there was any tendency to be modeled in the data. The Ljung-Box (Ljung and Box 1978) and Mann-Kendall (Mann 1945) tests showed non-significant results ($p - value > 0.05$) for all cities. In this way, the assumption of independence was met, as well as there wasn't necessary to model trends in none of the cities. Table 7 presents the $HPD_{95\%}$ intervals of the *GEV* distribution parameters for the different cities, as well as the results of the convergence criteria of the corresponding posterior density chains.

Table 7 shows that all the posterior chains of the parameters met the conditions proposed in the three convergence criteria evaluated. So, it is pointed out that there are no indications of a lack of convergence in the chains generated for the parameters. By Geweke's criterion, it presented

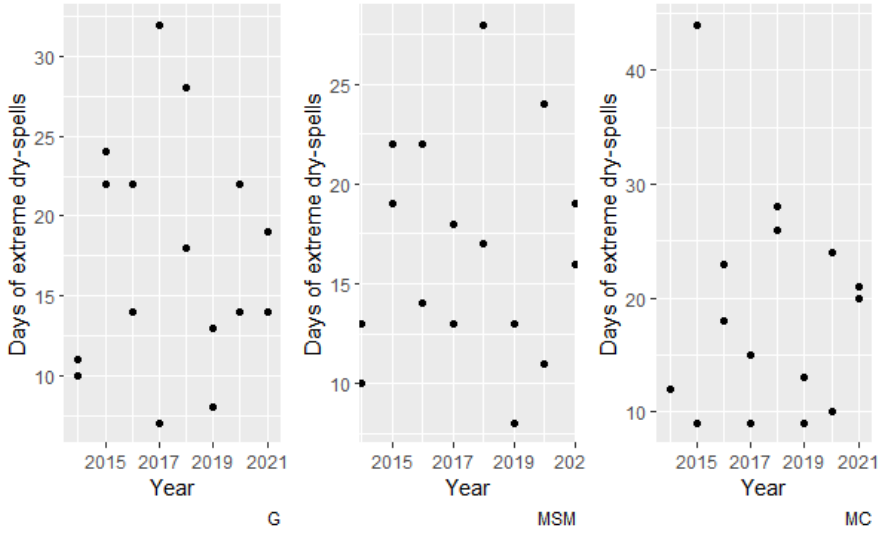


Figure 3. Values of series of second-order maximums of extreme dry-spells in the R1 period (Sep-Dec) in the cities of Guaxupé (G), Monte Santo de Minas (MSM) and Monte Carmelo (MC).

Table 7. Results of the $HPD_{95\%}$ intervals of the parameters of the GEV distribution according to the city and the results of the criteria for the analysis of convergence of the chains of the posterior distribution of the parameters, namely Raftery and Lewis (R&L), Geweke (GW) and Heidelberger-Welch (H&W).

GEV's prior	City	Parameter	$HPD_{95\%}$	Median	GW	R&L	H&W
Informative	Guaxupé	μ	[12.49 , 17.71]	15.02	0.54	1.00	0.87
		σ	[4.39 , 9.47]	6.63	0.07	0.99	0.63
		ϵ	[-0.41 , 0.26]	-0.09	0.49	0.99	0.16
	MSM	μ	[12.78 , 17.26]	14.91	0.78	1.00	0.70
		σ	[3.56 , 7.55]	5.97	0.92	1.02	0.60
		ϵ	[-0.41 , 0.24]	0.01	0.16	0.97	0.91
	MC	μ	[12.50 , 17.77]	14.88	0.98	1.01	0.51
		σ	[4.09 , 10.04]	7.83	0.37	1.04	0.22
		ϵ	[-0.19 , 0.44]	0.21	0.40	0.98	0.75
Non-Informative	Guaxupé	μ	[10.22 , 17.52]	13.69	0.64	0.99	0.28
		σ	[3.67 , 9.49]	6.16	0.12	1.02	0.87
		ϵ	[-0.46 , 0.58]	0.00	0.06	0.99	0.99
	MSM	μ	[11.16 , 17.11]	14.11	0.80	1.01	0.97
		σ	[3.03 , 7.48]	4.92	0.43	1.04	0.79
		ϵ	[-0.48 , 0.45]	-0.07	1.83	1.03	0.27
	MC	μ	[9.04 , 15.04]	11.22	0.30	4.67	0.61
		σ	[1.02 , 8.24]	3.80	0.22	1.05	0.43
		ϵ	[0.08 , 3.53]	0.92	0.83	1.01	0.27

$|GW| < 1.96$, indicating that there is no evidence of a lack of convergence. Regarding the Raftery and Lewis criterion, all were less than 5, which indicates that the number of iterations was sufficient to define the stationary distribution. The p -value of the Heidelberger and Welch criterion was not

significant ($p > 0.05$), attesting that the series is stationary.

As pointed out in section 3.3.3, all the generated chains were applied as possible values for the *GEV* distribution parameters. Using the *GEV*'s distributions fitted for each city, it was obtained the exceedance probabilities for each quantile that represent the maximum dry spells observed in the *R1* period for each analyzed year. The maximum observed dry spells in each year and region can be seen in Figure 4 and the $HPD_{95\%}$ interval for the exceedance probability for each maximum observed dry spells is presented in Figure ?? . After that, it was fitted the proposed loss predictor model.

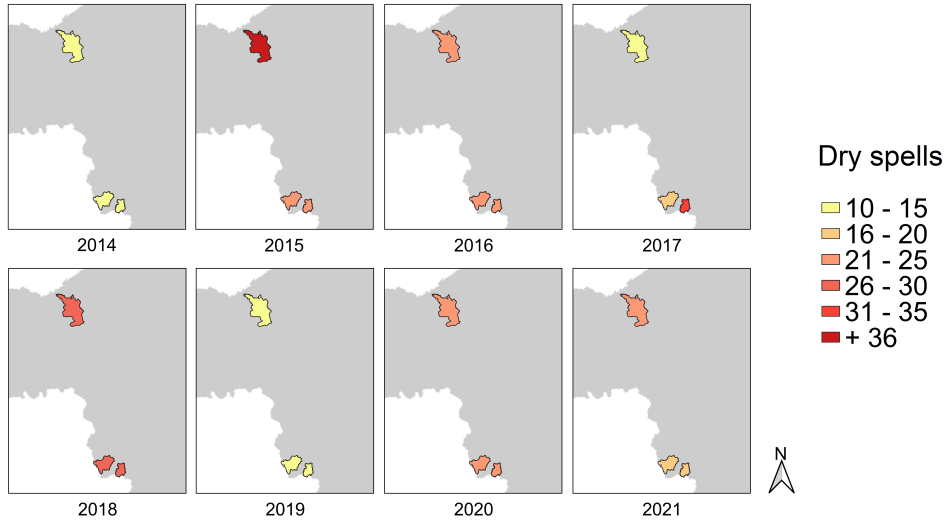


Figure 4. Sequences of the maximum consecutive days with less than 3mm of rain in the cities of Guaxupé, Monte Santo de Minas, and Monte Carmelo in the years from 2014 to 2021.

4.2 Loss predictor model

According to the results in Table 8, the adjustment of the logistic model demonstrated that our initial hypothesis was accepted. There is, the greater the probability of exceeding extreme dry spells, the greater are also the chances of observing losses in production as a result. Besides that, according to the scale proposed by Jeffreys for the Bayes factor, the tested hypothesis was decisively accepted.

Table 8. Credibility interval $HPD_{95\%}$ for the logistic model coefficients, the posterior median for results, the Bayes factor in favor of H_1 and $exp(\beta)$.

GEV's prior	Coefficient	$HPD_{95\%}$	Median	BF	$exp(\beta)$
Informative	β_0	[-3.81 , -2.90]	-3.39	-	[2.21% , 5.50%]
	β_1	[1.59 , 3.63]	2.59	>100	[94.5% , 97.79%]
Non-Informative	β_0	[-5.51 , -2.93]	-3.98	-	[0.40% , 5.35%]
	β_1	[1.71 , 8.23]	4.47	>100	[94.66% , 99.59%]

The interpretation of β_0 in Table 8 suggests that for the prior informative for the *GEV* distribution, only 2.21% ($exp(-3.81)$) to 5.50% ($exp(-2.90)$), of the observed losses are due to factors other than extreme dry spells. By complementary, it can be inferred that 94.5% to 97.79% of losses in coffee production for the evaluated cities are related to the observed extreme dry spells. Considering prior non-informative for the *GEV* distribution, 0.40% ($exp(-5.51)$) to 5.35% ($exp(-2.93)$) of losses are

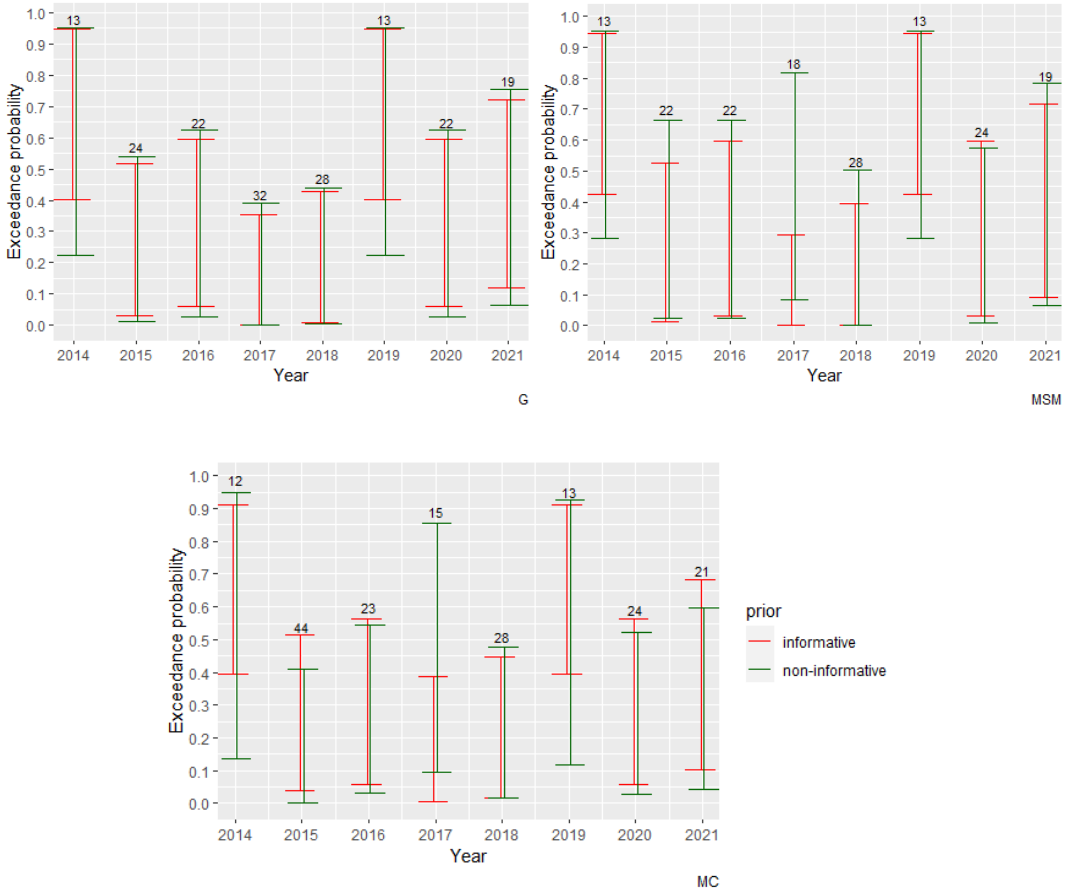


Figure 8. HPD_{95%} interval for the exceedance probability of the observed dry spells in Guaxupé (G), Monte Santo de Minas (MSM), and Monte Carmelo (MC) in the years from 2014 to 2021.

explained by factors other than extreme dry spells, while 94.66% to 99.59% of them are related to them.

These results confirm that, for the analyzed data, the most influential factor in the final coffee productivity is the water availability (Picini et al. 1999, Martins et al. 2015 and Aparecido, Rolim, and Souza 2015). Despite this, it is reinforced that the results obtained here refer only to the analyzed data, and cannot be generalized. This is because the research was carried out considering only one source of meteorological and coffee production data for the cities studied. The achievement of the generalization of this result would depend on a larger area of study, as well as the use of more data.

The results even show that the credibility interval for the regression coefficient related to the use of informative prior is more precise. It demonstrates a gain in the inferential process by its use. The initial hypothesis proposed was attested by both priors structures. So, the analysis was carried out to verify which trigger to be used in the parametric contract for the protection of coffee in case of extreme dry spells in flowering results in the lowest basic risk and viable cost for the agents.

4.3 Choice of trigger level: accuracy analysis

For all of the 15,000 fitted loss predictor models, the highest accuracy found was 87,5%. So, it defined the interval of triggers that represented the highest accuracy in each model. Then, the lower

limit of the most accurate interval of triggers for each model was selected. The selected value of each model represents the trigger in which the model starts to present the highest accuracy. Table 9 and Figure 9 show the $HPD_{95\%}$ interval for the posterior distribution of the most accurate trigger, for informative GEV (a) and non-informative prior for GEV (b).

Table 9. Credibility interval $HPD_{95\%}$ for the posterior distribution of the trigger at the accuracy of 87.5%.

GEV's prior	$HPD_{95\%}$	Median
Informative	[0.21, 0.29]	0.26
Non-Informative	[0.13, 0.31]	0.23

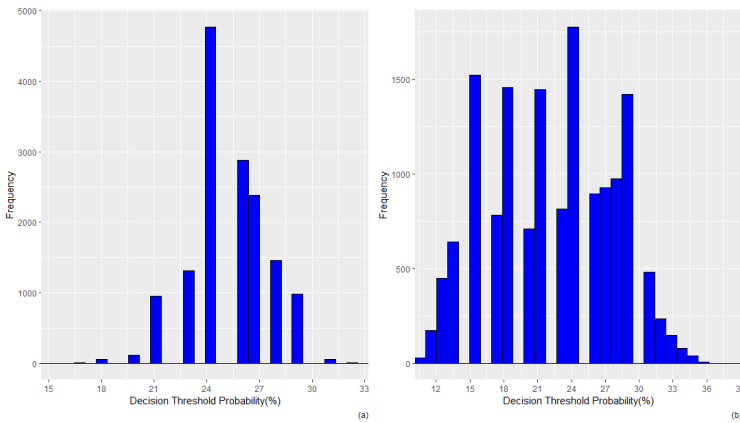


Figure 9. Posterior distribution of the trigger at the 87.5% accuracy considering informative prior for GEV (a) and non-informative prior for GEV (b).

If it is considered informative prior to the GEV distribution, with 95% of credibility the most accurate trigger is between the triggers 21% and 29%. If it is considered non-informative prior, the most accurate trigger is between the triggers 13% to 31%. It is seen that the use of prior information presented a narrower interval of choice for triggers. This is a positive finding, as it tends to make it easier for the manager to choose the ideal and safest contractual trigger.

4.4 Choice of trigger level: cost analysis

The lowest found cost of the proposed contract for the insurer was \$2,458.33. Considering informative prior for GEV parameters 2,200 simulations found this lowest cost. For non-informative prior it was 2,910 simulations. So, it was possible to define the posterior distribution of the trigger when the cost to the insurer is the lowest. Its results are shown in Table 10 and Figure 10.

Table 10. Credibility interval $HPD_{95\%}$ for the posterior distribution of the trigger when the cost to the insurer is the lowest found - \$2,458.33.

GEV's prior	$HPD_{95\%}$	Median
Informative	[0.20, 0.27]	0.24
Non-Informative	[0.13, 0.27]	0.18

As demonstrated in Table 10, if it is considered informative prior for GEV parameters, with 95% of credibility, the triggers that show the lower cost for insured are between the triggers of 20% and

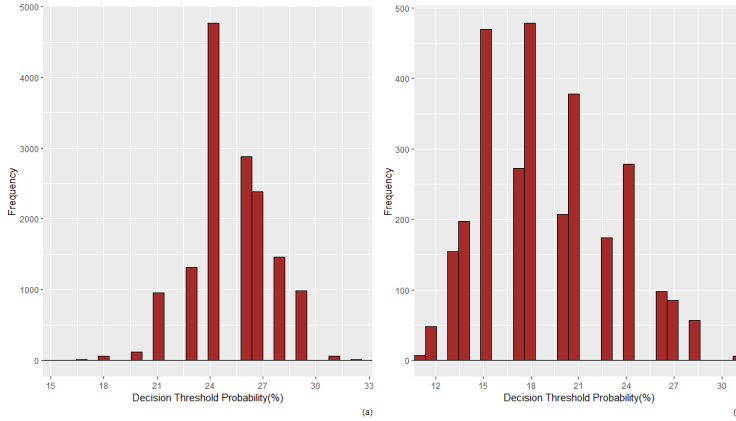


Figure 10. Posterior distribution of the trigger when the cost to the insurer is the lowest found - \$2,458.33, considering informative prior (a) and non-informative prior (b).

27%. If it is considered non-informative prior for *GEV* parameters, it is between the triggers 13% and 27%. Again, it was seen that greater precision is observed in the results related to the informative prior. In summary, for the model with informative prior, with 95% credibility, the highest accuracy is between the triggers of 21% and 29% and the lowest cost for the insurer is between the triggers of 20% and 27%. So, it can be considered that the triggers that tend to result in a contract safer and cheaper for the insurer occur when the model points to a probability of loss at the intersection between these two intervals, that is, for triggers between 21% and 27%.

Likewise, considering the model with non-informative prior, the highest accuracy is between the triggers of 13% and 31% and the lowest cost for the insurer is between the triggers of 13% and 27%. So, the triggers between 13% and 27% are considered safer and the best price for the insurer. In this way, it was verified among the triggers from 21% to 27%, for the informative prior model, and from 13% to 27%, for the non-informative prior model, which would be the total costs incurred by the insured in these triggers. Figure 11 presents the median and mean premium costs for informative prior and non-informative prior.

As shown in Figure 11, for the informative prior model the median premium cost is lower for the insured among the 24% to 27% triggers. In addition, the lowest mean premium cost value happens for the insured at the 27% trigger. Already for the non-informative prior model, the median premium cost is lower for the insured among the 20% to 27% and the lowest mean premium cost value happens for the insured at 27%. It is important to note that the median and mean premium values are smaller in the model that is considered prior informative for *GEV* parameters. It means that the lowest credibility interval previously found for this structure of prior presented in Table 8 has a positive influence on the final cost of the proposed contract.

Considering the evaluated triggers for each model, the median total cost of the proposed contract for the insured is shown in Figure 12.

Figure 12 confirms that the lowest cost for the insured occurs when the contractual indemnification trigger is set at 27% in both models. As presented in Table 9, this trigger also results in a high level of accuracy, reducing the basis risk of the contract. Finally, Figure 13 shows the median total cost and its *HPD*_{95%} interval for the insurer and the insured, across all possible triggers to be considered.

In fact, concerning the model fitting with informative prior for *GEV* parameters, Figure 13 demonstrated that the 27% trigger, in addition to demonstrating good level accuracy, also results in lower costs for the insured. Besides that, in relation to insurers it is possible to verify that within

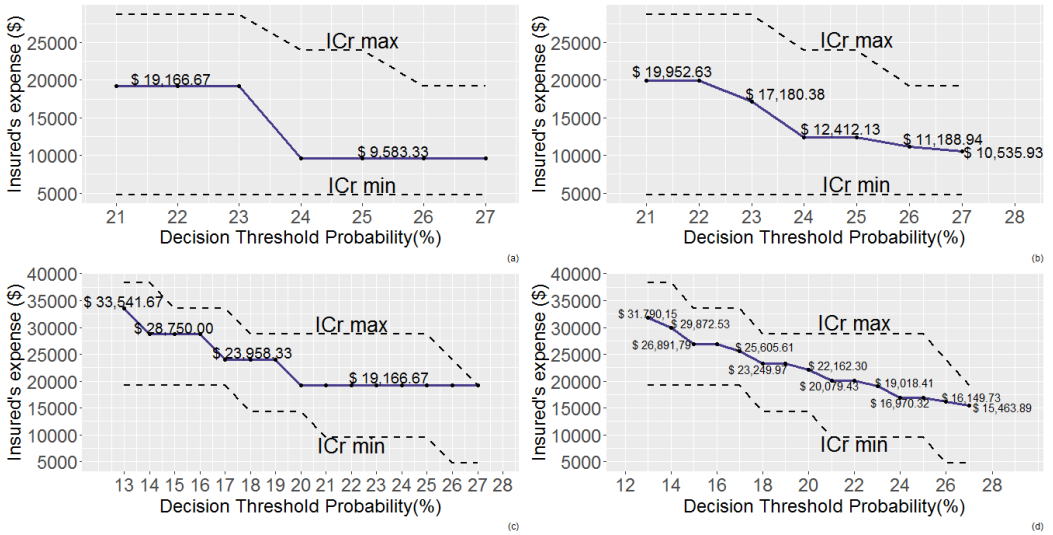


Figure 11. Median (a) and mean (b) premium costs and their $HPD_{95\%}$ interval to the insured to informative prior model and median (c) and mean (d) cost and their $HPD_{95\%}$ interval to the insured to the non-informative prior model.

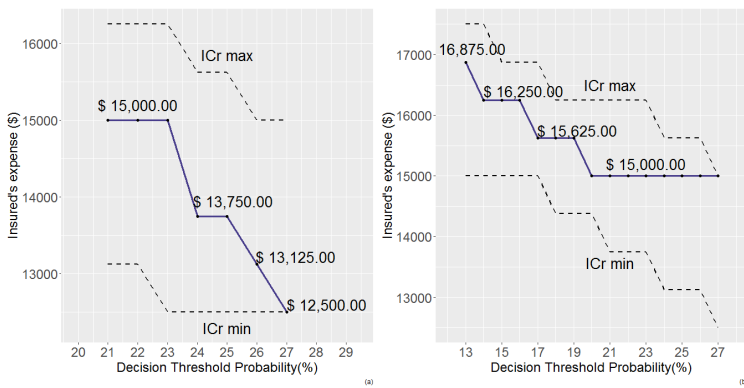


Figure 12. Median total cost and their $HPD_{95\%}$ interval to the insured considering the triggers 21% to 27%, for informative (a) and non-informative (b) prior model.

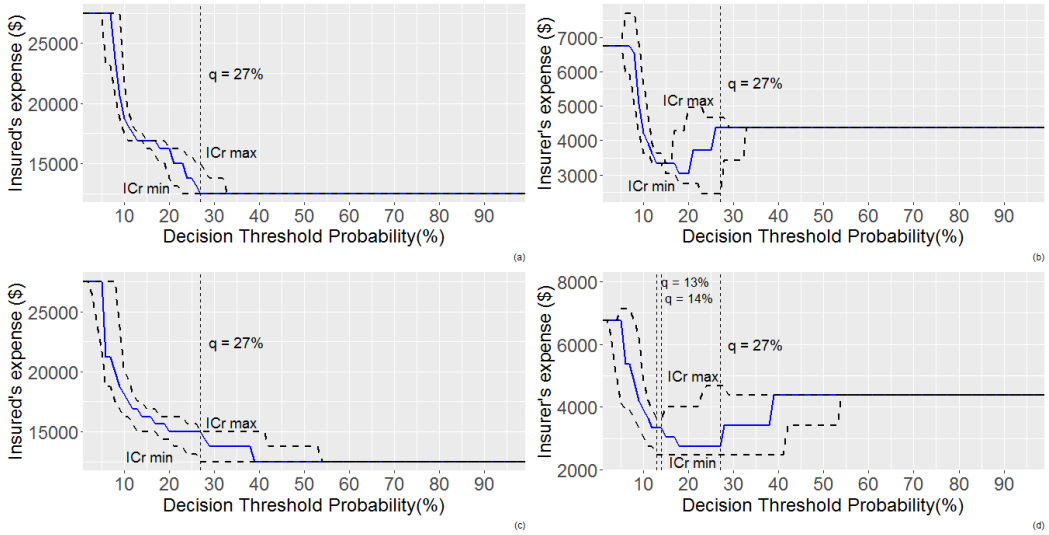


Figure 13. Median total cost considering informative prior model to the insured (a) and insurer (b) and considering non-informative prior to the insured (c) and insurer (d) and their $HPD_{95\%}$ interval.

the credibility interval, the trigger 27% presents the lowest limit for the costs and also presents the lowest maximum costs for the insured, among the triggers from 21% to 27%. Thus, unlike the work carried out by Figueiredo et al. (2018) which proposed a different optimal trigger for the insured and insurer, here a common trigger was established for both the insurer and the insured.

Regarding the model with non-informative prior for *GEV* parameters, it was seen inside the credibility interval the trigger of 27% found the lowest minimum and maximum cost for the insured at the triggers of 13% to 27%. In this model, according to the $HPD_{95\%}$ the insurer has the lowest cost minimum and maximum costs in the trigger of 13% and 14%. So, considering the non-informative prior model there are divergences between the best contractual trigger for the insurer and the insured, as well as in Figueiredo et al. (2018).

Thus, as well as in Pai et al. (2022) the use of a Bayesian assumptions model was able to capture the relationship between crop loss and the occurrence of extreme weather events with reasonable accuracy, which allows the proposed model to be used as a reasonable payment mechanism for parametric insurance with reduced basis risk.

5. Conclusion

In this research, a modeling strategy was proposed for parametric agricultural insurance contracts aimed at protecting crops against the occurrence of extreme events. The method proposed in this work was created by merging several existing methods, such as analysis of extremes, logistic regression, Bayesian inference, and risk analysis.

The use of Bayesian inference brought gains in the precision of the results, through the incorporation of previous information, as well as in the interpretation of the results. This is because the credibility interval allowed quantifying uncertainties with a reasonable degree of precision, at the various decision-making levels. It helps the management of parametric insurance contracts to take the best decisions related to this product.

A case study was carried out applying the proposed method to design a parametric insurance contract to protect coffee production from dry spells during the flowering period. The case study had limited application as there was a paucity of data. However, even with this limitation, it was

possible to establish an optimal trigger that minimized the base risk. The optimal contractual trigger, in this case, was when the model pointed to a 27% probability of loss. In this trigger, the proposed model was able to predict loss events with 87.5% accuracy. In addition, contract costs for both the insured and the insurer have also been reduced in the ideal trigger.

The main conclusion was that the methodological procedures proposed in this work can be adapted to the elaboration of agricultural parametric insurance contracts for different cultures and climatic events, even when data is scarce.

Acknowledgement

We thank CAPES for funding this research and we also thank Cooxupé for providing the data.

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Financing Code 001.

References

- Afshar, Mehdi H, Timothy Foster, Thomas P Higginbottom, Ben Parkes, Koen Hufkens, Sanjay Mansabdar, Francisco Ceballos, and Berber Kramer. 2021. Improving the performance of index insurance using crop models and phenological monitoring. *Remote Sensing* 13 (5): 924.
- Aguirre, Alberto Frank Lázaro, Denismar Alves Nogueira, and Luiz Alberto Beijo. 2020. Análise da temperatura máxima de piracicaba (sp) via distribuição gev não estacionária: uma abordagem bayesiana. *Revista Brasileira de Climatologia* 27:496–517.
- Ahmad, Touqeer, Ishfaq Ahmad, Irshad Ahmad Arshad, and Nicolas Bianco. 2022. A comprehensive study on the bayesian modelling of extreme rainfall: a case study from pakistan. *International Journal of Climatology* 42 (1): 208–224.
- Amaral, Ana Maria Santana, Adriano Bortolotti Silva, Marina Ariele Angelocci, Fernando Ferrari Putti, Thamiris Lentz de Almeida Coelho, Wanessa Tavares Campos Corsini, Fábio dos Santos Corsini, and José Messians Miranda. 2017. Certifica minas café certification in coffee plantations in the alfenas region–south of minas gerais. *Coffee Science* 12 (1): 114–123.
- Aparecido, Lucas Eduardo de Oliveira, Glauco de Souza Rolim, and Paulo Sergio Souza. 2015. Sensitivity of newly transplanted coffee plants to climatic conditions at altitudes of minas gerais, brazil. *Australian Journal of Crop Science* 9 (2): 160–167.
- Arruda, Flávio Bussmeyer, Mara de Andrade Marinho Weill, Ângela Iaffé, Emilio Sakai, and Regina Célia de Matos Pires. 2000. Estudo da influência do clima e do consumo hídrico na produção de cafeeiros (coffea arabica l.) em pindorama, sp.
- Box, George EP, and George C Tiao. 2011. *Bayesian inference in statistical analysis*. John Wiley & Sons.
- Butturi-Gomes, Davi, Luiz Alberto Beijo, and Fabricio Goecking Avelar. 2019. On modeling the maximum duration of dry spells: a simulation study under a bayesian approach. *Theoretical and Applied Climatology* 137:1337–1346.
- Camargo, Ângelo Paes de, and Marcelo Bento Paes de Camargo. 2001. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do brasil. *Bragantia* 60:65–68.
- Carter, Michael, and Tara Chiu. 2018. A minimum quality standard to ensure index insurance contracts do no harm.
- Carvalho, José Ruy Porto, Eduardo Delgado Assad, Silvio Roberto Medeiros Evangelista, and Hilton da Silveira Pinto. 2013. Estimation of dry spells in three brazilian regions—analysis of extremes. *Atmospheric Research* 132:12–21.
- Castro, P Neto, and E de A Vilela. 1986. Veranico: um problema de seca no periodo chuvoso. *Informe Agropecuario (Brasil)*.(Jun 12 (138): 59–62.
- Ceballos, Antonio, José Martinez-Fernández, and Miguel Ángel Luengo-Ugidos. 2004. Analysis of rainfall trends and dry periods on a pluviometric gradient representative of mediterranean climate in the duero basin spain. *Journal of Arid Environments* 58 (2): 215–233.
- CEPEA. 2022. Mercado de trabalho/cepea: pib-agro/cepea: pib do agro cresce 8,36% em 2021; participação no pib brasileiro chega a 27,4%. Accessed August 11, 2022. <https://bityli.com/TUWgAMY>.
- CNA. 2021. Mercado em foco - análise dos custos e benefícios da contratação de seguro do cafezal, <https://www.cnabrazil.org.br/assets/arquivos/dtec.mercado-em-foco5.seguro-cafe.VF.pdf>.

- Coles, Stuart. 2001. *An introduction to statistical modeling of extreme values*. Vol. 208. Springer.
- Coles, Stuart, and Luis Pericchi. 2003. Anticipating catastrophes through extreme value modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52 (4): 405–416.
- CONAB. 2021. Café conjuntura semanal, accessed July 22, 2022. <https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuario-e-extrativista/analises-do-mercado/historico-de-conjunturas-de-cafe>.
- Dalhaus, Tobias, Oliver Musshoff, and Robert Finger. 2018. Phenology information contributes to reduce temporal basis risk in agricultural weather index insurance. *Scientific reports* 8 (1): 1–10.
- Discacciati, Andrea, Nicola Orsini, and Sander Greenland. 2015. Approximate bayesian logistic regression via penalized likelihood by data augmentation. *The Stata Journal* 15 (3): 712–736.
- Draper, Norman R, and Harry Smith. 1998. *Applied regression analysis*. Vol. 326. John Wiley & Sons.
- Embrapa. 2009. Agrometeorologia dos cultivos: o fator meteorológico na produção agrícola, accessed May 30, 2023. https://www.embrapa.br/documents/1355291/37056285/Bases+climatol%5C%C3%5C%B3gicas_G.R.CUNHA_Livro_Agrometeorologia+dos+cultivos.pdf/13d616f5-cbd1-7261-b157-351eaa31188d?version=1.0.
- . 2020. A agricultura brasileira, accessed March 4, 2021. <https://www.embrapa.br/vii-plano-diretor/a-agricultura-brasileira>.
- . 2021. O agro brasileiro alimenta 800 milhões de pessoas, diz estudo da embrapa, accessed March 4, 2021. <https://www.embrapa.br/busca-de-noticias/-/noticia/59784047/o-agro-brasileiro-alimenta-800-milhoes-de-pessoas-diz-estudo-da-embrapa>.
- Fernandes, Antônio Alves Tôrres, Dalson Britto Figueiredo Filho, Enivaldo Carvalho da Rocha, and Willber da Silva Nascimento. 2021. Leia este artigo se você quiser aprender regressão logística. *Revista de Sociologia e Política* 28.
- Ferreira, Ana, and Laurens De Haan. 2015. On the block maxima method in extreme value theory: pwm estimators.
- Figueiredo, Rui, Mario LV Martina, David B Stephenson, and Benjamin D Youngman. 2018. A probabilistic paradigm for the parametric insurance of natural hazards. *Risk Analysis* 38 (11): 2400–2414.
- Fornazier, Armando, Paulo Marcelo Souza, and Nivaldo José Ponciano. 2012. A importância do seguro rural na redução de riscos da agropecuária. *Revista de Estudos Sociais* 14 (28): 39–52.
- Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2013. *Bayesian data analysis*. CRC press.
- Geweke, John F, et al. 1991. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Technical report. Federal Reserve Bank of Minneapolis.
- Gianola, Daniel, and Rohan L Fernando. 1986. Bayesian methods in animal breeding theory. *Journal of Animal Science* 63 (1): 217–244.
- Heidelberger, Philip, and Peter D Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31 (6): 1109–1144.
- Heimfarth, Leif Erec, and Oliver Musshoff. 2011. Weather index-based insurances for farmers in the north china plain: an analysis of risk reduction potential and basis risk. *Agricultural Finance Review* 71 (2): 218–239.
- Jeffreys, Harold. 1998. *The theory of probability*. OuP Oxford.
- Jenkinson, Arthur F. 1955. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81 (348): 158–171.
- Lichtenberg, Erik, and Eva Iglesias. 2022. Index insurance and basis risk: a reconsideration. *Journal of Development Economics* 158:102883. issn: 0304-3878. <https://doi.org/https://doi.org/10.1016/j.jdeveco.2022.102883>. <https://www.sciencedirect.com/science/article/pii/S0304387822000505>.
- Ljung, Greta M, and George EP Box. 1978. On a measure of lack of fit in time series models. *Biometrika* 65 (2): 297–303.
- Mann, Henry B. 1945. Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, 245–259.
- Martins, Everton, Lucas Eduardo de Oliveira Aparecido, Lucas Paulino S Santos, José Marcos Angélico de Mendonça, and Paulo Sergio de Souza. 2015. Influência das condições climáticas na produtividade e qualidade do café produzido na região do sul de minas gerais.

- Miller. 2020. Parametric insurance solutions", accessed August 2, 2022. https://www.miller-insurance.com/-/media/Bulletins/NA_bulletin/Nov-20/Parametric-insurance-solutions-from-Miller.pdf.
- O'Brien, Sean M, and David B Dunson. 2004. Bayesian multivariate logistic regression. *Biometrics* 60 (3): 739–746.
- Pai, Jeffrey, Yunxian Li, Aijun Yang, and Chenxu Li. 2022. Earthquake parametric insurance with bayesian spatial quantile regression. *Insurance: Mathematics and Economics* 106:1–12.
- Paulino, Carlos Daniel, Bento George C Murteira, and Maria Antônia Amaral Turkman. 2003. *Estatística bayesiana*. FCG.
- Picini, Angélica Giarolla, Marcelo Bento Paes de Camargo, Altino Aldo Ortolani, Luiz Carlos Fazuoli, and Paulo Boller Gallo. 1999. Desenvolvimento e teste de modelos agrometeorológicos para a estimativa de produtividade do cafeeiro. *Bragantia* 58:157–170.
- Prokopchuk, Olena, Ihor Prokopchuk, Grzegorz Mentel, and Yuriy Bilan. 2020. Parametric insurance as innovative development factor of the agricultural sector of economy. *AGRIS on-line Papers in Economics and Informatics* 12 (665-2022-413): 69–86.
- Radu, Nicoleta, and Felicia Alexandru. 2022. Parametric insurance—a possible and necessary solution to insure the earthquake risk of romania. *Risks* 10 (3): 59.
- Rafferty, Adrian E, S Lewis, JM Bernardo, JO Berger, AP Dawid, and AFM Smith. 1992. Bayesian statistics. *Oxford Sci. Publ*, 323–349.
- Ragulina, Galina, and Trond Reitan. 2017. Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the bayesian approach. *Hydrological Sciences Journal* 62 (6): 863–879.
- Reboita, Michelle Simões, Marcelo Rodrigues, Luiz Felipe Silva, and Maria Amélia Alves. 2015. Aspectos climáticos do estado de minas gerais. *Revista Brasileira de Climatologia* 17.
- Ruiz-Cardenas, Ramiro. 2015. A cafeicultura e sua relação com o clima, accessed May 30, 2023. <https://toolbox.coffeeandclimate.org/wp-content/uploads/Folheto-Cartilha-cafe-y-clima-Brasil.pdf>.
- Sengupta, Rohini, and Carolyn Kousky. 2020. Parametric insurance for disasters. *Wharton Risk Center Primer. Wharton Risk Center, Philadelphia*.
- Shirsath, Paresh, Shalika Vyas, Pramod Aggarwal, and Kolli N Rao. 2019. Designing weather index insurance of crops for the increased satisfaction of farmers, industry and the government. *Climate Risk Management* 25:100189.
- Skees, Jerry R, et al. 2012. The potential of parametric insurance solutions for managing natural disaster risk to reduce poverty and improve economic development in emerging. *Lexington (US):[sn]*.
- Soares, Lidiany dos Santos, Tiago Teruel Rezende, Luiz Alberto Beijo, and Kleso Silva Franco Júnior. 2021. Interaction between climate, flowering and production of dry coffee (coffea arabica l) in minas gerais. *Coffee Science* 16:e161786.
- Sum, Christine, Raghav Ohri, Shi An The, and Shyam Bhayani. 2020. Parametric landscape case studies, accessed November 25, 2020. <https://www.actuaries.org/sites/default/files/2020-12/20112507gicParametricInsurance.pdf>.
- Swaminathan, Hariharan, and Janice A Gifford. 1985. Bayesian estimation in the two-parameter logistic model. *Psychometrika* 50:349–364.
- Teh, Tse-Ling, and Christopher Woolnough. 2019. A better trigger: indices for insurance. *Journal of Risk and Insurance* 86 (4): 861–885.
- Turner, Stuart. 2020. Parametric insurance policies: do they have to be weather-related?, accessed September 1, 2020. <https://www.airmic.com/sites/default/files/Stuart-Turner-Risk-Leadership-Dissertation-White-Paper.pdf>.
- Wakefield, Jon. 2013. *Bayesian and frequentist regression methods*. Vol. 23. Springer.
- World Bank. 2005. Managing agricultural production risk : innovations in developing countries, accessed June 6, 2023. <https://openknowledge.worldbank.org/entities/publication/01aadfc3-86f1-587f-a06c-a2b8966ec273>.
- . 2017. Index insurance is having a development impact where it's needed most, accessed August 16, 2017. <https://blogs.worldbank.org/voices/index-insurance-having-development-impact-where-it-s-needed-most>.
- . 2023. Agriculture and food, accessed March 31, 2023. <https://www.worldbank.org/en/topic/agriculture/overview#:~:text=Agriculture%5C%20can%5C%20help%5C%20reduce%5C%20poverty,and%5C%20work%5C%20mainly%5C%20in%5C%20farming.>
- Zurich. 2021. Construction weather parametric insurance protecting your profits from extreme climate events, accessed August 9, 2021. <https://bityli.com/qAzxoW>.

5. CONSIDERAÇÕES FINAIS

Nesta pesquisa foi proposto um método para a estruturação de seguros agrícolas paramétricos que visem a proteção das lavouras contra a ocorrência de eventos climáticos extremos. O método proposto trouxe como alicerce a abordagem inferencial bayesiana. Esta abordagem possibilitou a quantificação das incertezas, por meio do estabelecimento de intervalos de credibilidade, nos vários níveis da tomada de decisão que envolvem o contrato de seguro paramétrico, auxiliando no processo decisório das seguradoras que elaboram este tipo de produto.

A realização do estudo de caso visando a aplicação do método proposto para a criação de um contrato de seguro paramétrico de proteção de lavouras de café contra a ocorrência de veranicos extremos na fase da florada, formação de chumbinhos e expansão dos frutos demonstrou resultados promissores. Isso porque, mesmo diante de uma série de dados limitados e com apenas 24 observações no total, foi possível o estabelecimento de um gatilho contratual com 87,5% de acurácia na previsão de perdas e que apresentou um custo razoável, tanto para a seguradora, como para o segurado.

A pesquisa demonstrou que a utilização de informação a *priori* para os parâmetros da distribuição GEV, melhorou o desempenho do modelo e tornou os intervalos de credibilidade de todas as variáveis estimadas mais precisas do que quando usada a distribuição a *priori* não informativa. Acredita-se que para os casos em que seja possível a obtenção de informações a *priori* para ser incorporada no modelo predictor de perdas os resultados se mostrarão ainda mais precisos.

Dessa forma, os procedimentos metodológicos propostos neste trabalho podem ser adaptados para a elaboração de contratos que tenham cobertura para diferentes culturas e eventos climáticos extremos, uma vez que se atuou na definição de gatilhos indenizatórios que compatibilizassem da melhor forma possível a ocorrência das perdas com o pagamento das indenizações nos contratos, reduzindo o risco base. Como sugestão para trabalhos futuros coloca-se a aplicação do método com a ampliação da base de dados e o uso de *prioris* informativas no modelo predictor de perdas.

REFERÊNCIAS

- AFSHAR, M. H. *et al.* Improving the performance of index insurance using crop models and phenological monitoring. **Remote Sens.**, v. 13, n. 5, p. 924-942, 2021.
- AGUIRRE, A. F. L.; NOGUEIRA, D. A.; BEIJO, L. A. Análise da temperatura máxima de Piracicaba (SP) via distribuição GEV não estacionária: uma abordagem bayesiana. **Revista Brasileira de Climatologia**, v. 27, p. 496–517, 2021.
- AHMAD, T. *et al.* A comprehensive study on the Bayesian modelling of extreme rainfall: A case study from Pakistan. **International Journal of Climatology**, v. 42, n. 1, p. 208-224, 2022.
- ALMEIDA, Gisele Carolina. **Uma abordagem Bayesiana para a modelagem dos ventos máximos de Sorocaba-SP e Bauru-SP**. 2018. 72 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, MG, 2018.
- ARRUDA, F. B. *et al.* Estudo da influência do clima e do consumo hídrico na produção de cafeeiros (*Coffea arabica* L.) em Pindorama, SP. **Bragantia**, v. 62, n.1, p. 782-785, 2000.
- APARECIDO, L. E. O.; ROLIM, G. S.; SOUZA, P. S. Sensitivity of newly transplanted coffee plants to climatic conditions at altitudes of Minas Gerais, Brazil. **Australian Journal of Crop Science**, Queensland, v.9, n. 2, p. 160-167, 2015.
- BARRETO, A. S. **Modelos de regressão: teoria e aplicações com o programa estatístico R**. Brasília: Edição do Autor, 2011.
- BAUTISTA, E. A. L.; ZOCCHI, S. S.; ANGELOCCI, L. R. A distribuição generalizada de valores extremos aplicada ao ajuste dos dados de velocidade máxima do vento em Piracicaba, São Paulo, Brasil. **Revista de Matemática e Estatística**, São Paulo - SP, v. 22, p. 95-111, 2004.
- BEIJO, L. A. **Construção de uma priori para os parâmetros do modelo de valores extremos generalizado baseada em quantis com distribuição Gumbell**. 2006. 98 f. Tese (Doutorado em Estatística Experimental e Agropecuária) - Universidade Federal de Lavras, Lavras, 2006.
- BOX, G. E. P.; TIAO, G. **Bayesian inference in statistical analysis**. New York: Wiley, 1992
- BUTTURI-GOMES, D.; BEIJO, L. A.; AVELAR, F. G. On modeling the maximum duration of dry spells: a simulation study under a Bayesian approach. **Theoretical and Applied Climatology**, v. 137, n. 1, p. 1337-1346, 2019.
- CAMARGO, M. B. P. de. The impact of climatic variability and climate change on arabic coffee crop in Brazil. **Bragantia**, Campinas, v. 69, p. 239-247, 2010.

CAMARGO, A. P.; CAMARGO, M. B. P. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. **Bragantia**, Campinas, v. 60, n. 1, p. 65-68, 2001.

CARTER, M.R., CHIU, T. A. Minimum quality standard to ensure index insurance contracts do no harm. **Feed the Future Innovation Lab for Assets and Market Access**, Davis, v.1, n.1, p.1-2, 2018.

CARVALHO *et al.* Estimation of dry spells in three Brazilian regions: analysis of extremes. **Atmospheric Research**, v. 132-133, n. 1, p. 12-21, 2013.

CEBALLOS, A.; MARTINEZ-FERNÁNDEZ, J.; LUENGO-UGIDOS, M. Á. Analysis of rainfall trends and dry periods on a pluviometric gradient representative of Mediterranean climate in the Duero Basin, Spain. **Journal of Arid Environments**, v. 58, n. 2, p. 215–233, 2004.

ESCOLA SUPERIOR DE AGRICULTURA LUIZ DE QUEIROZ. Centro de Estudos Avançados em Economia Aplicada. **Mercado de trabalho: PIB do agro cresce 8,36% em 2021 e participação no PIB brasileiro chega a 27,4%**. Piracicaba, 2022. Disponível em: <https://bitly.com/TUWgAMY>. Acesso em: 11 ago. 2022.

CHENG, Q; VARSHNEY, P. K.; ARORA, M. K. Logistic regression for feature selection and soft classification of remote sensing data. **IEEE Geoscience and Remote Sensing Letters**, v. 3, n. 4, p. 491-494, 2006.

COLES, S. G. **An introduction to statistical modeling of extreme values**. Bristol: Springer, 2001.

COMPANHIA NACIONAL DE ABSTECIMENTO (BRASIL). **Café: conjuntura semanal - de 2/8/21-6/8/21**. Brasília, 2021. Disponível em: <https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuario-e-extrativista/analises-do-mercado/historico-de-conjunturas-de-cafe>. Acesso em: 22 jul. 2022.

COMPANHIA NACIONAL DE ABSTECIMENTO (BRASIL). **Safra de café 2022 pode chegar a 53,4 milhões de sacas, impactada por clima adverso**. Brasília, 2021. Disponível em: https://cast.conab.gov.br/post/2022-05-19_2_lev_de_cafe/. Acesso em: 22 jul. 2022.

CONFEDERAÇÃO NACIONAL DA AGRICULTURA (BRASIL). **Mercado em foco - análise dos custos e benefícios da contratação de seguro do cafezal**. Brasília, 2021. Disponível em: <https://www.cnabrazil.org.br/assets/arquivos/dtec.mercado-em-foco5.seguro-cafe.VF.pdf>. Acesso em: 4 mar. 2021.

CORREIA FILHO, W. L. F.; ALVES, L. E. R.; DOS SANTOS, M. N. Caracterização de extremos anuais de precipitação para o estado de Sergipe. **Revista de Geografia (Recife)**, v. 34, n. 1, p. 49-70, 2017.

DA SILVA, R. S.; DO NASCIMENTO, F. F. Extreme Value Theory Applied to r Largest Order Statistics Under the Bayesian Approach. **Revista Colombiana de Estadística**, v. 42, n. 2, p. 143-166, 2019.

DALHAUS, T.; MUSSHOF, O.; FINGER, R. Phenology information contributes to reduce temporal basis risk in agricultural weather index insurance. **Scientific Reports**, v. 8, n. 1, p. 1-10, 2018.

DE OLIVEIRA, L. H. A.; CARVALHO, J. V. F. We are living on the edge! Gerenciando sinistros de extrema severidade com o apoio da teoria de valores extremos. *In*: USP INTERNATIONAL CONFERENCE IN ACCOUNTING, 20., 2020, São Paulo. **Anais eletrônicos** [...] São Paulo: Universidade de São Paulo, 2020. Disponível em: <https://congressosp.fipecafi.org/anais/Anais2020/ArtigosDownload/1902.pdf>. Acesso em: 22 jun. 2022.

DOBSON, A. J. **An introduction to generalized linear models**. 2nd ed. New York: CRC Press, 2002.

DOMÍNGUEZ-ALMENDROS, S.; BENÍTEZ-PAREJO, N.; GONZALEZ-RAMIREZ, A. R. Logistic regression models. **Allergologia e Immunopathologia**, v. 39, n. 5, p. 295-305, 2011.

DOS REIS, C. G. **Espacialização da probabilidade de ocorrência de veranicos extremos em regiões de Minas Gerais: uma abordagem Bayesiana**. 2020. 67 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Alfenas, Alfenas, MG, 2021.

EMBRAPA. **Agrometeorologia dos cultivos**: o fator meteorológico na produção agrícola. Brasília, Instituto Nacional de Meteorologia, 2009.

EMBRAPA. **O agro brasileiro alimenta 800 milhões de pessoas, diz estudo da Embrapa**. Brasília, 2021. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/59784047/o-agro-brasileiro-alimenta-800-milhoes-de-pessoas-diz-estudo-da-embrapa>. Acesso em: 4 mar. 2021.

FIGUEIREDO, R. *et al.* A probabilistic paradigm for the parametric insurance of natural hazards. **Risk Analysis**, v. 38, n. 11, p. 2400-2414, 2018.

FERNANDES, A. A. T. *et al.* Leia este artigo se você quiser aprender regressão logística. **Revista de Sociologia e Política**, v. 28, 2020.

FORNAZIER, A.; SOUZA, P. M.; PONCIANO, N. J. A Importância do Seguro Rural na Redução de Riscos da Agropecuária. **Revista de Estudos Sociais**, v. 14, n. 28, p. 39-52, 2012.

GELFAND, A. E. *et al.* Illustration of Bayesian inference in normal data models using Gibbs sampling. **Journal of the American Statistical Association**, Alexandria, v. 85, n. 410, p. 972-985, June 1990.

GELMAN, A. *et al.* **Bayesian data analysis**. London: Chapman Hall, 2003. 668 p.

GEWEKE, J. **Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion) in Bayesian statistics.** 4nd. ed. Oxford: Oxford University Press, 1992.

GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217-244, 1986.

GOOD, I. J. **Probability and the weighing of evidence.** London: Charles Griffin, 1950.

HEIMFARTH, L. E.; MUSSHOFF, O. Weather index- based insurances for farmers in the North China Plain: An analysis of risk reduction potential and basis risk. **Agricultural Finance Review**, v. 71, n. 2, p. 218-239, 2011.

HEIDELBERGER, P.; E WELCH, P. D. Simulation run length control in the presence of an initial transient. **Operations Research**, v. 31, p. 97-109, 1983.

JEFFREYS, H. **Theory of probability.** Oxford: Clarendon Press, 1998.

JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, v. 81, n. 348, p. 158-171, 1955.

JOHNSON, L. Paying ex gratia: Parametric insurance after calculative devices fail. **Geoforum**, v. 125, p. 120-131, 2021.

JÚLIO, W. S. *et al.* Identificação de períodos secos e chuvosos e a influência na produção de café de montanha nas matas de Minas. **CLIMEP - Climatologia e Estudos da Paisagem**, v. 9, n.1, p. 1-17, 2016.

LEMA, D. *et al.* Willingness to pay for weather-based index insurance in milk production. **Revista de Investigación en Modelos Financieros**, v. 1, p. 52-69, 2019.

LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297-303, 1978.

MANN, H. B. Nonparametric tests against trend. **Econometrica Journal of the Econometric Society**, v.13, n. 3, p. 245–259, 1945.

MARTINS, E. *et al.* Influência das condições climáticas na produtividade e qualidade do cafeeiro produzido na região do sul de Minas Gerais. **Coffee Science**, Lavras, v. 10, n. 4, p. 499 - 506, 2015.

MARTINS, T. B. *et al.* Predição da precipitação máxima no município de Silvanópolis-MG: Abordagens clássica e bayesiana. **IRRIGA**, v. 23, n. 3, p. 467-479, 2018.

MILLER. **Parametric insurance solutions:** guaranteed, objective and transparent protection. Londres, 2020. Disponível em: https://www.miller-insurance.com/-/media/Bulletins/NA_bulletin/Nov-20/Parametric-insurance-solutions-from-Miller.pdf . Acesso em: 02 ago. 2022.

MIQUELLUTI, D. L.; OZAKI, V. A. Identifying potential regions for a precipitation index insurance product in Paraná - Brazil: a hierarchical clustering approach. **Revista Brasileira de Climatologia**, v. 29, p. 78–98, 2021. Disponível em: <https://ojs.ufgd.edu.br/index.php/rbclima/article/view/15142>. Acesso em: 18 maio 2023.

MIQUELLUTI, D. L.; OZAKI, V. A.; MIQUELLUTI, D. J. An application of geographically weighted quantile LASSO to weather index insurance design. **Revista de Administração Contemporânea**, v. 26, n.3, p. 1-17, 2022.

NEWE SEGUROS. **Seguro paramétrico terá subvenção a partir de 2021 e promete ser mais barato.** Rio de Janeiro, 2020. Disponível em: <https://neweseguros.com.br/seguro-parametrico-tera-subvencao-a-partir-de-2021-e-promete-ser-mais-barato/>. Acesso em: 08 set. 2021.

NOGUEIRA, D.; SÁFADI, T.; FERREIRA, D. Avaliação de critérios de convergência univariados para o método de Monte Carlo via Cadeias de Markov. **Revista Brasileira de Estatística**, v. 65, n. 224, 2004.

OZAKI, V. A. O papel do seguro na gestão do risco agrícola e os empecilhos para o seu desenvolvimento. **Revista Brasileira de Risco e Seguro**, v. 2, n. 4, p. 75-92, 2007.

PAMPEL, F. C. **Logistic regression:** a primer. Thousand Oaks: Sage Publications, 2000.

PAULINO, D. C.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística Bayesiana.** Lisboa: FCG, 2003.

PICINI, A. G. *et al.* Desenvolvimento e teste de modelos agrometeorológicos para a estimativa de produtividade do cafeeiro. **Bragantia**, Campinas, v. 58, n. 1, p. 157-170, 1999.

PROKOPCHUK, O. *et al.* Parametric insurance as innovative development factor of the agricultural sector of economy. **AGRIS on-line Papers in Economics and Informatics**, v. 12, n. 3, p. 69-86, 2020.

RADU, N.; ALEXANDRU, F. Parametric Insurance: A Possible and Necessary Solution to Insure the Earthquake Risk of Romania. **Risks**, v. 10, n. 3, p. 59-75, 2022.

RAFTERY, A. E.; LEWIS, S. **How many iterations in the Gibbs sampler, in Bayesian Statistics.** 4. ed. Oxford: Oxford University, 1992.

RUIZ-CÁRDENAS, R. *et al.* **A cafeicultura e sua relação com o clima**. Lavras: Edição do autor, 2015. Disponível em: <https://toolbox.coffeeandclimate.org/wp-content/uploads/Folheto-Cartilha-cafe-y-clima-Brasil.pdf>. Acesso em: 25 jul. 2022

SHIRSATH, P. *et al.* Designing weather index insurance of crops for the increased satisfaction of farmers, industry and the government. **Climate Risk Management**, v. 25, n.1, p. 100189-100201, 2019.

SOARES, S. S. *et al.* Interaction between climate, flowering and production of dry coffee (*Coffea arabica* L.) in Minas Gerais. **Coffee Science**, v. 16, p.e161786, 2021.

SOUZA, S. **Seguros: contabilidade, atuária e auditoria**. São Paulo: Saraiva, 2007.

SUN, C. *et al.* **The Parametric Landscape & Case Studies**. Singapura: Institute and Faculty of Actuaries, 2020. Disponível em: <https://www.actuaries.org.sg/sites/default/files/2020-12/20112507gicParametricInsurance.pdf>. Acesso em: 17 ago. 2022.

TAVARES, M. G.; SANTOS, E. B. Período de retorno das precipitações máximas diárias no estado do Rio de Janeiro. **Revista Brasileira de Climatologia**, v. 30, p. 274-290, 2022.

TURNER, S. **Parametric insurance Policies: Do they have to be weather-related?** Armic: White Paper, 2020. Disponível em: <https://www.airmic.com/sites/default/files/Stuart-Turner-Risk-Leadership-Dissertation-White-Paper.pdf>. Acesso em: 01 ago. 2022.

WALENDORFF, R. Seguro paramétrico inédito protege cacau no sul da Bahia. **Valor Econômico**, 17 ago. 2021. Disponível em: <https://valor.globo.com/agronegocios/noticia/2021/08/17/seguro-parametrico-inedito-protege-cacau-no-sul-da-bahia.ghtml>. Acesso em: 01 ago. 2022.

WAKEFIELD, J. **Bayesian and frequentist regression methods**. New York: Springer, 2013.

WORLD BANK. **Index insurance is having a development impact where it's needed most**. Washington, 2017. Disponível em: <https://blogs.worldbank.org/voices/index-insurance-having-development-impact-where-it-s-needed-most>. Acesso em: 01 ago. 2022.

ZURICH. **Zurich Construction Weather Parametric Insurance Protecting your profits from extreme climate events**. Berna, 2021. Disponível em: <https://bityli.com/qAzxoW>. Acesso em: 08 set. 2021.

APÊNDICE A – script do código R para o ajuste das distribuições GEV, considerando a distribuição a priori informativa

```

set.seed(1)

#Pacotes necessários
library(readxl)
library(evd)
library(R2OpenBUGS)
library(readxl)
library(Kendall)
library(coda)
library(rstanarm)
library(plotrix)
library(ggplot2)
library(hnp)
library(plotly)
library(patchwork)
library(magrittr)

##### Importando os dados #####
extremos <- dados referentes ao maior veranico observado em R1 e o
vetor perda
guaxupe <- série de máximos de segunda ordem dos veranicos máximos
observados em Guaxupé no período R1
msm <- série de máximos de segunda ordem dos veranicos máximos
observados em Monte Santo de Minas no período R1
mc <- série de máximos de segunda ordem dos veranicos máximos
observados em Monte Carmelo no período R1

##### Ajuste da GEV #####

# Estabelecendo os parâmetros da distribuição à priori
priori <- fgev(muzambinho$R1);priori #definir a média
priori$var.cov #definir a variância (usei a variância flexionada em
3x)

## Criando o modelo para gerar as cadeias posteriori da GEV

sink("Gev_RII.txt")
cat("
  model {

    # Verossimilhança
    for (i in 1:n) {
      y[i] ~ dgev(mu, sigma, eta)
    }

    # Prior
    mu ~ dnorm(16.06,0.3022408) # dnorm(media, (1/(3*variância)))
  }

```

```

    sigma ~ dlnorm(1.732165,1.719625) #
dlnorm(log(media),1/(log(3*variância)))
    eta ~ dnorm(-0.05865, 15.23657) # dnorm(media, (1/(3*variância))
}
    ",fill=TRUE)
sink()

#Guaxupé
dados_bug_guaxupe<-
list(y=guaxupe$R1,n=length(guaxupe$R1));dados_bug_guaxupe
g<-fgev(guaxupe$R1) #Estimativas de verossimilhança

inits_guaxupe <- function(){ list(mu=g$estimate[1],
sigma=g$estimate[2], eta=g$estimate[3])}

params_guaxupe <- c("mu","sigma", "eta")
nc = 3 #Numero de cadeias
ni = 8000 #Tamanho da cadeia quase final
nb = 3000 #Numero de simulação que serão descartadas
nt = 4#Thinning rate

# Iniciando o Amostrador para a geração das cadeias posteriori dos
parâmetros da GEV

gev.bayes_guaxupe = bugs(data = dados_bug_guaxupe, inits =
inits_guaxupe,
                        parameters =c(params_guaxupe),
                        model = "Gev_RII.txt",
                        n.thin = nt, n.chains = nc,
                        n.burnin = nb, n.iter = ni, codaPkg=FALSE,
debug=T)
print(gev.bayes_guaxupe, dig = 4) #Mostra as estimativas médias para
os parâmetros e níveis de retorno

post_gb_guaxupe<-as.mcmc(gev.bayes_guaxupe$sims.matrix[,]) # salva a
saída como cadeia mcmc

#Intervalo de credibilidade HPD
HPDinterval(post_gb_guaxupe)

#Critérios de Convergência das cadeias posteriori dos parâmetros e
níveis de retorno
geweke.diag(post_gb_guaxupe) # Geweke
raftery.diag(post_gb_guaxupe) # Raftery & Lewis
heidel.diag(post_gb_guaxupe) # Heidelberger & Welch

### Verificando as probabilidades de ocorrência para cada quantidade
de dias sem chuva
e_mu_guaxupe <- unlist(gev.bayes_guaxupe$sims.matrix[,1]) #Valor
médio do parâmetro posicao da GEV
e_sig_guaxupe <- unlist(gev.bayes_guaxupe$sims.matrix[,2]) #Valor
médio do parâmetro escala da GEV
e_shape_guaxupe <- unlist(gev.bayes_guaxupe$sims.matrix[,3]) #Valor
médio do parâmetro forma da GEV

```

```

#### Verificando a probabilidade para os valores extremos observados
extremos_guaxupe <- extremos$R1[1:8] #Valores extremos de estiagem
observados no ciclo R1
prob_guaxupe_gev <-
matrix(,length(e_mu_guaxupe),length(extremos_guaxupe))
for(i in 1:length((e_mu_guaxupe))){
  mu_g <- gev.bayes_guaxupe$sims.matrix[i,1]
  sig_g <- gev.bayes_guaxupe$sims.matrix[i,2]
  eta_g <- gev.bayes_guaxupe$sims.matrix[i,3]
  prob_guaxupe_gev[i,1] <- pgev(extremos_guaxupe[1],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
  prob_guaxupe_gev[i,2] <- pgev(extremos_guaxupe[2],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
  prob_guaxupe_gev[i,3] <- pgev(extremos_guaxupe[3],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
  prob_guaxupe_gev[i,4] <- pgev(extremos_guaxupe[4],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
  prob_guaxupe_gev[i,5] <- pgev(extremos_guaxupe[5],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
  prob_guaxupe_gev[i,6] <- pgev(extremos_guaxupe[6],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
  prob_guaxupe_gev[i,7] <- pgev(extremos_guaxupe[7],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
  prob_guaxupe_gev[i,8] <- pgev(extremos_guaxupe[8],loc= mu_g,
scale=sig_g, shape=eta_g, lower.tail=FALSE)
}

#Monte Santo de Minas
dados_bug_msm<- list(y=msm$R1,n=length(msm$R1));dados_bug_msm
g<-fgev(msm$R1) #Estimativas de verossimilhança

inits_msm <- function(){ list(mu=g$estimate[1], sigma=g$estimate[2],
eta=g$estimate[3])}

params_msm <- c("mu","sigma", "eta")
nc = 3 #Numero de cadeias
ni = 8000 #Tamanho da cadeia quase final
nb = 3000 #Numero de simulação que serão descartadas
nt = 4#Thinning rate

# Iniciando o Amostrador para a geração das cadeias posteriori dos
parâmetros da GEV

gev.bayes_msm = bugs(data = dados_bug_msm, inits = inits_msm,
parameters =c(params_msm),
model = "Gev_RII.txt",
n.thin = nt, n.chains = nc,
n.burnin = nb, n.iter = ni, codaPkg=FALSE,
debug=T)

print(gev.bayes_msm, dig = 4) #Mostra as estimativas médias para os
parâmetros e níveis de retorno

post_gb_msm<-as.mcmc(gev.bayes_msm$sims.matrix[,]) # salva a saída
como cadeia mcmc

```

```

#Intervalo de credibilidade HPD
HPDinterval(post_gb_msm)

#Critérios de Convergência das cadeias posteriori dos parâmetros e
níveis de retorno
geweke.diag(post_gb_msm) # Geweke
raftery.diag(post_gb_msm) # Raftery & Lewis
heidel.diag(post_gb_msm) # Heidelberger & Welch

### Verificando as probabilidades de ocorrência para cada quantidade
de dias sem chuva
e_mu_msm <- unlist(gev.bayes_msm$sims.matrix[,1]) #Valor médio do
parâmetro posicao da GEV
e_sig_msm <- unlist(gev.bayes_msm$sims.matrix[,2]) #Valor médio do
parâmetro escala da GEV
e_shape_msm <- unlist(gev.bayes_msm$sims.matrix[,3]) #Valor médio do
parâmetro forma da GEV

#### Verificando a probabilidade para os valores extremos observados
extremos_msm <- extremos$R1[9:16] #Valores extremos de estiagem
observados no ciclo R1
prob_msm_gev <- matrix(,length(e_mu_msm),length(extremos_msm))
for(i in 1:length((e_mu_msm))){
  mu_msm <- gev.bayes_msm$sims.matrix[i,1]
  sig_msm <- gev.bayes_msm$sims.matrix[i,2]
  eta_msm <- gev.bayes_msm$sims.matrix[i,3]
  prob_msm_gev[i,1] <- pgev(extremos_msm[1],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
  prob_msm_gev[i,2] <- pgev(extremos_msm[2],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
  prob_msm_gev[i,3] <- pgev(extremos_msm[3],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
  prob_msm_gev[i,4] <- pgev(extremos_msm[4],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
  prob_msm_gev[i,5] <- pgev(extremos_msm[5],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
  prob_msm_gev[i,6] <- pgev(extremos_msm[6],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
  prob_msm_gev[i,7] <- pgev(extremos_msm[7],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
  prob_msm_gev[i,8] <- pgev(extremos_msm[8],loc= mu_msm,
scale=sig_msm, shape=eta_msm, lower.tail=FALSE)
}

#Monte Carmelo
dados_bug_mc<- list(y=mc$R1,n=length(mc$R1));dados_bug_mc
g<-fgev(mc$R1) #Estimativas de verossimilhança

inits_mc <- function(){ list(mu=g$estimate[1], sigma=g$estimate[2],
eta=g$estimate[3])}

params_mc <- c("mu","sigma", "eta")
nc = 3 #Numero de cadeias
ni = 8000 #Tamanho da cadeia quase final

```

```

nb = 3000 #Numero de simulação que serão descartadas
nt = 4#Thinning rate

# Iniciando o Amostrador para a geração das cadeias posteriori dos
parâmetros da GEV

gev.bayes_mc = bugs(data = dados_bug_mc, inits = inits_mc,
                    parameters =c(params_mc),
                    model = "Gev_RII.txt",
                    n.thin = nt, n.chains = nc,
                    n.burnin = nb, n.iter = ni, codaPkg=FALSE,
debug=T)

print(gev.bayes_mc, dig = 4) #Mostra as estimativas médias para os
parâmetros e níveis de retorno

post_gb_mc<-as.mcmc(gev.bayes_mc$sims.matrix[,]) # salva a saída
como cadeia mcmc

#Intervalo de credibilidade HPD
HPDinterval(post_gb_mc)

#Critérios de Convergência das cadeias posteriori dos parâmetros e
níveis de retorno
geweke.diag(post_gb_mc) # Geweke
raftery.diag(post_gb_mc) # Raftery & Lewis
heidel.diag(post_gb_mc) # Heidelberger & Welch

### Verificando as probabilidades de ocorrência para cada quantidade
de dias sem chuva
e_mu_mc <- unlist(gev.bayes_mc$sims.matrix[,1]) #Valor médio do
parâmetro posicao da GEV
e_sig_mc <- unlist(gev.bayes_mc$sims.matrix[,2]) #Valor médio do
parâmetro escala da GEV
e_shape_mc <- unlist(gev.bayes_mc$sims.matrix[,3]) #Valor médio do
parâmetro forma da GEV

#### Verificando a probabilidade para os valores extremos observados
extremos_mc <- extremos$R1[17:24] #Valores extremos de estiagem
observados no ciclo mc
prob_mc_gev <- matrix(,length(e_mu_mc),length(extremos_mc))
for(i in 1:length((e_mu_mc))){
  mu_mc <- gev.bayes_mc$sims.matrix[i,1]
  sig_mc <- gev.bayes_mc$sims.matrix[i,2]
  eta_mc <- gev.bayes_mc$sims.matrix[i,3]
  prob_mc_gev[i,1] <- pgev(extremos_mc[1],loc= mu_mc, scale=sig_mc,
shape=eta_mc, lower.tail=FALSE)
  prob_mc_gev[i,2] <- pgev(extremos_mc[2],loc= mu_mc, scale=sig_mc,
shape=eta_mc, lower.tail=FALSE)
  prob_mc_gev[i,3] <- pgev(extremos_mc[3],loc= mu_mc, scale=sig_mc,
shape=eta_mc, lower.tail=FALSE)
  prob_mc_gev[i,4] <- pgev(extremos_mc[4],loc= mu_mc, scale=sig_mc,
shape=eta_mc, lower.tail=FALSE)
  prob_mc_gev[i,5] <- pgev(extremos_mc[5],loc= mu_mc, scale=sig_mc,
shape=eta_mc, lower.tail=FALSE)
}

```

```
    prob_mc_gev[i,6] <- pgev(extremos_mc[6],loc= mu_mc, scale=sig_mc,  
shape=eta_mc, lower.tail=FALSE)  
    prob_mc_gev[i,7] <- pgev(extremos_mc[7],loc= mu_mc, scale=sig_mc,  
shape=eta_mc, lower.tail=FALSE)  
    prob_mc_gev[i,8] <- pgev(extremos_mc[8],loc= mu_mc, scale=sig_mc,  
shape=eta_mc, lower.tail=FALSE)  
}
```

APÊNDICE B – script do código em R para o ajuste dos modelos logísticos

```
##### Regressão Logística #####
vetor_perda_obs # 1 é "perda" e 0 é "não perda"
cidade

#Organizando variável explicativa dos modelos logísticos

#Guaxupé
dados_g <-
matrix(,length(prob_guaxupe_gev[1,]),length(prob_guaxupe_gev[,1]))
for(i in 1:length(prob_guaxupe_gev[,1])){
  dados_g[,i] <- prob_guaxupe_gev[i,]
}
dados_g[,1:3]

#MSM
dados_msm <-
matrix(,length(prob_msm_gev[1,]),length(prob_msm_gev[,1]))
for(i in 1:length(prob_msm_gev[,1])){
  dados_msm[,i] <- prob_msm_gev[i,]
}
dados_msm[,1:3]

#MC
dados_mc <- matrix(,length(prob_mc_gev[1,]),length(prob_mc_gev[,1]))
for(i in 1:length(prob_mc_gev[,1])){
  dados_mc[,i] <- prob_mc_gev[i,]
}
dados_mc[,1:3]

#Organizando dados da regressão

dados_R1 <-
matrix(,length(extremos$L[1:24]),length((prob_guaxupe_gev[,1])))
dim(dados_R1)
dados_R1[1:8,] <- dados_g
dados_R1[9:16,] <- dados_msm
dados_R1[17:24,] <- dados_mc

dadosR1 <- data.frame(cidade, vetor_perda_obs, dados_R1) #Base de
dados da regressão
View(dadosR1)

#Modelo logístico bayesiano
sink("log.simples_b.txt")
cat("
model{
  for( i in 1 : N ) {
    L[i] ~ dbern(p[i])
    logit(p[i]) <- beta0 + beta1*prob[i]
  }
}
```



```

beta0 ~ dnorm(0,0.0001)
beta1 ~ dnorm(0,0.0001)

#Previsoes
for( i in 1 : N ) {
  previsoes[i] <-
exp(beta0+beta1*prob[i]) / (1+exp(beta0+beta1*prob[i]))
}
}

",fill=TRUE)
sink()

prob <- dadosR1[,3: ncol(dadosR1)]
L <- dadosR1[,2]
N = length(L)
data = list("prob","L","N")

inits = function(){
  list(beta0=1,beta1=1)
}
theta<-c("beta0","beta1")

estimativas <- matrix(,ncol(dadosR1)-2,26) #cada linha representa um
modelo e cada coluna representa uma variável prevista

#15 mil é a quantidade de simulações realizadas
#26 colunas porque são dois betas estimados e 24 previsões

# Rodando o modelo logístico
for (i in 1:(ncol(dadosR1)-2)){
  prob <- dadosR1[,3:(ncol(dadosR1))]
  prob <- prob[,i]
  L <- dadosR1[,2]
  N = length(L)
  data = list("prob","L","N")
  logistica = bugs(data,inits,
                  model.file="log.simples_b.txt",
                  parameters=c(theta,"previsoes"),
                  n.chains=3,n.iter=8000,n.burnin=3000,n.thin=4,
                  codaPkg=F, debug=F)

  for (k in 1:2){
    a <- unlist(logistica$mean[k]) #sao os betas
    estimativas[i,k] <- a
  }
  for (j in 1:24){
    b <- unlist(logistica$mean$previsoes[j])
    estimativas[i,j+2] <- b
  }
  if ((i%100!=0)==FALSE){
    print(i)
  }
}
}

```