

PRODUTO 10

RELATORIO FINAL

Contratada:

Escola Superior de Agricultura Luiz de Queiroz – ESALQ
Fundação de Estudos Agrários Luiz de Queiroz – FEALQ

Execução:



Contratante:



Inversión Pública & Adaptación al
Cambio Climático en América Latina

SUMARIO

1. INTRODUÇÃO GERAL	5
2. RECOMENDAÇÃO DE DATA DE SEMEADURA BASEADA EM INDICADOR TEMPORAL DE PRODUTIVIDADE E CENÁRIOS DE MUDANÇAS CLIMÁTICAS ...	10
2.1 INTRODUÇÃO	10
2.2 CENÁRIOS CLIMÁTICOS E DE CONCENTRAÇÃO DE CARBONO	12
2.3 RESOLUÇÃO ESPACIAL	15
2.4 SIMULAÇÕES COM OS MODELOS AGRÍCOLAS	16
2.5 RESULTADOS.....	17
2.6 CONSIDERAÇÕES FINAIS	19
3. SEÇÃO 2 - DETERMINAÇÃO DAS ZONAS HOMOGÊNEAS DE RISCO - ESTRATIFICAÇÃO MUNICIPAL DO RISCO BASEADO EM MULTIMODELOS AGROMETEOROLÓGICOS E CENÁRIOS DE MUDANÇAS CLIMÁTICAS.....	20
3.1 INTRODUÇÃO	20
3.2 ANÁLISE DE RISCO CLIMÁTICO.....	21
3.3 RESULTADOS.....	22
3.4 CONSIDERAÇÕES FINAIS.....	28
4. SEÇÃO 3- PROBABILIDADE DE PERDA – ESTRATIFICAÇÃO MUNICIPAL DO RISCO BASEADO EM MODELOS ATUARIAIS	29
4.1 INTRODUÇÃO	29
4.2 A PROBABILIDADE DE PERDA	29
4.3 A DISTRIBUICAO NORMAL ASSIMÉTRICA.....	30
4.4 A DISTRIBUICAO SKEW-T	31
4.5 A DISTRIBUICAO ODD LOG-LOGÍSTICA	32
4.6 ESCOLHA DE MODELOS	33
4.7 VARIÁVEIS GERADAS PELOS MODELOS DSSAT E APSIM.....	34
4.8 SOBRE AS TABELAS	36
4.8.1 ARQUIVOS .TXT DAS PROBABILIDADES DE PERDA, MODELO DSSAT ..	36
4.8.2 ARQUIVOS .TXT DAS PROBABILIDADES DE PERDA, MODELO APSIM... 38	
4.8.3 SOBRE OS MAPAS, MODELOS DSSAT E APSIM	38
4.9 OBSERVAÇÕES FINAIS.....	40
4.10 REFERÊNCIAS	41
5. SEÇÃO 4 - ÍNDICE DE PERDA E ESTRATIFICAÇÃO MUNICIPAL DO RISCO BASEADO EM ÍNDICES DE PERDA.....	42
5.1 INTRODUÇÃO	42
5.2 O ÍNDICE DE PERDA	42
5.3 VARIÁVEIS GERADAS PELOS MODELOS DSSAT E APSIM.....	42
5.4 SOBRE AS TABELAS	45

5.4.1	ARQUIVOS .TXT DOS ÍNDICES DE PERDA, MODELO DSSAT	45
5.4.2	ARQUIVOS .TXT DOS ÍNDICES DE PERDA, MODELO APSIM.....	46
5.4.3	SOBRE OS MAPAS, MODELOS DSSAT E APSIM	47
5.5	OBSERVAÇÕES FINAIS.....	56
6.	SEÇÃO 5 - PROPOSTA DE METODOLOGIA DE PRECIFICAÇÃO.....	57
6.1	INTRODUÇÃO	57
6.2	CONTEXTUALIZAÇÃO	57
6.3	CARACTERIZAÇÃO DOS PRODUTOS DE SEGURO AGRÍCOLA E PROAGRO 62	
6.4	METODOLOGIA ATUARIAL DE PRECIFICAÇÃO DE RISCO.....	65
6.5	ASPECTOS ATUARIAIS DO SEGURO AGRÍCOLA	67
6.6	ESTUDOS ATUARIAIS APLICADOS AO BRASIL	73
6.7	A DISTRIBUICAO NORMAL ASSIMÉTRICA.....	76
6.8	A DISTRIBUICAO SKEW-T	78
6.9	A DISTRIBUICAO NÃO-PARAMÉTRICA.....	83
6.10	CRITÉRIOS PARA ESCOLHA DE DISTRIBUIÇÕES	85
6.11	PRECIFICAÇÃO DO SEGURO DE FATURAMENTO: MODELANDO A PRODUTIVIDADE E O PREÇO CONJUNTAMENTE	87
6.12	CÓPULAS	89
6.13	INFERÊNCIA DAS CÓPULAS.....	91
6.14	MEDIDAS DE AJUSTE DOS MODELOS DE CÓPULAS.....	92
6.15	CONSIDERAÇÕES FINAIS.....	98
6.16	REFERÊNCIAS	101
7.	SEÇÃO 6 - ANÁLISE CONTRAFACTUAL – CONSIDERANDO O EFEITO DA MUDANÇA CLIMÁTICA	106
7.1	INTRODUÇÃO	106
7.2	NOTA METODOLÓGICA	107
7.3	SOBRE OS MAPAS	111
7.4	REPRESENTAÇÃO TABULAR DO ÍNDICE DE PERDA.....	120
8.	SEÇÃO 7 - DESCRIÇÃO METODOLÓGICA PARA OBTENÇÃO DOS PRODUTOS 2 A 7 (SEÇÕES 1 A 6 DESTE RELATÓRIO).....	127
8.1	INTRODUÇÃO	127
8.2	NOTA EXPLICATIVA.....	128
8.3	NOTA METODOLÓGICA DO PRODUTO 2.....	128
8.4	NOTA METODOLÓGICA DO PRODUTO 3.....	139
8.5	NOTA METODOLÓGICA DO PRODUTO 4.....	141
8.6	NOTA METODOLÓGICA DO PRODUTO 5.....	144
8.7	NOTA METODOLÓGICA DO PRODUTO 6.....	144

8.8	NOTA METODOLÓGICA DO PRODUTO 7.....	146
9.	FLUXOGRAMA DO PROJETO	149
10.	ANEXO 1 - CÓDIGOS EM R REFERENTES AO SEÇÃO 5 (PRODUTO 6).....	151
11.	ANEXO RESUMO DAS Etapas da metodologia de precificação (SEÇÃO 5 – PRODUTO 6).....	162
11.1	TRATAMENTO PRELIMINAR DOS DADOS	163
11.2	AJUSTE E TESTES DAS DISTRIBUIÇÕES DE PROBABILIDADE.....	166
11.3	CÁLCULO DA PROBABILIDADE DE PERDA.....	171
11.4	CÁLCULO DA TAXA DE PRÊMIO	172

1. INTRODUÇÃO GERAL

No Plano Agrícola e Pecuário da safra 2018-2019 foram anunciados aproximadamente R\$ 195 bilhões que serão direcionados ao setor na forma de crédito, seguro rural e apoio à comercialização. Há muitos anos, a política agrícola está sedimentada na equalização de juros com recursos do Tesouro Nacional. Por outro lado, as demais políticas, seguro rural e apoio à comercialização, ficaram à margem quando se compara o total de gastos do Governo.

A política de crédito rural, baseada na equalização da taxa de juros, foi importante em momentos de altas taxas de juros. Por meio desta, o setor se modernizou, houve aumento do PIB setorial, criação de empregos, geração de divisas, diversificação da pauta de produtos exportados, entre vários outros fatores positivos. Porém, após décadas de juros altos, o país passa por um momento de estabilidade macroeconômica, com taxas de inflação controladas e o menor patamar de juros dos últimos anos.

Nesse contexto, questiona-se a eficácia do uso de recursos públicos na política de equalização de juros em detrimento da política de seguro rural, essencial para a proteção do produtor contra a ocorrência de intempéries climáticas ou reduções de sua renda, e que ainda possui um orçamento baixo ante a sua demanda.

A baixa massificação do seguro rural compromete a recuperação financeira do produtor em caso de quebra de safra ocasionada por eventos climáticos adversos. Há necessidade de renegociar as dívidas com o agente financeiro geralmente por longos anos reduzindo, dessa forma, sua capacidade de reinvestimento. Na prática, o produtor rural tem acesso ao crédito com juros equalizados, em troca de altas garantias junto ao agente financeiro e pouca ou nenhuma proteção contra eventos climáticos adversos.

Apesar da elevada participação dos bancos públicos na concessão do crédito rural, observa-se, nos últimos três anos, um aumento da participação dos agentes financeiros

privados no “*funding*” do agronegócio, bem como novas seguradoras interessadas em participar do programa de subvenção ao prêmio do seguro rural (PSR). Evidencia-se um novo cenário para a condução da política agrícola, com maior participação privada e apoio ao seguro rural frente a redução de gastos com a equalização de juros do crédito rural.

Essa nova tendência de mercado vem associado a necessidade de entender melhor os riscos associados às operações de crédito e seguro rural. Dessa forma, para mensurar os riscos e reduzir as perdas de uma multiplicidade de culturas agrícolas, o Governo realiza um estudo de zoneamento de risco climático, com base em um indicador que expressa a condição hídrica da cultura (índice de satisfação de necessidade de água – ISNA), e a disponibiliza ao mercado.

Apesar de ter sua importância, o ZARC possui limitações. A principal delas refere-se ao uso de apenas um indicador climático (ISNA) na modelagem agrometeorológica. Embora seja importante, não é o único risco coberto pelas seguradoras. Nesse contexto, o projeto intitulado “Investimento Público e Adaptação à Mudança do Clima na América Latina (IPACC II)”, PN: 2014.9049.9-002, tem por objetivo apresentar uma metodologia alternativa aquela utilizada pelo ZARC.

A grande contribuição do projeto é levar em conta diversos fatores climáticos e resultar em uma estimativa de produtividade que pode ser utilizada no lugar do ISNA. Em outras palavras, para efeitos de quantificação de perdas, considera-se não a sua causa, mas modela-se diretamente a consequência do efeito climático na lavoura, mensurado pela produtividade agrícola.

Ademais, o projeto extrapolou a abordagem climatológica e incluiu no escopo fatores atuariais. A vantagem de trabalhar com modelos atuariais é a possibilidade de gerar resultados que podem ser diretamente aplicados e entendidos pelo mercado segurador. Dessa forma, gerou-se quatro tipos de produtos, complementares ao ZARC, que podem ser

utilizados por instituições financeiras, seguradoras e resseguradoras em suas operações de crédito e seguro rural.

Os produtos contemplam recomendações de épocas de plantio e zonas homogêneas de risco para a soja, no estado do Paraná, e que poderá ser ampliada para outros estados e culturas agrícolas. Uma das preocupações ao longo do projeto foi criar algoritmos que pudessem ser rapidamente replicados para outras culturas e regiões do país, com uma metodologia inovadora e confiável, gerando resultados com escalabilidade e acurácia.

Além da criação de quatro novos produtos de zoneamento de risco climático-atuarial, o projeto abordou a questão das mudanças climáticas e o impacto desse fenômeno nos resultados dos produtos criados. Para isso, considerou-se a seleção dos Modelos de Circulação Global (GCMs) baseados no protocolo proposto pelo Projeto AgMIP que, por sua vez, vale-se do conceito de Representative Climate Futures (RCF). Além disso, dado que as projeções climáticas de longo prazo carregam incertezas e dado que a finalidade de estudos de vulnerabilidade climática como no presente projeto é a geração de cenários agrícolas futuros, a abordagem RCF é suficientemente robusta para permitir a avaliação de cenários futuros para a cultura da soja para o Estado do Paraná. Com base nesta abordagem, foram gerados um cenário otimista (com o GCM de menor variação em relação a linha de base), intermediários e pessimistas (com base no GCM de maior variação em relação a linha de base) de produtividade futura da cultura da soja.

A variável de verificação utilizada como referência para a geração dos RCFs foi a temperatura do ar, por se tratar de uma variável climática que expressa a quantidade de energia no ambiente para a realização dos processos atmosféricos e que está fortemente correlacionada com a chuva e com a própria irradiância solar global. Vale lembrar que a temperatura, radiação solar e chuva são as variáveis de entrada para os modelos agrícolas utilizados neste projeto (DSSAT e APSIM, descritos mais a frente neste documento).

Assim, com base na abordagem RCF e tendo a temperatura do ar como variável-critério, a partir de uma análise das bases de dados do Paraná foram selecionados três GCMs, representado, como já mencionado, o modelo com a maior variação de temperatura em relação a linha de base, o modelo com a menor variação e, por fim, um modelo utilizando uma variação intermediária entre dois anteriores, dentre os 20 disponíveis AGMIP.

Além desses três GCMs, incluiu-se um quarto modelo (HadGem2). Dessa forma, trabalhou-se com quatro bases de dados climáticos originados dos seguintes GCM's:

1. GCM de variação mínima em relação a LB: MIROC5;
2. GCM de variação média em relação a LB: BNU-ESM;
3. GCM de variação máxima em relação a LB: CanESM2.
4. GCM adicional: HadGem2

Para cada um dos 4 GCMs selecionados, considerou-se 2 níveis de concentração de CO₂ atmosférico, de modo a representar as forçantes radioativas descritas nos Representative Concentration Pathways (RCPs) 4.5 e 8.5. Assim, para efeito de simulação, foram assumidos os níveis de concentrações de CO₂ de 526 ppm (RCP 4.5) e 628 ppm (RCP 8.5), constituindo-se dessa forma, para cada local selecionado, 9 bases climáticas (1 atual e 8 futuros).

O quinto e último produto gerado no projeto foi a proposta de precificação de risco visando propor melhorias na tarifação do Programa de Garantia da Atividade Agropecuária (PROAGRO). No Brasil, até 2016, o PROAGRO definia uma alíquota única para todas as culturas agrícolas em todo o território nacional. A partir de 2017, houve uma mudança na tarifação. As alíquotas passaram a se diferenciar por culturas, mas ainda assim, permaneceram as mesmas para todo o território nacional.

Por exemplo, alíquota de 4,5% é definida para todos os produtores de soja do país, ou seja, independente do risco associado a cada propriedade, a alíquota é a mesma.

Equivocadamente, o PROAGRO considera o produtor situado no norte do Paraná com o mesmo risco do produtor localizado em Lucas do Rio Verde, no Mato Grosso.

Quando os riscos não são adequadamente precificados surgem problemas de desequilíbrio atuarial da carteira e seleção adversa, ou seja, quando há anti-seletividade por problemas de assimetria de informação. Na prática, quando a taxa de prêmio é fixada conforme o risco médio em uma carteira composta por produtores de alto e baixo risco, aqueles com risco maior terão uma probabilidade maior de demandar pelo seguro do que os produtores de baixo risco.

Assim, apenas aqueles de alto risco terão incentivos em demandar pelo seguro. Como consequência a concentração da carteira com indivíduos de alto risco aumenta consideravelmente a chance de tornar a carteira deficitária. Na safra seguinte, o mercado tende a elevar a taxa a fim de compensar as perdas. O problema torna-se ainda pior, e assim, sucessivamente.

Percebe-se, dessa forma, que definir uma taxa única para todo o território nacional baseado em uma média global, para cada cultura agrícola, apenas fortalece o problema da seleção adversa, com severas implicações para o equilíbrio da carteira e para a continuidade do seguro agrícola. O histórico do PROAGRO revela tais dificuldades.

O procedimento correto nessas circunstâncias é a segmentação das taxas por unidade segurada (município ou propriedade rural) e tipo de cultura, levando-se em conta suas peculiaridades edafoclimáticas, genética e de manejo. Porém, dadas as limitações de dados estatísticos que reflitam a estrutura de risco em nível de propriedade recorre-se ao uso de dados agregados em nível municipal como uma proxy dos dados individualizados.

Para fins práticos, calculou-se as taxas por municípios do Paraná, para a cultura da soja. Antes, porém, é importante destacar um ponto. As perdas financeiras no PROAGRO podem ocorrer de duas formas distintas: i) perdas de receita pela má especificação da taxa, ou

seja, não se distingue entre municípios de alto, médio e baixo risco, associando a todos uma única taxa. Dessa forma, perde-se receita, pois a taxa atual de 4,5% é menor do que aquela que deveria ser cobrada; e, ii) perdas de receita devido a migração de produtores ao seguro privado, pois não estão dispostos a pagar pelo PROAGRO e enxergam no seguro privado uma solução mais barata, devido a subvenção ao prêmio.

Nesse contexto, a estimativa da perda de receita do PROAGRO foi da ordem de R\$ 88,3 milhões. Ao passo que expectativa de perda de receita em função da migração para o seguro privado é da ordem de R\$ 53,8 milhões. Se considerarmos as duas perdas, atinge-se um valor máximo (teto) de perda da ordem de R\$ 142,1 milhões. Esse valor representa a perda de receita apenas para a soja em um único estado. Cabe ressaltar que o PROAGRO atua com muitas outras culturas em praticamente todo o território nacional, portanto a perda de receita é muito maior do que o valor supramencionado.

As seções seguintes mostram todos os produtos separados por capítulos. Nesse material encontra-se a metodologia e também os resultados na forma tabular, gráficos e mapas, bem como o material complementar, tais como, códigos computacionais e descrição do passo a passo de todos os cálculos.

2. RECOMENDAÇÃO DE DATA DE SEMEADURA BASEADA EM INDICADOR TEMPORAL DE PRODUTIVIDADE E CENÁRIOS DE MUDANÇAS CLIMÁTICAS

2.1 INTRODUÇÃO

De acordo com os Termos de Referência do Projeto intitulado “Investimento Público e Adaptação à Mudança do Clima na América Latina (IPACC II)”, PN: 2014.9049.9-002: “No marco de implementação do projeto IPACC II, as contrapartes principais do Brasil, o Ministério de Planejamento, Desenvolvimento e Gestão (MP), o Ministério da Fazenda (MF)

e o Ministério do Meio Ambiente (MMA), identificaram o setor agrícola como sendo um dos setores que aloca altos níveis de investimento público e privado, e ao mesmo tempo sendo um dos setores de maior vulnerabilidade aos efeitos da mudança do clima. Tais condições motivaram a execução de uma iniciativa piloto, conjuntamente com o Ministério de Agricultura, Pecuária e Abastecimento (MAPA), com a finalidade de inserir a gestão do risco climático nos instrumentos de análise de investimentos no setor e, desta forma, complementar o instrumento atualmente utilizado Zoneamento Agrícola de Risco Climático (ZARC) buscando aprimorá-lo a partir da criação de recomendações de novas metodologias científicas, como suporte à política de gestão agrícola”.

O objetivo geral do projeto é desenvolver um conjunto de recomendações visando reduzir as perdas agrícolas associadas aos efeitos da mudança do clima e que contribuam para uma alocação eficiente dos investimentos públicos e privados na agricultura. Nesta seção, trata-se do objetivo específico referente ao Produto 2 do projeto, que aborda a “Recomendação de data de semeadura baseada em indicador temporal de produtividade e cenários de mudanças climáticas”.

2.2 CENÁRIOS CLIMÁTICOS E DE CONCENTRAÇÃO DE CARBONO

A abordagem utilizada para seleção dos Modelos de Circulação Global (GCMs) a serem considerados neste projeto baseou-se no protocolo proposto pelo Projeto AgMIP que, por sua vez, vale-se do conceito de *Representative Climate Futures* (RCF) proposto por Whetton et al, 2012)¹. A justificativa para o uso da abordagem RCFs está em se reduzir o tempo e a capacidade computacional necessária para as simulações e assim produção de uma grande quantidade de dados similares entre si. Além disso, dado que as projeções climáticas de longo prazo carregam incertezas e dado que a finalidade de estudos de vulnerabilidade climática como no presente projeto é a geração de cenários agrícolas futuros, a abordagem RCF é suficientemente robusta para permitir a avaliação de cenários futuros para a cultura da soja para o Estado do Paraná, assim como largamente utilizado na literatura científica² e proposto do termo de referência do presente projeto. Com base nesta abordagem, foram gerados um cenário otimista (com o GCM de menor variação em relação a linha de base),

¹ Whetton, P., Hennessy, K., Clarke, J., McInnes, K., Kent. D. 2012. Use of Representative Climate Futures in impact and adaptation assessment. *Climatic Change*. 115(3–4), 433–442.

² Inúmeros estudos científicos publicados em periódicos internacionais de alto impacto valem-se da mesma abordagem usada neste projeto, citando-se como exemplos os seguintes:

- Marin, F.R., J.W. Jones, A. Singels, F. Royce, E.D. Assad, G.O. Pellegrino, and F. Justino. 2013. Climate change impacts on sugarcane attainable yield in Southern Brazil. *Climatic Change* 117: 227–239.
- Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., ... Winter, J. M. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170, 166–182. <https://doi.org/10.1016/j.agrformet.2012.09.011>
- Singels, A., Jones, M., Marin, F., Ruane, A., & Thorburn, P. (2014). Predicting Climate Change Impacts on Sugarcane Production at Sites in Australia, Brazil and South Africa Using the Canegro Model. *Sugar Tech*, 16(4). <https://doi.org/10.1007/s12355-013-0274-1>
- Wilby, R.L., S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. Intergovernmental Panel on Climate Change (IPCC) Supporting Material, available from the Data Distribution Centre of the IPCC Task Group on Data and Scenario Support for Impacts and Climate Analysis (TG CIA), http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf.

intermediários e pessimistas (com base no GCM de maior variação em relação a linha de base) de produtividade futura da cultura da soja.

A variável de verificação utilizada como referência para a geração dos RCFs foi a temperatura do ar, por se tratar de variável climática que expressa a quantidade de energia no ambiente para a realização dos processos atmosféricos e que está fortemente correlacionada com a chuva e com a própria irradiância solar global. Vale lembrar que a temperatura, radiação solar e chuva são as variáveis de entrada para os modelos agrícolas utilizados neste projeto (DSSAT e APSIM, descritos mais a frente neste documento). Assim, com base na abordagem RCF e tendo a temperatura do ar como variável-critério, a partir de uma análise das base de dados do Paraná foram selecionados três GCMs, representado, como já mencionado, o modelo com a maior variação de temperatura em relação a linha de base, o modelo com a menor variação e, por fim, um modelo utilizando uma variação intermediária entre dois anteriores, dentre os 20 disponíveis AGMIP. Além desses três GCMs, incluiu-se um quarto modelo (HadGem2), por solicitação da equipe gestora do por ocasião da reunião de março de 2018. Assim, trabalhou-se com quatro bases de dados climáticos originados dos seguintes GCM's:

1. GCM de variação mínima em relação a LB: MIROC5;
2. GCM de variação média em relação a LB: BNU-ESM;
3. GCM de variação máxima em relação a LB: CanESM2.
4. GCM adicional: HadGem2

Para cada um dos 4 GCMs selecionados, considerou-se 2 níveis de concentração de CO₂ atmosférico, de modo a representar as forçantes radioativas descritas nos *Representative Concentration Pathways* (RCPs) 4.5 e 8.5. Assim, para efeito de simulação, foram assumidos

os níveis de concentrações de CO₂ de 526 ppm (RCP 4.5) e 628 ppm (RCP 8.5), constituindo-se dessa forma, para cada local selecionado, 9 bases climáticas (1 atual e 8 futuros), como mostrado na Tabela 1. A Tabela 2 apresenta a variações médias de temperatura e chuva associadas a cada um dos GCM's selecionados.

Tabela 1. Relação de GCM's e RCP's utilizados nas simulações com os modelos agrícolas.

Base climática	Clima Atual	Cenário Futuro/GCM	RCP
1	Linha de base	--	--
2	--	MIROC5	4.5
3	--	MIROC5	8.5
4	--	BNU-ESM	4.5
5	--	BNU-ESM	8.5
6	--	CanESM2	4.5
7	--	CanESM2	8.5
8	--	HadGem2	4.5
9	--	HadGem2	8.5

Tabela 2. Variação em relação a linha de base de temperatura do ar e chuva total para o Estado do Paraná associada a cada CGM selecionados a partir da metodologia proposta por Hudson & Ruane (2013)³.

³ Hudson, N; Ruane, A. 2013. Guide for Running AgMIP Climate Scenario Generation Tools with R in Windows. Disponível em <http://www.agmip.org/wp-content/uploads/2013/10/Guide-for-Running-AgMIP-Climate-Scenario-Generation-with-R-v2.3.pdf>

Base climática	Variação de temp. (°C)	Variação de chuva (%)
2	0.9752	8.653%
3	1.6816	3.780%
4	1.7951	-5.737%
5	2.2237	-5.482%
6	2.8691	0.136%
7	3.9698	4.974%
8	2.3108	18.774%
9	3.1621	6.821%

2.3 RESOLUÇÃO ESPACIAL

Referente a resolução espacial dos mapas finais, considerando-se que a base de dados climáticos obtidos para a simulação da linha de base (densidade de 142 pontos de dados distribuídos no Estado do Paraná), isso permitiu gerar mapas com resolução de até 1,6 km de resolução⁴, portanto com maior nível de informação espacial do que o proposto inicialmente no termo de referência, que era de 5 km. Os métodos de downscaling podem ser de natureza dinâmica (e.g. valendo-se de modelo climáticos regionais) ou estatísticos (i.e. utilizando estatísticas climáticas históricas e informações sobre mudanças climáticas projetadas) ou ainda uma combinação de ambos. Assim, em acordo com os protocolos da principal referência científica mundial sobre estudos no campo da modelagem agrícola aplicada à mudança do clima - o programa AgMIP (www.agmip.org) e conforme descrito em Hudson & Ruane (2013)⁵ e Ruane et al. (2013)⁶, neste projeto utilizou-se da abordagem estatística para a

⁴ Como disponível nas ferramentas de cartografia do software ArcGIS e detalhado em <https://www.esri.com/arcgis-blog/products/product/mapping/choosing-an-appropriate-cell-size-when-interpolating-raster-data/>

⁵ Hudson, N; Ruane, A. 2013. Guide for Running AgMIP Climate Scenario Generation Tools with R in Windows. Disponível em <http://www.agmip.org/wp-content/uploads/2013/10/Guide-for-Running-AgMIP-Climate-Scenario-Generation-with-R-v2.3.pdf>

geração de cenários climáticos futuros que, por sua vez, baseiam-se na premissa de que o clima regional é condicionado por dois fatores principais: as condições climáticas de larga escala e as feições fisiografias (topografia, uso do solo, continentalidade/oceanidade), de modo que a informação climática é determinada, a princípio, por um modelo estatístico que relaciona as variáveis climáticas de larga escala (no caso deste projeto, os dados oriundos dos GCMs com as variáveis locais (base de dados que compõem a linha de base de cada uma das 142 estações meteorológicas utilizadas). Assim, os dados dos GCMs alimentam um modelo estatístico que estima o clima regional, considerando a estrutura temporal da série da linha de base. O ponto forte desta abordagem é permitir a obtenção de dados para localidades específicas (pontos ou bacias hidrográficas de pequena escala), como é o caso da abordagem utilizada neste projeto. Ademais, este método já leva em conta a correção de possíveis erros sistemáticos dos modelos, fazendo a geração dos cenários futuros a partir da perturbação estatística originada a partir dos GCM's sobre a série climática da linha de base, e não usando os dados futuros gerados pelos modelos propriamente ditos, o que permite fazer a comparação entre cenários agrícolas futuros e atuais com a devida coerência técnica.

2.4 SIMULAÇÕES COM OS MODELOS AGRÍCOLAS

Após a constituição da base climática, foi feita organização da base de solos do Paraná em três tipos de solos conforme classe textural (arenoso, médio, argiloso) também foram estabelecidos foram três tipos de materiais genéticos (ciclo precoce, médio e tardio) representando os diferentes tipos de comprimentos de ciclo da soja no Estado do Paraná. Os modelos parametrizados foram *DSSAT/CropGro-Soybean* e o modelo *APSIM-Soybean* e todas as simulações foram feitas com ambos os modelos em paralelo. Um terceiro cenário agrícola futuro foi gerado a partir da média dos dois modelos (DSSAT e APSIM). Por fim,

⁶ Ruane, A.C., J.M.Winter, S.P.McDermid, N.I.Hudson. 2013. AgMIP Climate Data and Scenarios for Integrated Assessment. Chapter 3. Handbook of Climate Change and Agroecosystems.

para cada cenário climático, cada tipo de solo e cada genótipo foram simuladas 16 datas de semeadura, em passos decendiais (ou seja, a cada 10 dias), entre agosto e dezembro de cada ano da base de dados (30 anos, entre 1980 e 2010).

2.5 RESULTADOS

Tais simulações geraram um banco de dados que foi analisado para a geração das zonas homogêneas de risco conforme descrito na próxima seção deste documento. Além disso, as simulações foram organizadas e os dados apresentados da forma de mapas de produtividade, risco climático e zonas homogêneas e disponibilizadas no repositório compartilhado com o comitê gestor do projeto.

Como já mencionado, os mapas foram gerados para cada data de semeadura, tipo de solo e material genético utilizando os dois modelos baseados em processos (DSSAT-Cropgro e APSIM-Soybean), além da média entre esses dois modelos, utilizando as bases climáticas oriundas da linha de base e dos oito GCM/RCP's selecionados. Entende-se, com base no que foi proposto no termo de referência do projeto e na descrição metodológica e produtos gerados, que este produto está finalizado. Um exemplo de mapa disponível no repositório do Sistema Tempocampo é apresentado nas Figuras 1 e 2 descrevem a variação espacial da produtividade simulada com a média dos modelos agrícolas utilizados na simulação, para duas datas de semeadura (21/10 e 21/12), solos de textura média e cultivares de ciclo longo (tardio).

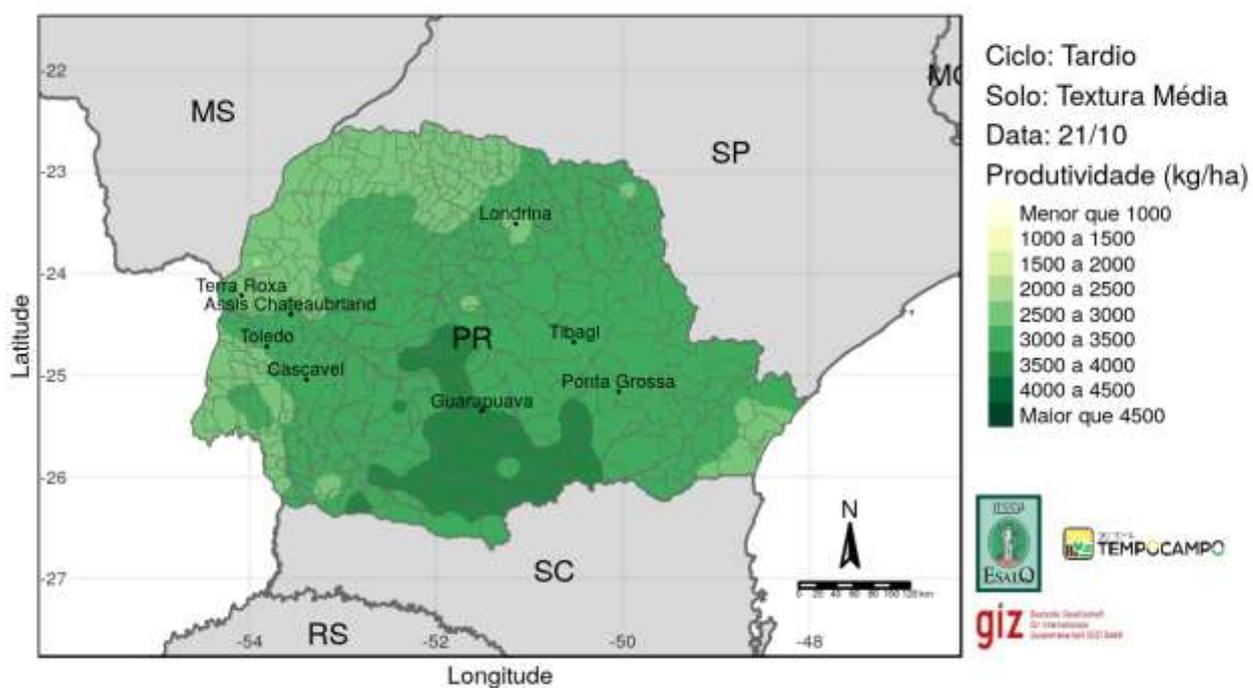


Figura 1. Mapa da produtividade média simulada com os modelos Apsim-Soybean e DSSAT/Cropgro semeada em 21/10 em solos de textura média e cultivares de ciclo longo.

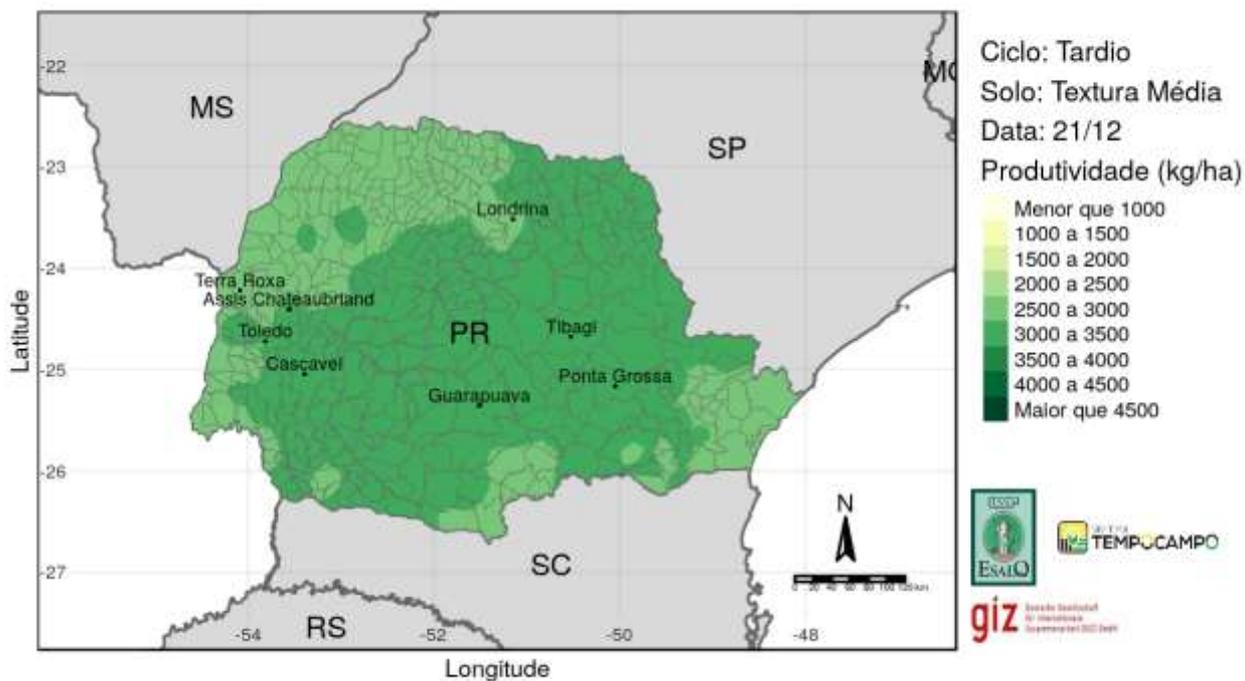


Figura 2. Mapa da produtividade média simulada com os modelos Apsim-Soybean e DSSAT/Cropgro semeada em 21/12 em solos de textura média e cultivares de ciclo longo.

2.6 NSIDERAÇÕES FINAIS

Os requisitos expressos no termo de referência relativo a este produto foram cumpridos integralmente e a qualidade das simulações é avaliada de forma positiva, ainda que esta análise seja ainda preliminar. Os resultados estão todos disponíveis no repositório do projeto compartilhado com o comitê gestor do projeto, onde é possível fazer o download integral das tabelas de dados e mapas referente aos produtos 2, 3, 4 e 5.

3. SEÇÃO 2 - DETERMINAÇÃO DAS ZONAS HOMOGÊNEAS DE RISCO - ESTRATIFICAÇÃO MUNICIPAL DO RISCO BASEADO EM MULTIMODELOS AGROMETEOROLÓGICOS E CENÁRIOS DE MUDANÇAS CLIMÁTICAS

3.1 INTRODUÇÃO

De acordo com os Termos de Referência do Projeto intitulado “Investimento Público e Adaptação à Mudança do Clima na América Latina (IPACC II)”, PN: 2014.9049.9-002: *“No marco de implementação do projeto IPACC II, as contrapartes principais do Brasil, o Ministério de Planejamento, Desenvolvimento e Gestão (MP), o Ministério da Fazenda (MF) e o Ministério do Meio Ambiente (MMA), identificaram o setor agrícola como sendo um dos setores que aloca altos níveis de investimento público e privado, e ao mesmo tempo sendo um dos setores de maior vulnerabilidade aos efeitos da mudança do clima. Tais condições motivaram a execução de uma iniciativa piloto, conjuntamente com o Ministério de Agricultura, Pecuária e Abastecimento (MAPA), com a finalidade de inserir a gestão do risco climático nos instrumentos de análise de investimentos no setor e, desta forma, complementar o instrumento atualmente utilizado Zoneamento Agrícola de Risco Climático (ZARC) buscando aprimorá-lo a partir da criação de recomendações de novas metodologias científicas, como suporte à política de gestão agrícola”.*

O objetivo geral do projeto é desenvolver um conjunto de recomendações visando reduzir as perdas agrícolas associadas aos efeitos da mudança do clima e que contribuam para uma alocação eficiente dos investimentos públicos e privados na agricultura. Como objetivo específico encontra-se o produto 3: “Determinação das zonas homogêneas de risco -

estratificação municipal do risco baseado em multimodelos agrometeorológicos e cenários de mudanças climáticas”.

3.2 ANÁLISE DE RISCO CLIMÁTICO

A estratificação municipal do risco foi feita considerando-se cinco níveis de risco (NR, 60%, 65%, 70%, 75% e 80%) aplicado sobre a produtividade média de cada município e, a partir dela fazendo uma análise frequencial, como mostrado na equação 1. Com ela, pode-se estabelecer qual o nível de risco para cada combinação (clima, solo, genética) e quantificar o nível de risco para cada data de semeadura. No presente caso, optou-se por não delimitar uma janela de semeadura fixa por entendermos que isso implicaria em perda de informações e empobrecimento dos resultados- na forma atual, basta ao tomador de decisão definir qual o nível de risco que se pretende admitir e então define-se a janela de semeadura para este nível de risco.

$$F_{i,j,k,l,r,q} = F(Y_{i,j,k,l,r} \geq Y_{c,j,k,l,r} * NR_q) = \frac{(1 - m_{i,j,k,l,r,q})}{n} * 100$$

em que F é a frequência (em porcentagem) da produtividade para um determinada data de semeadura, solo, clima e genética ser maior do que o produto da produtividade média (30 anos e todas as datas de semeadura) (Y_c) para um determinado solo, clima e genética aquela data, pelo nível de risco selecionado (NR), sendo dado em cinco (q) níveis (60%, 65%, 70%, 75% e 80%). i é data da semeadura, j é o município, k é o clima (linha de base ou climas futuros), l é o tipo de solo, r é o material genético.

Este produto foi concluído e os resultados estão disponíveis no mesmo repositório do portal do Sistema Tempocampo, sendo lá apresentados como Tabelas de Risco. Tais tabelas estão organizadas para cada combinação de solo, grupo genético e clima, apresentando em cada uma delas o nível de risco referente a cada data de semeadura, permitindo assim recomendar as datas mais adequadas. Aqui, diferentemente da abordagem tradicional do ZARC de indicar ou não cada uma das datas, apresentadas o nível de risco na forma de classes, permitindo assim ao usuário selecionar quais níveis de risco pretende considerar na sua análise. Caso deseje, há também tabelas (no mesmo formato) apresentando o valor absoluto do risco climático (como opção no seletor *critério* do website). Assim, considerando-se os mapas de produtividade gerados no produto 2 e somando-se aqui os mapas gerados para cada nível de risco (e considerando-se a ramificação em termos de modelos, GCM's, RCP's, solos, genéticas, e datas de semeadura) tem-se no repositório do Sistema Tempocampo aproximadamente 23.000 mapas referentes a todas as combinações simuladas. Entende-se, com base no que foi proposto no termo de referência do projeto e na descrição metodológica e produtos gerados, que este produto está finalizado.

3.3 RESULTADOS

Tais simulações geraram um banco de dados que foi analisado para a geração das zonas homogêneas de risco conforme descrito na próxima seção deste documento. Considerando-se os mapas de produtividade gerados no Produto 2 e somando-se aqui os mapas gerados para cada nível de risco, tem-se no repositório compartilhado com o comitê gestor do projeto um conjunto de 23.000 mapas, além de um conjunto de tabelas correspondentes a cada um dos mapas.

Como já mencionado, os mapas foram gerados para cada data de semeadura, tipo de solo e material genético utilizando os dois modelos baseados em processos (DSSAT-Cropgro e APSIM-Soybean), além da média entre esses dois modelos, utilizando as bases climáticas oriundas da linha de base e dos oito GCM/RCP's selecionados. Entende-se, com base no que foi proposto no termo de referência do projeto e na descrição metodológica e produtos gerados, que este produto está finalizado. Um exemplo de mapa de risco disponível no repositório do Sistema Tempocampo é apresentado nas Figuras 1 e 2, que mostram a variação espacial do risco climático para duas datas de semeadura (21/9 e 21/10) dadas as mesmas condições de contorno para as simulações (solo, genética, clima e níveis de risco). Vale destacara que a metodologia de mapeamento atende integralmente aos requisitos para a recomendação das datas de semeadura (assumindo-se ZARC como referência técnica para esse procedimento de recomendação de datas), uma vez que cada município é classificado integralmente em uma única classe de risco levando-se em conta a área de cada faixa de risco dentro do município. Esta técnica de classificação é importante para os casos em que um município é dividido em duas classes de risco diferentes. Nestes casos, tomou-se como classe aquela de menor risco e que representasse pelo menos 20% da área do município. As Tabelas 1 e 2 mostram, por sua vez, mostram um exemplo de variação dos valores absolutos de risco e as classes de risco, respectivamente, para cada data de semeadura e para todos os municípios do Estado do Paraná.

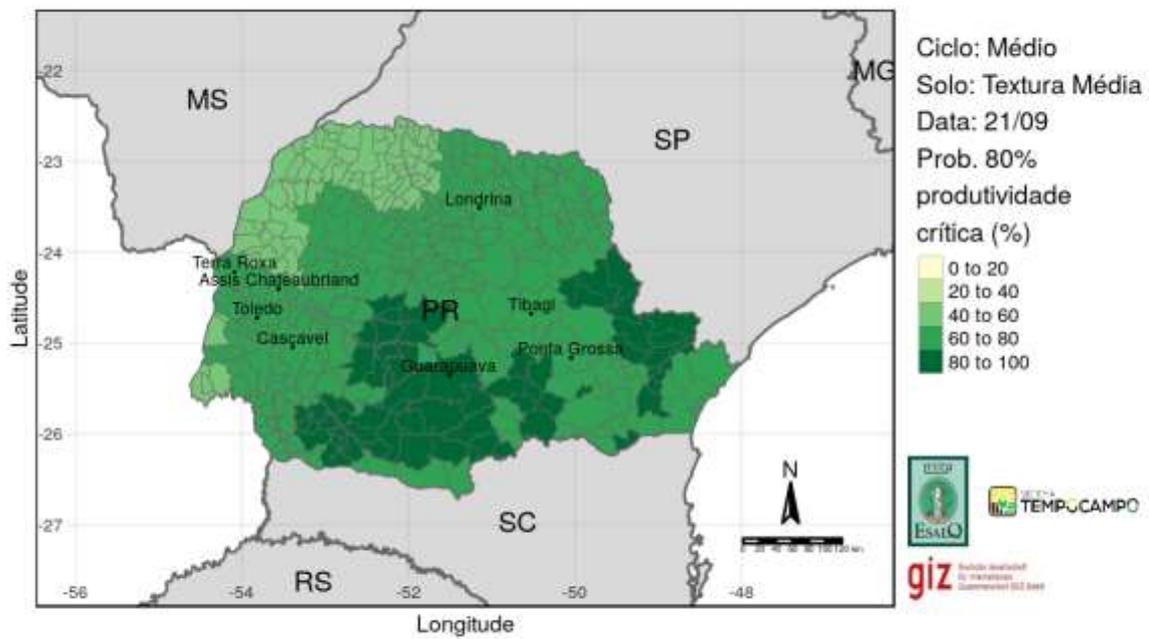


Figura 1. Mapa de risco climático calculado a partir da produtividade média dos dois modelos, com data de semeadura em 21/9 em solos de textura média e cultivares de ciclo médio.

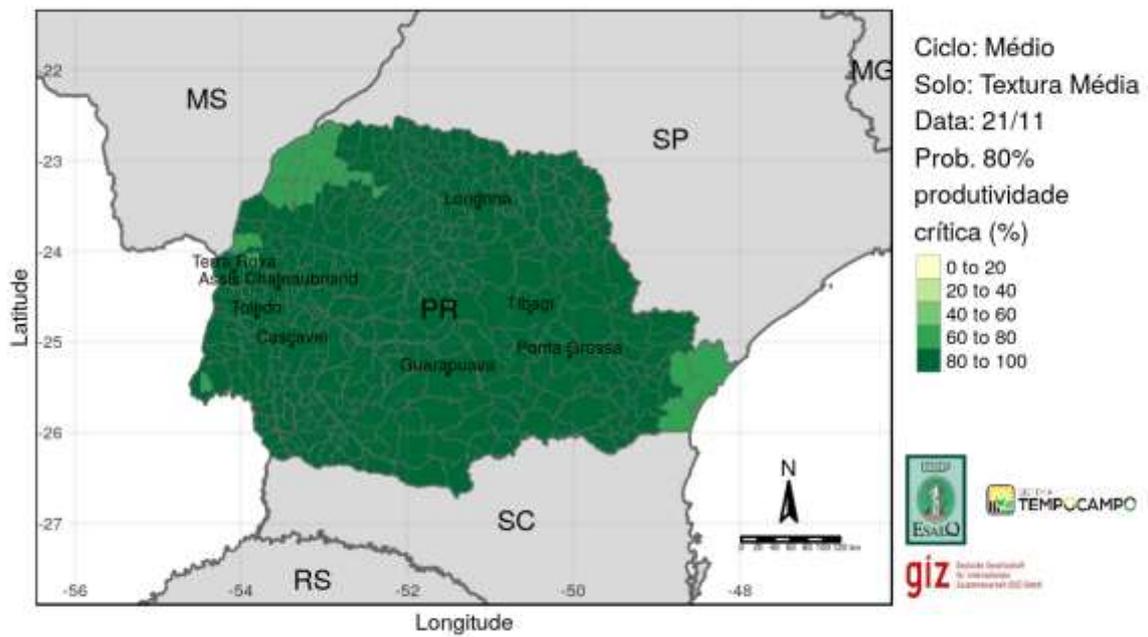


Figura 2. Mapa de risco climático calculado a partir da produtividade média dos dois modelos, com data de semeadura em 21/11 em solos de textura média e cultivares de ciclo médio.

Tabela 1. Variação da frequência de produtividades superiores ao nível de risco assumido para cada município do Estado do Paraná e para as 16 datas de semeadura.



Modelo
 Média(DSSAT e APSIM)

Genótipo
 Tardio

Solo
 Textura Argilosa

Critério
 Probabilidade

Cenário
 Linha de Base

Produtividade Crítica
 60%

Mostrar
 Voltar Mapas

Município	GEOCODIGO	01/08	11/08	21/08	01/09	11/09	21/09	01/10	11/10	21/10	01/11	11/11	21/11	01/12	11/12	21/12	30/12
ABATIÁ-PR	4100103	55.02	60.81	68.90	79.26	85.34	89.15	92.37	97.78	98.12	98.33	98.33	98.02	98.73	96.66	97.05	94.67
ADRIANÓPOLIS-PR	4100202	74.44	75.63	75.85	83.53	88.00	95.37	94.64	97.98	97.89	98.31	98.18	98.67	98.53	98.47	94.78	94.51
AGUDOS DO SUL-PR	4100301	74.64	78.04	79.72	86.23	91.16	94.50	94.76	98.18	98.30	98.32	98.33	98.63	98.65	98.25	98.58	98.18
AL MIRANTE TAMANDARÉ-PR	4100400	68.35	72.52	73.19	74.41	77.80	94.55	92.43	94.14	96.21	96.81	98.01	98.34	98.43	93.75	93.19	89.84
ALTAMIRA DO PARANÁ-PR	4100459	72.46	72.68	75.98	74.64	81.16	87.87	93.01	96.63	97.50	97.20	97.60	98.26	98.26	98.22	94.35	92.89
ALTO PARAÍSO-PR	4128625	52.10	67.32	58.17	61.89	67.00	74.18	83.95	92.47	96.87	97.91	98.33	98.00	97.27	96.67	94.41	91.73

Tabela 2. Classes de frequência de produtividades superiores ao nível de risco assumido para cada município do Estado do Paraná e para as 16 datas de semeadura.



Modelo

Média(DSSAT e APSIM)

Genótipo

Tardio

Solo

Textura Argilosa

Critério

Classes

Cenário

Linha de Base

Produtividade Crítica

60%

Mostrar

Voltar Mapas

Município	GEODIDIGO	01/08	11/08	21/08	01/09	11/09	21/09	01/10	11/10	21/10	01/11	11/11	21/11	01/12	11/12	21/12	30/12
ABATIÁ-PR	4100103	3	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1
ADRIANÓPOLIS-PR	4100202	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
AGUDOS DO SUL-PR	4100301	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1
ALMIRANTE TAMANDARÉ-PR	4100400	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1
ALTAMIRA DO PARANÁ-PR	4100459	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1
ALTO PARAISSÓ	4100535	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1

Classe de Risco

1 = Muito Baixo 2 = Baixo 3 = Médio 4 = Alto 5 = Muito Alto

3.4 CONSIDERAÇÕES FINAIS

Os requisitos expressos no termo de referência relativo a este produto foram cumpridos integralmente e a qualidade das simulações é avaliada de forma positiva, ainda que esta análise seja ainda preliminar. Os resultados estão todos disponíveis no repositório compartilhado com o comitê gestor do projeto, onde se pode fazer o download integral das tabelas de dados e mapas referente aos produtos 2, 3, 4 e 5.

4. SEÇÃO 3- PROBABILIDADE DE PERDA – ESTRATIFICAÇÃO MUNICIPAL DO RISCO BASEADO EM MODELOS ATUARIAIS

4.1 INTRODUÇÃO

De acordo com os Termos de Referência do Projeto intitulado “Investimento Público e Adaptação à Mudança do Clima na América Latina (IPACC II)”, PN: 2014.9049.9-002: “No marco de implementação do projeto IPACC II, as contrapartes principais do Brasil, o Ministério de Planejamento, Desenvolvimento e Gestão (MP), o Ministério da Fazenda (MF) e o Ministério do Meio Ambiente (MMA), identificaram o setor agrícola como sendo um dos setores que aloca altos níveis de investimento público e privado, e ao mesmo tempo sendo um dos setores de maior vulnerabilidade aos efeitos da mudança do clima. Tais condições motivaram a execução de uma iniciativa piloto, conjuntamente com o Ministério de Agricultura, Pecuária e Abastecimento (MAPA), com a finalidade de inserir a gestão do risco climático nos instrumentos de análise de investimentos no setor e, desta forma, complementar o instrumento atualmente utilizado Zoneamento Agrícola de Risco Climático (ZARC) buscando aprimorá-lo a partir da criação de recomendações de novas metodologias científicas, como suporte à política de gestão agrícola.”

O objetivo geral deste projeto é desenvolver um conjunto de recomendações visando reduzir as perdas agrícolas associadas aos efeitos da mudança do clima e que contribuam para uma alocação eficiente dos investimentos públicos e privados na agricultura. Como objetivo específico encontra-se o produto 6:

“Indicar, em mapas e tabelas, a probabilidade de perda: avaliação da probabilidade de perda por meio de resultados gerados pelos multimodelos de produtividade potencial”.

4.2 A PROBABILIDADE DE PERDA

A probabilidade de perda foi gerada tendo-se como base as informações de produtividade agrícola dos produtos 2 e 3 supracitados. A probabilidade de perda foi calculada a partir da seguinte expressão:

$$\text{Probabilidade de Perda (PP)} = \text{Prob}(Y \leq y_g)$$

em que Y é a produtividade agrícola e y_g é a produtividade garantida, ou seja, o nível de cobertura multiplicado pela produtividade estimada. O nível de cobertura variou de 60 a 80%, em múltiplos de 5%. O ajuste de distribuições é fundamental para mensurar com o maior nível de precisão a probabilidade de perda. Nesse sentido, utilizou-se as seguintes distribuições paramétricas: Normal Assimétrica, Skew-T e a distribuição Odd Log-logística, por apresentarem robustez e flexibilidade na detecção de assimetrias e bimodalidades. As seções seguintes mostram as características das distribuições supracitadas.

4.3 A DISTRIBUICAO NORMAL ASSIMÉTRICA

Considerou-se a distribuição Normal Assimétrica (DNA) pela sua flexibilidade em acomodar assimetrias presentes nas distribuições dos dados. A DNA foi proposta por Azzalini (1985) e é uma generalização da distribuição Normal. A vantagem está no fato de se captar assimetrias tanto negativas quanto positivas. Tais assimetrias tem impacto direto no calculo da probabilidade de perda e, conseqüentemente, no cálculo da taxa de prêmio.

Nesse contexto, considere f a função densidade de probabilidade (f.d.p.) simétrica ao redor de zero e G função distribuição cumulativa (f.d.c.) contínua e absoluta, tal que $g = G'$ é simétrica ao redor de zero (Ozaki e Silva, 2009). Então,

$$2 f(y) G(\gamma y), y \in R \quad (1)$$

É uma f.d.p. para qualquer $\gamma \times \in R$. Da equação (1), a DNA, com parâmetro de locação μ , escala σ e forma γ , é definida como:

$$\phi(y|\mu, \sigma, \gamma) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\frac{\gamma(y-\mu)}{\sigma}\right) \quad (2)$$

em que ϕ e Φ são respectivamente a f.d.p. e a f.d.c. de uma variável aleatória Normal padrão. O grau de assimetria da DNA é dada pela equação

$$\varepsilon = \delta^3 \left[\frac{4}{\pi} - 1 \right] \left[1 - \frac{2\delta^2}{\pi} \right]^{-3/2} \sqrt{2/\pi} \quad (3)$$

onde $\delta = \gamma(1 + \gamma^2)^{-1/2}$ e $-0.99527 < \varepsilon < 0.99527$ com casos limites ($\gamma \rightarrow \pm\infty$) sendo uma distribuição Meio Normal. De fato, utilizou-se a abordagem de Henze (1986), representando a DNA como uma mistura de média variância de uma Normal e uma Meia Normal.

4.4 A DISTRIBUICAO SKEW-T

Quando se tem dados com um comportamento assimétrico e presença de valores extremos ou discrepantes, a distribuição Skew-Normal pode não se ajustar bem. Em tais situações, uma alternativa mais robusta ao modelo Skew-Normal é proposta por Azzalini e Capitanio (2003), a versão assimétrica da distribuição t-Student (ST), denominada t-Student Assimétrica.

Uma variável X é chamada “Skew-t Student”, com parâmetros de locação μ , $\mu \in \mathbb{R}$, escala σ , $\sigma \in \mathbb{R}_+$, assimetria v , $v \in \mathbb{R}$, e curtose τ , $\tau \in \mathbb{R}_+$, se sua f.d.p é dada por:

$$f(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} t\left(\frac{x-\mu}{\sigma}\right) T\left[v\sqrt{\lambda}\left(\frac{x-\mu}{\sigma}\right)\right], \quad (4)$$

em que $z = \left(\frac{x-\mu}{\sigma}\right)$, $w = v\lambda^{1/2}z$, $\lambda = \frac{\tau+1}{\tau+z^2}$, $tZ1$ e $TZ1$ são respectivamente, a f.d.p da t-Student padrão com τ graus de liberdade e a fda da t-Student padrão com $\tau + 1$ graus de liberdade. Como notação, adota-se $Z \sim ST(\mu, \sigma, v, \tau)$. Note que, quando $v = 0$, a distribuição é equivalente a t-Student simétrica.

4.5 A DISTRIBUICAO ODD LOG-LOGÍSTICA

Uma nova classe de família de distribuições contínuas recentemente introduzidas por Cruz (2016) é chamada odd log-logistic-F (OLLF). A inferência e estimação pelo método da máxima verossimilhança para grandes amostras são também abordados para dados de sobrevivência censurados. Esta nova classe de distribuições é interessante, pois através de suas simulações pode-se capturar a bimodalidade dos dados.

Dado uma função base de probabilidade acumulada qualquer com $F(x; \xi)$ um vetor de parâmetros ξ , a função distribuição cumulativa (f.d.c.) da distribuição odd log-logistic-F (OLL-F) com um parâmetro de forma $\alpha > 0$ é definido por:

$$G(x; \xi) = \int_0^{\frac{F(x; \xi)}{\bar{F}(x; \xi)}} \frac{\alpha x^{\alpha-1}}{(1+x^\alpha)^2} dx = \frac{F(x; \xi)^\alpha}{F(x; \xi)^\alpha + \bar{F}(x; \xi)^\alpha} \quad (5)$$

em que $\bar{F}(x, \xi) = 1 - F(x, \xi)$ e $\alpha = \frac{\log \frac{G(x, \xi)}{G(x, \xi)}}{\log \frac{F(x, \xi)}{F(x, \xi)}}$.

O parâmetro α representa o quociente do log da razão de chances para a distribuição base. Se $F(x, \xi) = \frac{x}{1+x}$, a equação (5) então se reduz a função de probabilidade da distribuição log-logística. Várias distribuições podem ser geradas da equação acima, como as apresentados nos trabalhos de Alizadeh et al. (2015), Cruz et al. (2015) e Braga et al. (2016).

A nova família de distribuição OLLF permite uma maior flexibilidade nas caudas da distribuição. Se considerarmos a função base $G(x; \xi)$ como a distribuição acumulada da normal $N(\mu, \sigma)$ Assim, para a nova classe de família odd log logistic normal (OLL-N) a equação (5) se reduz a:

$$G(x; \mu, \sigma, \alpha) = \int_0^{\frac{\Phi(x; \mu, \sigma, \alpha)}{\bar{\Phi}(x; \mu, \sigma, \alpha)}} \frac{\alpha x^{\alpha-1}}{(1+x^\alpha)^2} dx = \frac{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right)}{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha} \quad (6)$$

em que $\bar{\Phi}(x; \mu, \sigma) = 1 - \Phi(x; \mu, \sigma)$.

A função densidade de probabilidade (f.d.p) é dada por:

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right) \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}}{\sigma \left\{ \Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha \right\}^2} \quad (7)$$

Note que $\alpha > 0$ é um parâmetro de forma. A partir de agora, assumimos que uma variável aleatória com função de densidade como da equação (7) é denotado por $X \sim \text{OLLN}(\alpha, \mu, \sigma)$.

4.6 ESCOLHA DE MODELOS

A escolha da melhor distribuição ajustada aos dados foi dada pelo critério de informação de Akaike (AIC) e pelo critério de informação Bayesiano (BIC). O AIC é um estimador da qualidade relativa de modelos estatísticos para um dado conjunto de dados (Akaike, 1974).

Dada uma coleção de modelos para os dados, o AIC estima a qualidade de cada modelo, em relação a cada um dos outros modelos. Assim, o AIC fornece um meio para a seleção de modelos. Suponha que tenhamos um modelo estatístico de alguns dados. Seja k o número de parâmetros estimados no modelo. Seja \hat{L} o valor máximo da função de verossimilhança para o modelo. Então o valor AIC do modelo é dado por

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

Dado um conjunto de modelos candidatos para os dados, o modelo preferido é aquele com o valor mínimo de AIC. Assim, o AIC recompensa a qualidade do ajuste (avaliado pela função de verossimilhança), mas também inclui uma penalidade que é uma função crescente do número de parâmetros estimados.

A penalidade desincentiva sobreajuste, dado que o aumento do número de parâmetros no modelo quase sempre melhora a qualidade do ajuste. Um critério relacionado com AIC, mas incorporando um formalismo bayesiano foi proposto por Akaike (1974, que minimiza o

critério de informação de Akaike, chamado de BIC (critério de informação Bayesiano), este penaliza mais fortemente modelos com um maior número de parâmetros do que o AIC tendendo, dessa forma, selecionar modelos com um menor número de parâmetros. O critério BIC é dado pela seguinte expressão:

$$BIC(k) = n \ln \hat{\sigma}_a^2 - (n-k) \ln \left(1 - \frac{k}{n} \right) + k \ln \left[\frac{\left(\frac{\hat{\sigma}_z^2}{\hat{\sigma}_a^2} - 1 \right)}{k} \right]$$

em que $\hat{\sigma}_a^2$ é a estimativa de máxima verossimilhança de σ_a^2 ; $\hat{\sigma}_z^2$ é a variância da série; k é o número de parâmetros; n é o tamanho da série (Morettin; Toloi, 2006; Wei, 2006).

4.7 VARIÁVEIS GERADAS PELOS MODELOS DSSAT E APSIM

As tabelas a seguir mostram as variáveis geradas pelos modelos DSSAT e APSIM e que serviram de base para os cálculos do índice de perda.

Tabela. Descrição das variáveis geradas pelo modelo Dssat.

arquivo	gcm	rcp	genética	data plantio	solo	município	produtividade	total
1	base	17	médio	16	3	142	38	259008
2	base	17	precoce	16	3	142	38	259008
3	base	17	tardio	16	3	142	38	259008
4	had	45	médio	16	3	142	30	204480
5	had	45	precoce	16	3	142	30	204480
6	had	45	tardio	16	3	142	30	204480
7	had	85	médio	16	3	142	30	204480
8	had	85	precoce	16	3	142	30	204480
9	had	85	tardio	16	3	142	30	204480
10	max	45	médio	16	3	142	30	204480
11	max	45	precoce	16	3	142	30	204480
12	max	45	tardio	16	3	142	30	204480

13	max	85	médio	16	3	142	30	204480
14	max	85	precoce	16	3	142	30	204480
15	max	85	tardio	16	3	142	30	204480
16	med	45	médio	16	3	142	30	204480
17	med	45	precoce	16	3	142	30	204480
18	med	45	tardio	16	3	142	30	204480
19	med	85	médio	16	3	142	30	204480
20	med	85	precoce	16	3	142	30	204480
21	med	85	tardio	16	3	142	30	204480
22	min	45	médio	16	3	142	30	204480
23	min	45	precoce	16	3	142	30	204480
24	min	45	tardio	16	3	142	30	204480
25	min	85	médio	16	3	142	30	204480
26	min	85	precoce	16	3	142	30	204480
27	min	85	tardio	16	3	142	30	204480

Tabela. Descrição das variáveis geradas pelo modelo Apsim.

arquivo	gcm	rcp	genética	data plantio	solo	município	produtividade	total
1	base	17	médio	16	3	142	9	61344
2	base	17	precoce	16	3	142	9	61344
3	base	17	tardio	16	3	142	9	61344
4	had	45	médio	16	3	142	30	204480
5	had	45	precoce	16	3	142	30	204480
6	had	45	tardio	16	3	142	30	204480
7	had	85	médio	16	3	142	30	204480
8	had	85	precoce	15	3	142	30	191700
9	had	85	tardio	16	3	142	30	204480
10	max	45	médio	16	3	142	30	204480
11	max	45	precoce	16	3	142	30	204480
12	max	45	tardio	16	3	142	30	204480
13	max	85	médio	16	3	142	30	204480

14	max	85	precoce	16	3	142	30	204480
15	max	85	tardio	16	3	142	30	204480
16	med	45	médio	16	3	142	30	204480
17	med	45	precoce	16	3	142	30	204480
18	med	45	tardio	16	3	142	30	204480
19	med	85	médio	16	3	142	30	204480
20	med	85	precoce	16	3	142	30	204480
21	med	85	tardio	16	3	142	30	204480
22	min	45	médio	16	3	142	30	204480
23	min	45	precoce	16	3	142	30	204480
24	min	45	tardio	16	3	142	30	204480
25	min	85	médio	16	3	142	30	204480
26	min	85	precoce	15	3	142	30	191700
27	min	85	tardio	16	3	142	30	204480

Nota:

1. base: cenário atual
2. min: GCM de variação mínima em relação a LB - MIROC5
3. med: GCM de variação média em relação a LB - BNU-ESM
4. max: GCM de variação máxima em relação a LB - CanESM2
5. had: GCM adicional - HadGem2

4.8 SOBRE AS TABELAS

4.8.1 ARQUIVOS .TXT DAS PROBABILIDADES DE PERDA, MODELO DSSAT

Os 27 arquivos dos índices de perda gerados pelo modelo DSSAT estão subdivididos da seguinte forma: 24 arquivos com os oito cenários futuros – (4 GCMs e RCPs, 4.5 e 8.5), juntamente com 3 tipos de genética (médio, precoce e tardio); e, 3 arquivos com o cenário atual e os mesmos 3 tipos de genética. A codificação dos nomes obedece a seguinte ordem: modelo, gcm, rcp, genética e produto. Dessa forma, d1_had_45_pre_pp refere-se ao modelo dssat, gcm adicional - HadGem2, rcp 45, genética precoce e produto probabilidade de perda.

Por sua vez, em cada um dos arquivos há 6.816 linhas e 7 colunas, respectivamente, local, probabilidade de perda em 5 níveis de cobertura, sigla da distribuição ajustada. As colunas 2 a 6 correspondem as estimativas das probabilidades de perda ($t + 1$) separadas, respectivamente, em 16 datas de plantio, 3 tipos de solo e 142 municípios, totalizando 6.816 linhas. As datas de plantio seguem a seguinte ordenação:

Ordem	Dia	Mês
1	1	agosto
2	11	agosto
3	21	agosto
4	31	agosto
5	10	setembro
6	20	setembro
7	30	setembro
8	10	outubro
9	20	outubro
10	30	outubro
11	9	novembro
12	19	novembro
13	29	novembro
14	9	dezembro
15	19	dezembro
16	29	dezembro

Por sua vez, os solos foram subdivididos de acordo com suas texturas:

- 1 – argilosa
- 2 – média
- 3 – arenosa

4.8.2 ARQUIVOS .TXT DAS PROBABILIDADES DE PERDA, MODELO APSIM

Os 27 arquivos das probabilidades de perda estão subdivididos da seguinte forma: 24 arquivos com os oito cenários futuros – (4 GCMs e RCPs, 4.5 e 8.5), juntamente com 3 tipos de genética (médio, precoce e tardio); e, 3 arquivos com o cenário atual e os mesmos 3 tipos de genética. A codificação dos nomes obedece a seguinte ordem: modelo, gcm, rcp, genética e produto. Dessa forma, ap_had_45_pre_ip refere-se ao modelo apsim, gcm adicional - HadGem2, rcp 45, genética precoce e produto índice de perda.

Por sua vez, em cada um dos arquivos há 6.816 linhas e 7 colunas, respectivamente, local, probabilidade de perda em 5 níveis de cobertura, sigla da distribuição ajustada. As colunas 2 a 6 correspondem as estimativas das probabilidades de perda ($t + 1$) separadas, respectivamente, em 16 datas de plantio, 3 tipos de solo e 142 municípios, totalizando 6.816 linhas. As datas de plantio e os tipos de solo seguem a ordenação anterior

4.8.3 SOBRE OS MAPAS, MODELOS DSSAT E APSIM

Ao todo, foram geradas aproximadamente 920.000 probabilidades de perda para cada um dos dois modelos – dssat e apsim, gcm, rcp, genética, data de plantio, tipo de solo, município e nível de cobertura. A quantidade enorme de resultados impõe uma dificuldade adicional no sentido de divulgá-los na forma tabular. Nesse contexto, para facilitar a visualização agregou-se os resultados em mapas. Assim, gerou-se aproximadamente 6500 mapas dos índices de perda para cada um dos dois modelos.

Tabela. Resultados do modelo Dssat.

Arquivo	Probabilidade de Perda (PP)	Mapas PP
1	34080	240
2	34080	240
3	34080	240
4	34080	240
5	34080	240
6	34080	240
7	34080	240

8	34080	240
9	34080	240
10	34080	240
11	34080	240
12	34080	240
13	34080	240
14	34080	240
15	34080	240
16	34080	240
17	34080	240
18	34080	240
19	34080	240
20	34080	240
21	34080	240
22	34080	240
23	34080	240
24	34080	240
25	34080	240
26	34080	240
27	34080	240

Tabela. Resultados do modelo Apsim.

Arquivo	Probabilidade de Perda (PP)	Mapas PP
1	34080	240
2	34080	240
3	34080	240
4	34080	240
5	34080	240
6	34080	240
7	34080	240
8	31950	225
9	34080	240

10	34080	240
11	34080	240
12	34080	240
13	34080	240
14	34080	240
15	34080	240
16	34080	240
17	34080	240
18	34080	240
19	34080	240
20	34080	240
21	34080	240
22	34080	240
23	34080	240
24	34080	240
25	34080	240
26	31950	225
27	34080	240

4.9 OBSERVAÇÕES FINAIS

Calculou-se as probabilidades de perda para 142 dos 399 municípios do Paraná. Com o intuito de criar zonas homogêneas com os índices de perda optou-se por interpolar os resultados utilizando a média das produtividades de perda municipais ponderadas por suas respectivas áreas. Assim, criou-se as zonas homogêneas agregadas em microrregiões do estado.

A principal limitação da interpolação dos dados municipais em microrregiões é a perda da informação em nível municipal. Ou seja, as probabilidades de perda foram transformadas em uma média do comportamento dos municípios o que leva a redução da variabilidade do comportamento do risco. Na prática, pode ocorrer de existirem municípios com probabilidades de perda maiores do que aquela apontada na microrregião, e vice-versa. O fato de ponderar a média pelo tamanho da área municipal ameniza o problema, na medida

em que dá um peso maior para os municípios mais representativos em função de seu tamanho.

4.10 REFERÊNCIAS

Akaike, H. (1974) A new look at the statistical model identification, *Automatic Control, IEEE Transactions on* 19, 716–723.

Alizadeh, M.; Emadi, M.; Doostparast, M.; Cordeiro, G.M.; Ortega, E.M.; Pescim, R.R. (2015) A new family of distributions: the kumaraswamy odd log-logistic, properties and applications, *Hacettepe Journal of Mathematics and Statistics*, 44.

Azzalini, A. A class of distributions wich includes the normal ones, *Scand. J. Statist.* 12 (1985), pp. 171-178.

Azzalini, A.; Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 367–389.

Braga, A. S.; Cordeiro, G.M.; Ortega, E.M.; Cruz, J.N. (2016) The odd log–logistic normal distribution: Theory and applications in analysis of experiments, *Journal of Statistical Theory and Practice*, 1–25.

Cruz, J.N. (2016) A nova família de distribuições odd log-logística: teoria e aplicações, Ph.D. diss., Escola Superior de Agricultura “Luiz de Queiroz.

Moretin, P.A.; Toloi, C.M. *Previsão de séries Temporais. Atual.* São Paulo. 1985.

Ozaki, V.A.; Silva, R. S. (2009) Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *Journal of Applied Statistics*, 36, 443-452.

Wei, W.W.S. *Time Series Analysis*

5. SEÇÃO 4 - ÍNDICE DE PERDA E ESTRATIFICAÇÃO MUNICIPAL DO RISCO BASEADO EM ÍNDICES DE PERDA

5.1 INTRODUÇÃO

Esta seção trata dos objetivos referentes ao produto 5 do projeto e tem como objetivos específicos o seguinte: “Indicar, em mapas e tabelas, o índice de perdas: análise atuarial com estratificação municipal do risco baseado em índices de perdas”.

5.2 O ÍNDICE DE PERDA

O índice de perda foi gerado tendo-se como base as informações de produtividade agrícola dos produtos 2 e 3 do Termo de Referência supramencionado. O mercado segurador, em geral, utiliza um índice de perda que é base para o cálculo das indenizações. Esse índice depende da produtividade garantida, calculada pela multiplicação do nível de cobertura e a produtividade esperada, e a produtividade observada no ano t. Os índices de perda foram gerados tendo-se por base a seguinte equação:

$$\text{Índice de Perda (IP)} = (y_g - y_o)/y_g$$

em que y_o é a produtividade observada no ano t. Para o cálculo dos índices de perda utilizou as mesmas nove bases climáticas, ou seja, o clima atual (linha de base) e os oito cenários futuros/gcm combinado com os 2 RCPs, 4.5 e 8.5, respectivamente, além dos dois modelos agrícolas Dssat e Apsim. As tabelas abaixo resumem a organização dos dados para calcular os índices de perda.

5.3 VARIÁVEIS GERADAS PELOS MODELOS DSSAT E APSIM

As tabelas a seguir mostram as variáveis geradas pelos modelos DSSAT e APSIM e que serviram de base para os cálculos do índice de perda.

Tabela 1. Descrição das variáveis geradas pelo modelo Dssat.

arquivo	gcm	rcp	genética	data plântio	solo	município	produtividade	total
1	base	17	médio	16	3	142	38	259008
2	base	17	precoce	16	3	142	38	259008
3	base	17	tardio	16	3	142	38	259008
4	had	45	médio	16	3	142	30	204480
5	had	45	precoce	16	3	142	30	204480
6	had	45	tardio	16	3	142	30	204480
7	had	85	médio	16	3	142	30	204480
8	had	85	precoce	16	3	142	30	204480
9	had	85	tardio	16	3	142	30	204480
10	max	45	médio	16	3	142	30	204480
11	max	45	precoce	16	3	142	30	204480
12	max	45	tardio	16	3	142	30	204480
13	max	85	médio	16	3	142	30	204480
14	max	85	precoce	16	3	142	30	204480
15	max	85	tardio	16	3	142	30	204480
16	med	45	médio	16	3	142	30	204480
17	med	45	precoce	16	3	142	30	204480
18	med	45	tardio	16	3	142	30	204480
19	med	85	médio	16	3	142	30	204480
20	med	85	precoce	16	3	142	30	204480
21	med	85	tardio	16	3	142	30	204480
22	min	45	médio	16	3	142	30	204480
23	min	45	precoce	16	3	142	30	204480
24	min	45	tardio	16	3	142	30	204480
25	min	85	médio	16	3	142	30	204480
26	min	85	precoce	16	3	142	30	204480
27	min	85	tardio	16	3	142	30	204480

Tabela 2. Descrição das variáveis geradas pelo modelo Apsim.

arquivo	gcm	rcp	genética	data plantio	solo	município	produtividade	total
1	base	17	médio	16	3	142	9	61344
2	base	17	precoce	16	3	142	9	61344
3	base	17	tardio	16	3	142	9	61344
4	had	45	médio	16	3	142	30	204480
5	had	45	precoce	16	3	142	30	204480
6	had	45	tardio	16	3	142	30	204480
7	had	85	médio	16	3	142	30	204480
8	had	85	precoce	15	3	142	30	191700
9	had	85	tardio	16	3	142	30	204480
10	max	45	médio	16	3	142	30	204480
11	max	45	precoce	16	3	142	30	204480
12	max	45	tardio	16	3	142	30	204480
13	max	85	médio	16	3	142	30	204480
14	max	85	precoce	16	3	142	30	204480
15	max	85	tardio	16	3	142	30	204480
16	med	45	médio	16	3	142	30	204480
17	med	45	precoce	16	3	142	30	204480
18	med	45	tardio	16	3	142	30	204480
19	med	85	médio	16	3	142	30	204480
20	med	85	precoce	16	3	142	30	204480
21	med	85	tardio	16	3	142	30	204480
22	min	45	médio	16	3	142	30	204480
23	min	45	precoce	16	3	142	30	204480
24	min	45	tardio	16	3	142	30	204480
25	min	85	médio	16	3	142	30	204480
26	min	85	precoce	15	3	142	30	191700
27	min	85	tardio	16	3	142	30	204480

5.4 SOBRE AS TABELAS

5.4.1 ARQUIVOS .TXT DOS ÍNDICES DE PERDA, MODELO DSSAT

Os 27 arquivos dos índices de perda gerados pelo modelo DSSAT estão subdivididos da seguinte forma: 24 arquivos com os oito cenários futuros – (4 GCMs e RCPs, 4.5 e 8.5), juntamente com 3 tipos de genética (médio, precoce e tardio); e, 3 arquivos com o cenário atual e os mesmos 3 tipos de genética. A codificação dos nomes obedece a seguinte ordem: modelo, gcm, rcp, genética e produto. Dessa forma, d1_had_45_pre_ip refere-se ao modelo dssat, gcm adicional - HadGem2, rcp 45, genética precoce e produto índice de perda.

Cada um dos 27 arquivos possui 6 colunas, respectivamente, local e 5 índices de perda para cada um dos 5 níveis de cobertura. Ao todo, são 224.928 linhas para o cenário atual e 170.400 linhas para os oito cenários futuros, ambos com 3 tipos de genética.

As estimativas dos índices de perda têm início em 1985, para o cenário atual, e com início em 2045, para os cenários futuros, separadas, respectivamente, em 16 datas de plantio, 3 tipos de solo e 142 municípios, totalizando 224.928 e 170.400 linhas dependendo do cenário. As datas de plantio seguem a seguinte ordenação:

Por sua vez, os solos foram subdivididos de acordo com suas texturas:

1 – argilosa

2 – média

3 – arenosa

Em geral, as tabelas seguem os mesmos padrões do produto 4. A principal diferença se refere ao fato dos índices de perda terem sido calculados não apenas para o último ano da série, mas para todos os anos, com início em $t + 5$. Devido à grande número de linhas para os

3 tipos de solo, optou-se por ocultar as tabelas e enviá-las como um link para efetuar o download de todos os resultados.

5.4.2 ARQUIVOS .TXT DOS ÍNDICES DE PERDA, MODELO APSIM

Os 27 arquivos dos índices de perda gerados pelo modelo APSIM estão subdivididos da seguinte forma: 24 arquivos com os oito cenários futuros – (4 GCMs e RCPs, 4.5 e 8.5), juntamente com 3 tipos de genética (médio, precoce e tardio); e, 3 arquivos com o cenário atual e os mesmos 3 tipos de genética. A codificação dos nomes obedece a seguinte ordem: modelo, GCM, RCP, genética e produto. Dessa forma, `ap_had_45_pre_ip` refere-se ao modelo APSIM, GCM adicional - HadGem2, RCP 45, genética precoce e produto índice de perda.

Cada um dos 27 arquivos possui 6 colunas, respectivamente, local e 5 índices de perda para cada um dos 5 níveis de cobertura, totalizando 27.264 linhas para o cenário atual e 170.400 linhas para os oito cenários futuros ambos com 3 tipos de genética.

As estimativas dos índices de perda têm início em 2013, para o cenário atual, e com início em 2045, para os cenários futuros, separadas, respectivamente, em 16 datas de plantio, 3 tipos de solo e 142 municípios, totalizando 27.264 e 170.400 linhas dependendo do cenário. As datas de plantio seguem a mesma ordenação e as texturas também.

Em geral, as tabelas seguem os mesmos padrões do produto 4. A principal diferença se refere ao fato dos índices de perda terem sido calculados não apenas para o último ano da série, mas para todos os anos, com início em $t + 5$. Devido ao grande número de linhas para os 3 tipos de solo, optou-se por ocultar as tabelas e enviá-las como um link para efetuar o download de todos os resultados.

5.4.3 SOBRE OS MAPAS, MODELOS DSSAT E APSIM

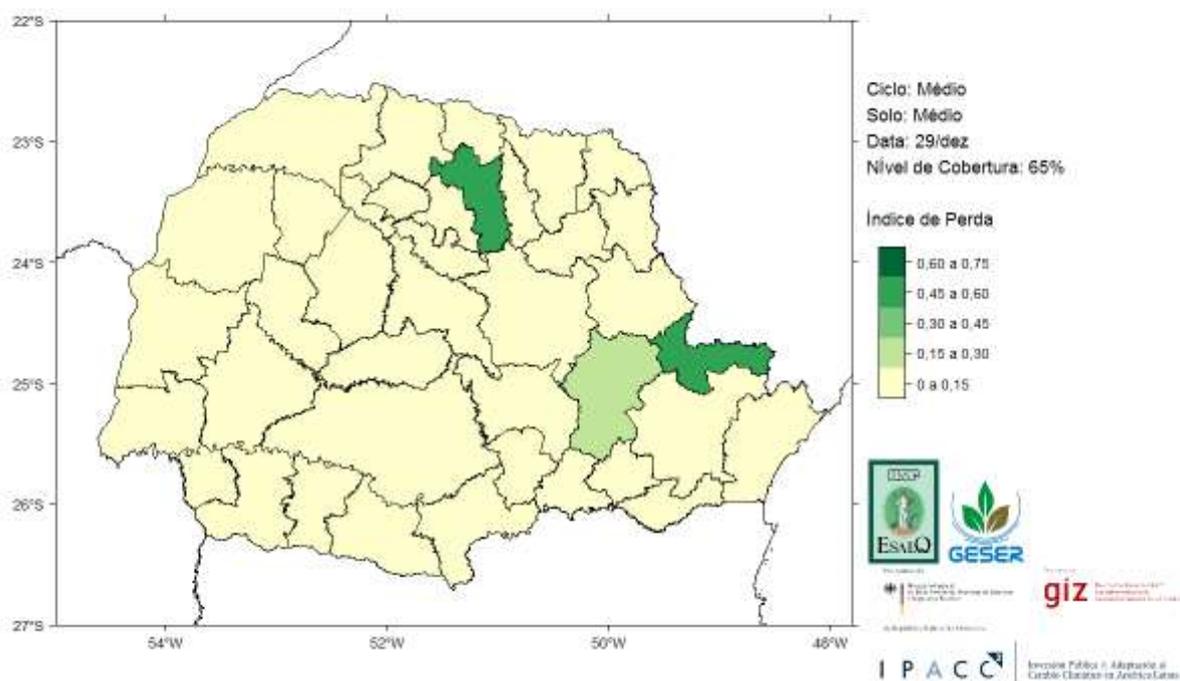
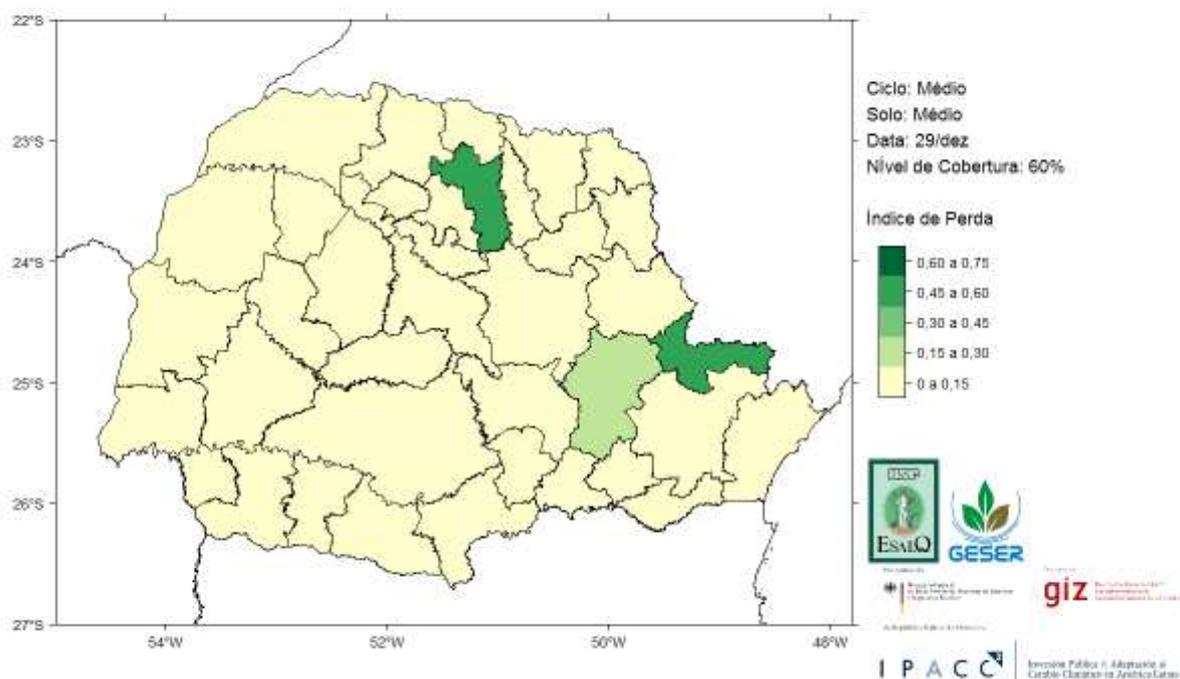
Ao todo, foram gerados aproximadamente 920.000 índices de perda para cada um dos dois modelos – DSSAT E APSIM, GCM, RCP, genética, data de plantio, tipo de solo, município e nível de cobertura. A quantidade enorme de resultados impõe uma dificuldade adicional no sentido de divulgá-los na forma tabular. Nesse contexto, para facilitar a visualização agregou-se os resultados em mapas. Assim, gerou-se aproximadamente 6500 mapas dos índices de perda para cada um dos dois modelos.

Tabela 3. Resultados do modelo Dssat.

Arquivo	Índice de Perda (IP)	Mapas IP
1	34080	240
2	34080	240
3	34080	240
4	34080	240
5	34080	240
6	34080	240
7	34080	240
8	34080	240
9	34080	240
10	34080	240
11	34080	240
12	34080	240
13	34080	240
14	34080	240
15	34080	240
16	34080	240
17	34080	240

18	34080	240
19	34080	240
20	34080	240
21	34080	240
22	34080	240
23	34080	240
24	34080	240
25	34080	240
26	34080	240
27	34080	240

Os mapas finalizados para uma determinada combinação de variáveis seguem o seguinte formato, respectivamente, para os níveis de cobertura de 60 a 80%, em múltiplos de 5%. Nos exemplos que seguem, têm-se o ciclo médio, solo com textura média, data de plantio 29 de dezembro e os respectivos níveis de cobertura. A agregação dos valores gerou intervalos entre 0 a 75% de índice de perda, dependendo da cobertura. Neste caso, o índice de perda é maior quanto maior o nível de cobertura.



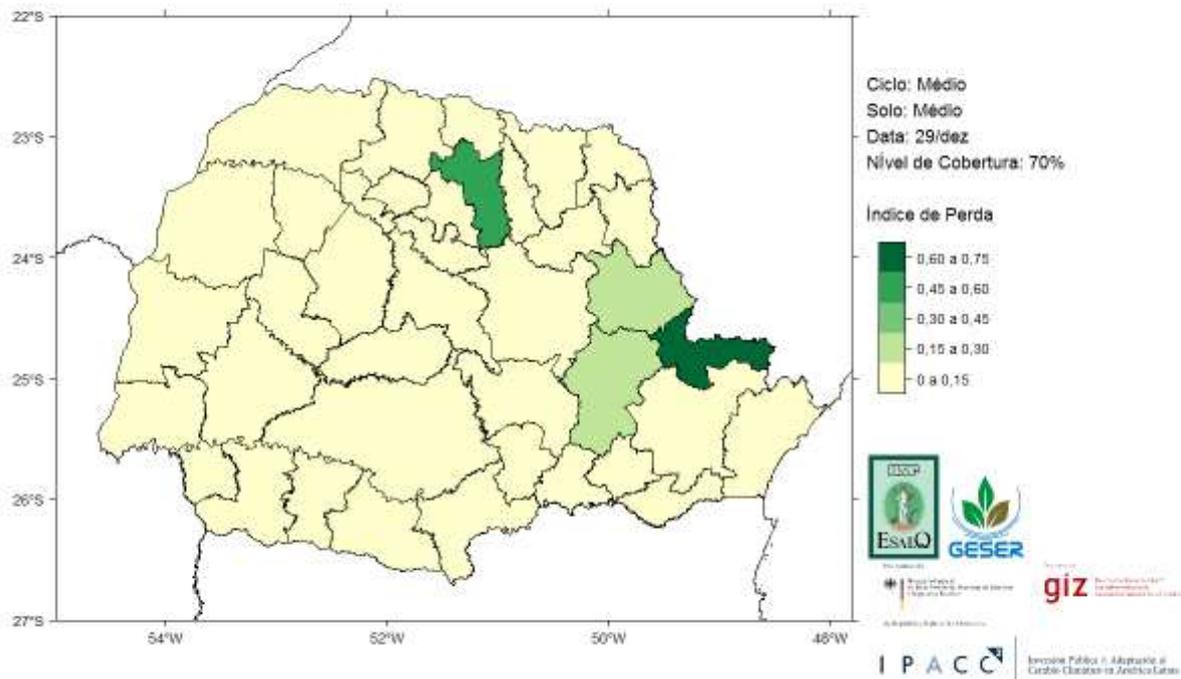
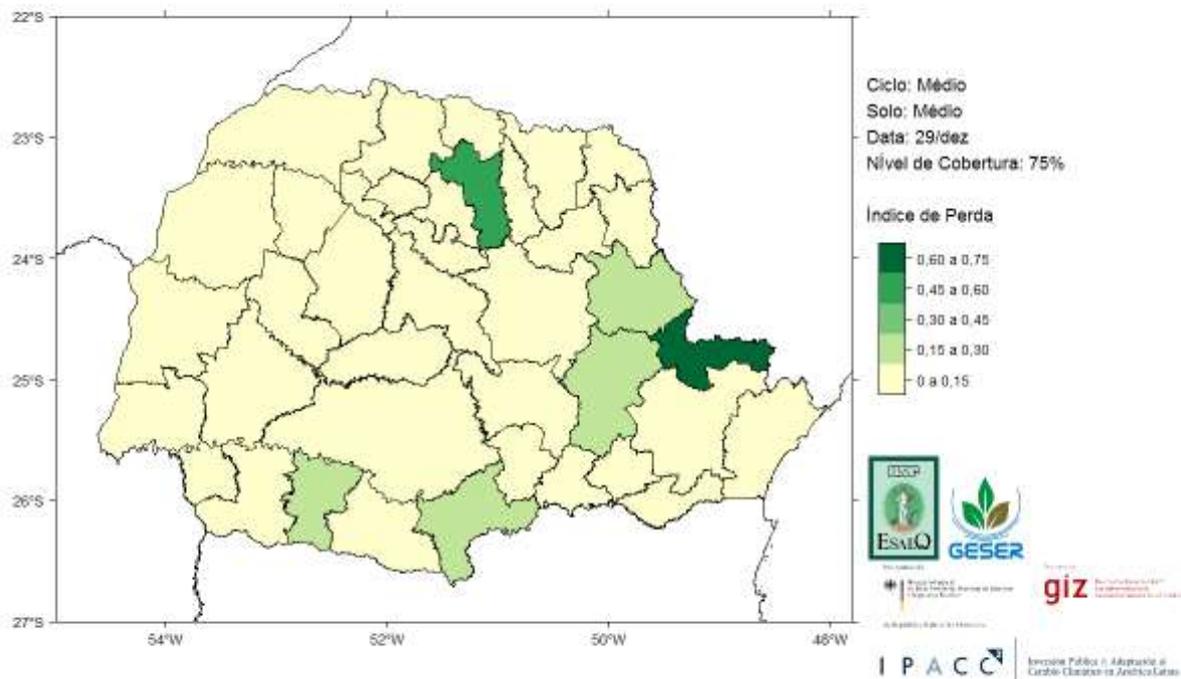


Figura 1. Representação espacial do índice de perda simulada para a cultura da soja no Paraná.



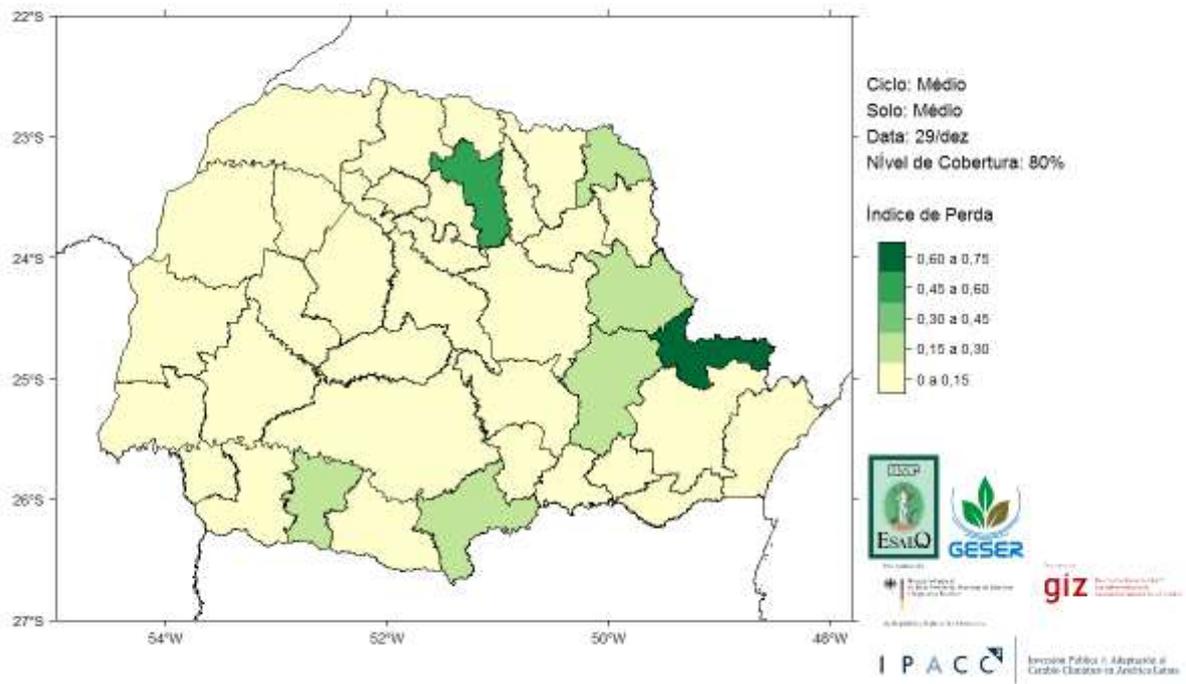
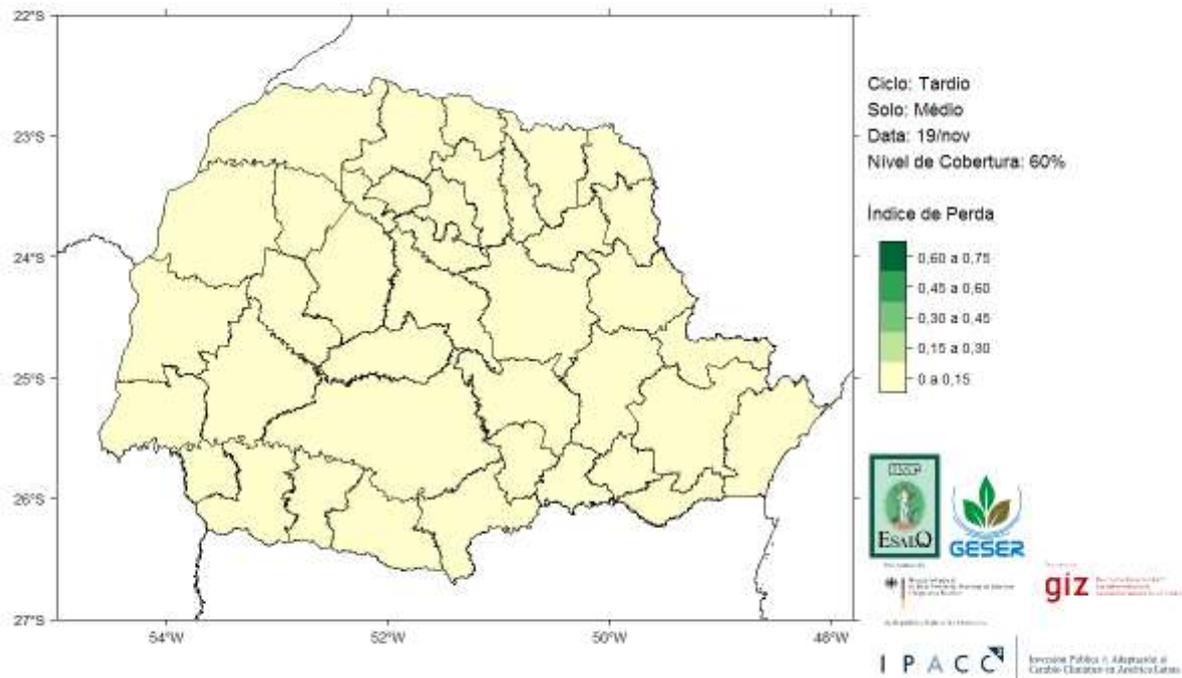


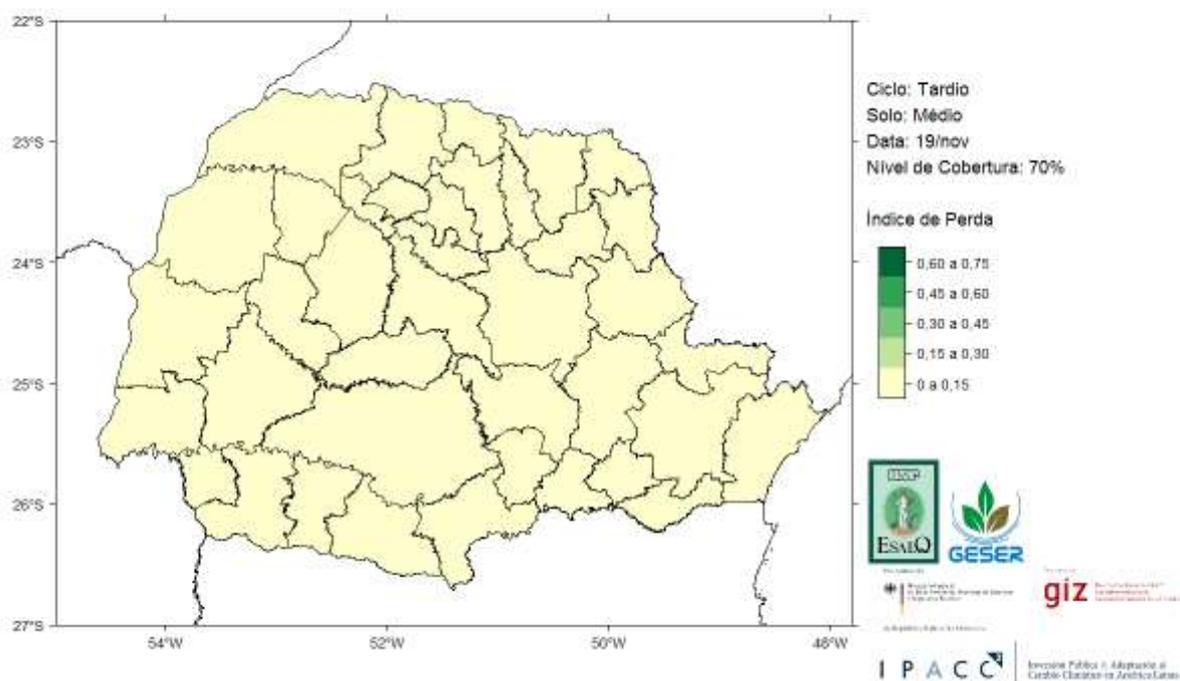
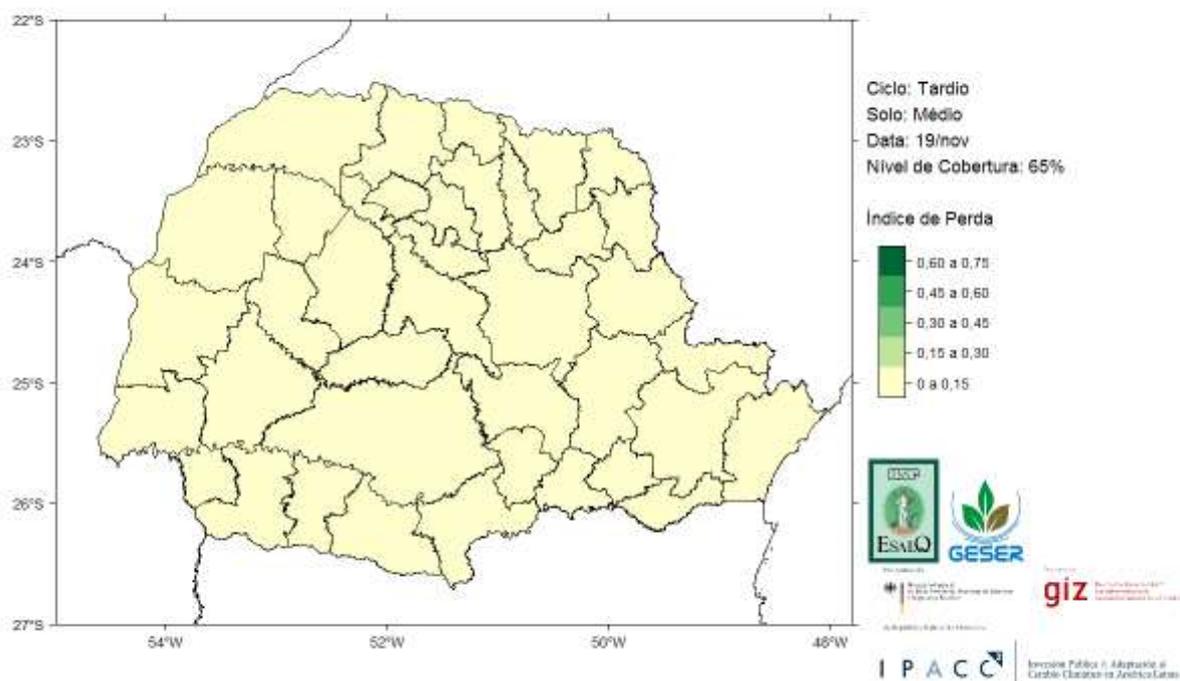
Tabela 5. Resultados do modelo Apsim.

Arquivo	Índice de Perda (IP)	Mapas IP
1	34080	240
2	34080	240
3	34080	240
4	34080	240
5	34080	240
6	34080	240
7	34080	240
8	31950	225
9	34080	240
10	34080	240
11	34080	240
12	34080	240
13	34080	240
14	34080	240
15	34080	240
16	34080	240
17	34080	240
18	34080	240
19	34080	240
20	34080	240
21	34080	240
22	34080	240
23	34080	240
24	34080	240
25	34080	240
26	31950	225
27	34080	240

Os mapas finalizados para uma determinada combinação de variáveis seguem o seguinte formato, respectivamente, para os níveis de cobertura de 60 a 80%, em múltiplos de

5%. Nos exemplos que seguem, tem-se o ciclo tardio, solo com textura média, data de plantio 19 de novembro e os respectivos níveis de cobertura. A agregação dos valores gerou intervalos entre 0 a 15% de índice de perda, dependendo da cobertura. Neste caso, o índice de perda é maior quanto maior o nível de cobertura.





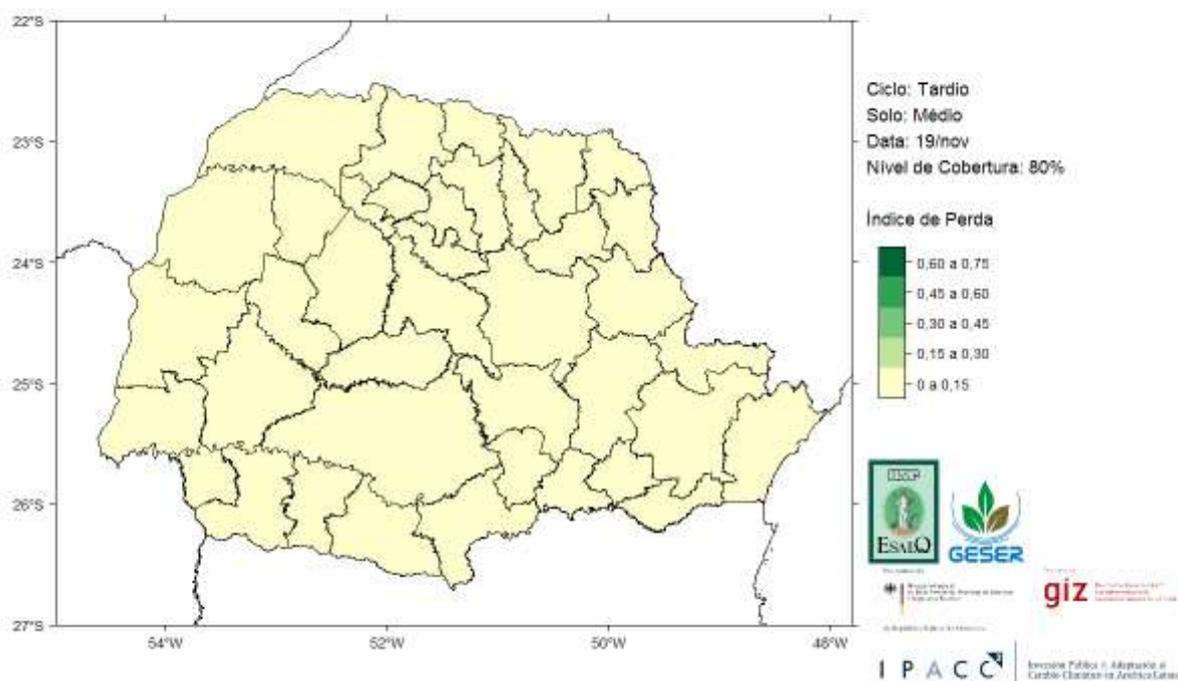
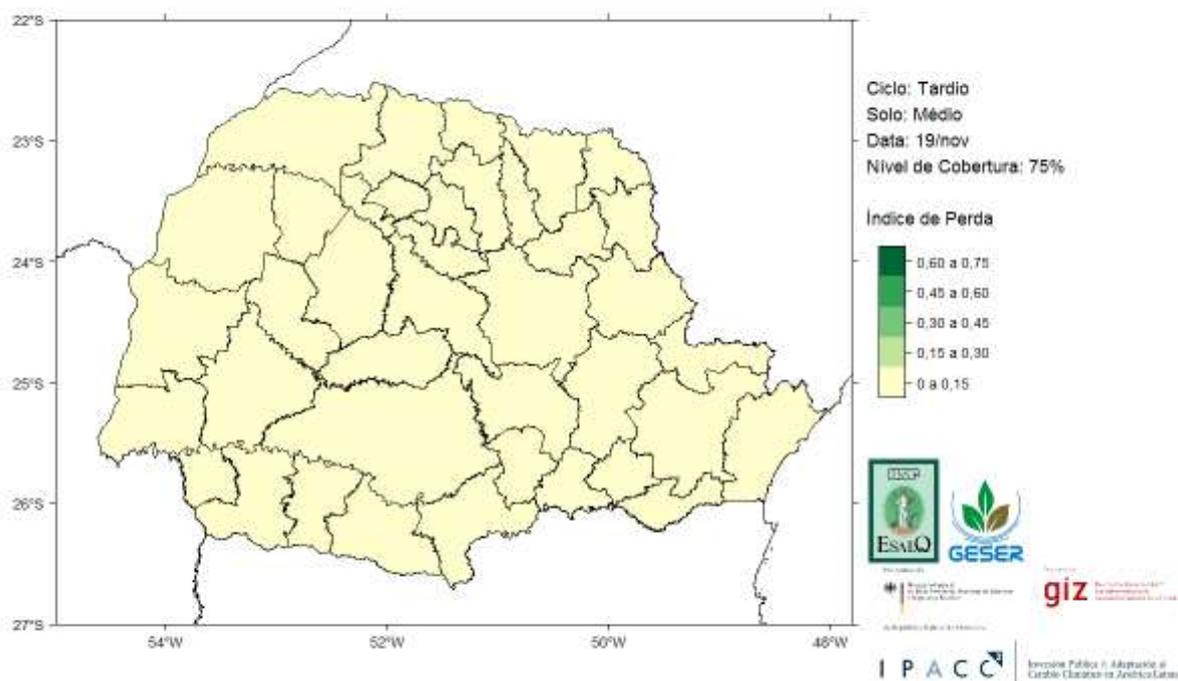


Figura 2. Representação espacial do índice de perda simulada para a cultura da soja no Paraná.

5.5 OBSERVAÇÕES FINAIS

Calculou-se os índices de perda para 142 dos 399 municípios do Paraná. Com o intuito de criar zonas homogêneas com os índices de perda optou-se por interpolar os resultados utilizando a média das produtividades de perda municipais ponderadas por suas respectivas áreas. Assim, criou-se as zonas homogêneas agregadas em microrregiões do estado.

A principal limitação da interpolação dos dados municipais em microrregiões é a perda da informação em nível municipal. Ou seja, os índices de perda foram transformados em uma média do comportamento dos municípios o que leva a redução da variabilidade do comportamento do risco. Na prática, pode ocorrer de existirem municípios com índices de perda maiores do que aquela apontada na microrregião, e vice-versa. O fato de ponderar a média pelo tamanho da área municipal ameniza o problema, na medida em que dá um peso maior para os municípios mais representativos em função de seu tamanho.

6. SEÇÃO 5 - PROPOSTA DE METODOLOGIA DE PRECIFICAÇÃO

6.1 INTRODUÇÃO

O objetivo desta seção está relacionada ao produto 6 do projeto e busca “*Desenvolver proposta metodológica de precificação do risco: apresentar proposta de metodologia de atuarial de tarifação do risco para o Proagro/Seguro Rural.*”

6.2 CONTEXTUALIZAÇÃO

O seguro agrícola teve início no país em meados do início do século XX no Estado de São Paulo e operacionalizado pelo Instituto Agrônômico do Estado. Sem a participação de seguradoras privadas, o prêmio do seguro contra granizo foi incorporado ao preço de venda da semente de algodão e formado uma espécie de fundo contra o granizo.

O Decreto-lei n. 1.186, de 3 de abril de 1939, criou o Instituto de Resseguros do Brasil (IRB) com o objetivo de regular as operações de resseguro, cosseguro e retrocessão e desenvolver a atividade seguradora de modo geral. A partir dessa data todas as operações de seguros eram obrigadas a ressegurar no IRB, inclusive no ramo rural.

Com o aumento de capacidade proporcionada pelo IRB abriu-se a oportunidade para se pensar na criação de uma empresa estatal que pudesse massificar o seguro agrícola em todo o território nacional. Já nessa época havia a preocupação de se elaborar um mecanismo de seguro agrícola que pudesse proteger não apenas o produtor agrícola e sua produção, mas também outras instituições importantes no processo produtivo, por exemplo, instituições financeiras que forneciam o crédito para o setor.

Apesar da importância da necessidade da criação do seguro agrícola, não havia um arcabouço legal para que pudesse nortear o mercado e o próprio Governo. Deste modo, no ano de 1954, teve início a regulamentação e o estabelecimento de normas para o ramo de seguro rural no país. A Lei n. 2.168, de janeiro de 1954, instituiu que o seguro rural se destinava à preservação das colheitas e dos rebanhos contra a eventualidade de risco que lhes são peculiares.

Permitiu que o IRB realizasse os estudos, levantamentos e planejamentos para a início do seguro rural em todo o país em colaboração com os serviços técnicos das repartições federais, estaduais, municipais, de autarquias e dos estabelecimentos bancários oficiais de financiamento à agricultura e pecuária, e também, elaborasse as condições das apólices e taxas de prêmios de seguros, e que, depois de provadas pelo Departamento Nacional de Seguros Privados e Capitalização, fossem colocadas em vigor, mediante decretos.

A mesma lei ainda autorizava o Poder Executivo a organizar uma sociedade de economia mista, com sede e foro na cidade do Rio de Janeiro, com o objetivo de desenvolver progressivamente operações de seguros rurais, sob a denominação de Companhia Nacional de Seguro Agrícola (CNSA). O Decreto n. 35.370, de 12 de abril de 1954, regulamentou as operações de seguro rural no país e conferiu a CNSA à exploração dessas operações.

Todas as responsabilidades não assumidas pelas companhias seguradoras ou pela CNSA eram resseguradas no IRB. Se as companhias não encontrassem resseguro das responsabilidades assumidas poderiam, através do IRB, realizar tal operação no exterior. As seguradoras poderiam, caso o IRB rejeitasse ou cancelasse tais responsabilidades, ressegurar em outras firmas no país ou na própria CNSA.

Ao longo de treze anos a CNSA encontrou uma série de dificuldades que culminaram na sua dissolução pelo Decreto-lei n. 73, de 21 de novembro de 1966. Um dos principais

problemas encontrados pela Companhia foi a falta de conhecimento da atividade a ser segurada.

Características importantes da produção agrícola e da pecuária não foram levadas em conta, por exemplo, as peculiaridades regionais tanto do ponto de vista da cultura e do manejo do produtor quanto das diferenças climáticas e edáficas, sem as quais não há como, de modo razoável, estabelecer prêmios acessíveis aos produtores e manter a carteira atuarialmente equilibrada. Na época não havia uma base de dados para se realizar os cálculos atuariais nem mesmo metodologias adequadas para realizar a precificação do risco.

Durante quase uma década os produtores ficaram expostos às oscilações e adversidades climáticas, sem qualquer mecanismo de seguros. Apenas em 1975, quase uma década depois, o Governo criou o Programa de Garantia da Atividade Agropecuária (PROAGRO) como um instrumento para garantir o crédito fornecido ao setor.

Pouco se aproveitou da malsucedida experiência da CNSA. Os anos iniciais do PROAGRO foram catastróficos no que se refere aos resultados atuariais, ou seja, a carteira foi altamente deficitária. Basicamente, os mesmos problemas da CNSA se repetiram no contexto do PROAGRO, principalmente na ausência de uma base de dados e de metodologias atuariais de precificação.

Até meados de 2000, o PROAGRO foi o único instrumento utilizado pelo setor para se proteger contra eventos climáticos adversos. Nesse período, destaca-se a atuação de algumas poucas seguradoras estatais pertencentes a governos estaduais, mas com uma abrangência bastante limitada.

No final de 1990, algumas seguradoras privadas começaram a iniciar suas operações no seguro rural, tendo em vista a possível privatização do IRB e a possibilidade de abertura do mercado ressegurador, até então monopólio do IRB.

A abertura do mercado foi importante na atração de resseguradoras internacionais aumentando a capacidade das seguradoras e o interesse por ramos não tradicionais, por exemplo, o ramo rural. Porém, a abertura ocorreria apenas em 2007. Apesar disso, o mercado privado operava a modalidade agrícola de forma cautelosa, devido a falta de experiência, ausência de dados estatísticos de risco e de metodologias atuariais específicas ao setor.

A sanção da Lei 10.823 de 2003, que criou o Programa de Subvenção ao Prêmio do Seguro Rural (PSR), foi um marco legal importante para o florescimento de um mercado privado de seguro rural. Novas seguradoras iniciaram suas operações após a sanção da lei supramencionada.

De forma geral, o Governo forneceu ao setor privado o apoio necessário para operar o seguro rural, mesmo sem uma base histórica necessária para a quantificação e precificação de risco, o que viria a ter início após anos mais tarde mediante o acúmulo de dados pelas companhias seguradoras. No entanto, essa base de dados é confidencial a cada seguradora e restrita apenas as operações realizadas por cada companhia.

A única fonte de dados pública consultada pelo mercado pertence ao Instituto Brasileiro de Geografia e Estatística (IBGE), por meio da “produção agrícola municipal – PAM”. Apesar de possuir uma série de variáveis coletadas por um período relativamente longo, por exemplo, produção agrícola, área e produtividade, os dados estão agregados em nível municipal o que dificulta sua utilização pelo mercado segurador, na medida em que agrega produtores de diferentes perfis de risco em uma única informação.

O mesmo ocorre com os relatórios circunstanciados do PROAGRO publicados pelo BACEN. Os dados presentes nesse relatório são agregados e com uma nomenclatura não usual pelo mercado segurador, dificultando sua comparação com as operações do mercado.

Ademais, variáveis importantes não são coletadas pelos operadores do programa, por exemplo, dados de produtividade observada que são essenciais para os cálculos atuariais. Mesmo assim, os dados individualizados poderiam ser utilizados como “proxy” dos dados de seguros, na medida em que as operações se assemelham em alguns aspectos. No entanto, os micro dados não são públicos por razões de confidencialidade.

Por fim, pode-se perceber que o processo histórico de evolução do seguro rural e PROAGRO revelam problemas que até hoje não foram superadas. A ausência de dados individualizados das unidades seguradas, de caráter público, e a uso de metodologias adaptadas de outros ramos, inadequadas às peculiaridades do setor agrícola, são dois dos principais problemas enfrentados para o desenvolvimento do seguro agrícola no país.

No que concerne a ausência de dados individualizados e públicos, o desenvolvimento de um banco de dados com informações úteis ao seguro rural é um esforço que precisa ser respaldado tanto pelo Governo quanto pela iniciativa privada e entidades representativas dos produtores.

Um grande passo foi dado com a criação do “Atlas do Seguro Rural” pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e disponibilizado no endereço (<http://indicadores.agricultura.gov.br/atlasdoseguro/index.htm>). Ele representa um grande avanço na publicidade dos dados das operações subvencionadas e esforço em fornecer maior transparência à política pública.

Nessa plataforma é possível encontrar uma diversidade de informações relacionadas às operações de seguro rural enquadradas no Programa de Subvenção ao Prêmio, por exemplo, número de produtores atendidos, número de apólices, área segurada, importância segurada, valor de prêmio, valor de subvenção, produtividade segurada, taxas de prêmio, agregadas por município, por seguradora, desde 2006, na forma tabular ou mapas. Falta ainda

incluir as variáveis de sinistros, por exemplo, número total de sinistros e valor das indenizações. Os microdados podem ser acessados por meio de solicitação ao MAPA.

O segundo problema apontado e objetivo principal deste relatório é descrever uma proposta de metodologia para a precificação de risco. Para isso, o relatório apresenta a seguinte estrutura. A próxima seção caracteriza os produtores de seguro agrícola e Proagro. A seção 3 apresenta a metodologia atuarial de precificação de risco detalhando em suas subseções as distribuições de probabilidade que podem ser utilizadas para se calcular as taxas. A seção 4 revela a formulação para se determinar as taxas de prêmio, no contexto univariado e bivariado. E, por fim, a seção 5 revela a forma de se modelar no caso bivariado (seguro renda).

6.3 CARACTERIZAÇÃO DOS PRODUTOS DE SEGURO AGRÍCOLA E PROAGRO

Os produtos mais conhecidos e comercializados pelo mercado segurador são o seguro agrícola de custeio e o seguro agrícola de produtividade. Ambos possuem muitas características semelhantes, mas também se diferenciam em outros aspectos.

As principais semelhanças referem-se aos riscos cobertos, níveis de cobertura, produtividade garantida e na forma de avaliação dos sinistros. Os principais riscos cobertos pelas apólices de seguros são:

- chuva excessiva;
- granizo;
- geada;
- incêndio;
- raio;

- seca;
- tromba d'água;
- ventos fortes;
- ventos frios; e
- variação excessiva de temperatura.

A gama de coberturas ofertadas varia conforme a seguradora. A outra variável, nível de cobertura (NC), é outro importante parâmetro do seguro agrícola. Ele é um percentual que varia entre 60 a 80%, em múltiplos de 5, e ser escolhido pelo produtor. A multiplicação do nível de cobertura pela produtividade esperada, em quilogramas por hectare ou sacas por hectare, resulta na produtividade garantida na apólice de seguros. Quanto maior o NC maior será a responsabilidade assumida pela seguradora e maior será a taxa cobrada pelo seguro. Por outro lado, o produtor terá uma proteção maior proporcionada pelo seguro.

A produtividade garantida é o gatilho que dispara o processo indenizatório. Se a produtividade observada ao longo da safra for menor do que a produtividade garantida o produtor terá o direito de receber a indenização. Existem duas situações nesse caso: perda total e perda parcial.

A primeira compreende as perdas decorrentes de um ou mais riscos cobertos que inviabilizem a continuidade da atividade. De outra forma, ocorre perda total quando ocorre morte em 100% das plantas na unidade segurada ou comprometimento total da produção. Por outro lado, as perdas parciais são caracterizadas quando os prejuízos decorrentes dos riscos cobertos não comprometem a continuidade da atividade.

Dessa forma, percebe-se que a variável que caracteriza o prejuízo é a produtividade agrícola. Apesar do produto cobrir perdas causadas por eventos climáticos adversos, o

prejuízo é mensurado pela produtividade agrícola que é consequência não apenas dos fatores climáticos mas também das condições de solo, relevo, tipo de cultivar e manejo do produtor.

Por outro lado, os dois tipos de seguro diferenciam-se quanto ao cálculo do limite máximo de indenização (LMI). Enquanto no seguro de custeio o LMI está associado ao custeio por hectare gasto na lavoura, o seguro de produtividade cobre a produtividade esperada multiplicada pela área total a ser segurada e por um preço pré-estabelecido em contrato que permanece fixo até o final da vigência do contrato. Em geral a responsabilidade assumida pela seguradora no seguro de produtividade é maior do que no seguro de custeio.

Por outro lado, o PROAGRO assemelha-se ao seguro de custeio, mas com algumas diferenças. As mais importantes para os objetivos deste relatório referem-se ao LMI total ou também chamado de “valor enquadrado”, no PROAGRO, e ausência de níveis de cobertura.

No PROAGRO TRADICIONAL o LMI varia entre 70 a 100%, dependendo da adesão ou não ao PROAGRO nos 36 meses anteriores à data de enquadramento, ou deferimento de cobertura (indenização) referente ao último enquadramento do mesmo empreendimento.

Outro aspecto relevante é a ausência de níveis de cobertura. Em geral, o seguro agrícola não possui franquia ou dedutível associado ao seguro de custeio e o seguro de produtividade. No entanto, a existência do NC nesses produtos funciona de forma semelhante à franquia ou dedutível. Ou seja, é a parte em que o segurado é o responsável pelo seu próprio risco.

Na ausência de franquia, um NC de 100% sobre a produtividade esperada desincentivaria as boas práticas agrícolas, tais como, o uso correto de defensivos e fertilizantes, aumentando a probabilidade de perda da produtividade.

Por sua vez, o PROAGRO possui um sistema de bônus/málus que premia ou penaliza dependendo da comunicação ou não da perda. Nesse caso, a taxa de prêmio do PROAGRO, denominada alíquota do adicional, é reduzida em 0,25 ponto percentual para cada ano safra sem comunicação de perda e acrescida em 0,50 ponto percentual quando ocorre a comunicação de perda. Na prática é o mesmo princípio do NC no caso do seguro agrícola no sentido de induzir o seguro a boas práticas de manejo e redução do risco moral.

6.4 METODOLOGIA ATUARIAL DE PRECIFICAÇÃO DE RISCO

Desde as primeiras iniciativas em operar o seguro agrícola, o desenvolvimento de métodos específicos que resultassem em taxas atuarialmente equilibradas foi um grande desafio. A atividade agrícola possui uma série de peculiaridades que dificulta sobremaneira a quantificação e precificação do risco.

As unidades seguradas podem apresentar correlação espacial e dependência temporal e raramente são homogêneas. Ademais é comum o setor apresentar prejuízos de natureza catastrófica ou generalizada. Esses fatores dificultam a precificação, e em última instância o próprio equilíbrio atuarial da carteira, pois violam alguns princípios de segurabilidade.

Geralmente, os ramos tradicionais do seguro vida e não vida possuem métodos atuariais já conhecidos e testados pelo mercado segurador (Booth et al., 1999; Hart et al., 1996). Porém, quando novos produtos são desenvolvidos ou novos nichos de atuação são encontrados, depara-se com problemas de ausência de dados estatísticos e também dificuldades de natureza metodológica dificultando o trabalho do atuário. Notoriamente, o seguro agrícola é uma das modalidades mais complexas de precificar o risco em virtude das peculiaridades supramencionadas.

Ao redor do mundo onde o seguro agrícola é operacionalizado o cálculo preciso e equilibrado da taxa de prêmio é fundamental para o sucesso do seguro. Isso porque o seguro está fundamentalmente baseado no conceito de mutualismo, em que os prêmios são proporcionais ao risco individual. Isso garante o equilíbrio atuarial da carteira. Outro ponto importante refere-se a severidade e a frequência dos sinistros. Todos esses fatores influenciam o processo de precificação.

Independente do modelo de governança - públicas, privadas ou mistas - uma das preocupações centrais no desenvolvimento de um mercado de seguro agrícola refere-se a precificação adequada do risco. Nos EUA, por exemplo, que possuem um dos mais antigos e desenvolvidos mercados de seguro agrícola do mundo a discussão acerca das melhorias e aprimoramentos metodológicos é realizada anualmente em eventos acadêmicos em que participam representantes do mercado (seguradoras e resseguradoras), governo (USDA) e pesquisadores nas áreas de economia, atuária, estatística e agronomia. Durante várias décadas os métodos atuariais aprimoraram-se conjugadas a evolução do próprio ramo de seguro.

Quando os riscos não são adequadamente precificados surgem problemas de desequilíbrio atuarial da carteira e seleção adversa, ou seja, quando há anti-seletividade por problemas de assimetria de informação. Na prática, quando a taxa de prêmio é fixada conforme o risco médio em uma carteira composta por produtores de alto e baixo risco, aqueles com risco maior terão uma probabilidade maior de demandar pelo seguro do que os produtores de baixo risco.

Assim, apenas aqueles de alto risco terão incentivos em demandar pelo seguro. Como consequência a concentração da carteira com indivíduos de alto risco aumenta consideravelmente a chance de tornar a carteira deficitária. Na safra seguinte, o mercado tende a elevar a taxa a fim de compensar as perdas. O problema torna-se ainda pior, e assim,

sucessivamente. Esse fenômeno é conhecido como “espiral da morte” na literatura acadêmica.

Percebe-se, dessa forma, que definir uma taxa única para todo o território nacional baseado em uma média global, para cada cultura agrícola, apenas fortalece o problema da seleção adversa, com severas implicações para o equilíbrio da carteira e para a continuidade do seguro agrícola. O histórico do PROAGRO revela tais dificuldades.

O procedimento correto nessas circunstâncias é a separação das taxas por tipo de unidade segurada (propriedade rural) e tipo de cultura, levando-se em conta suas peculiaridades edafoclimáticas, genética e de manejo. Porém, dadas as limitações de dados estatísticos que reflitam a estrutura de risco em nível de propriedade exaurindo a multiplicidade dos riscos envolvidos recorre-se ao uso de dados agregados em nível municipal como uma proxy dos dados individualizados.

O interessante nesse caso é que a metodologia utilizada com os dados agregados pode ser aplicada também aos dados individualizados por propriedade rural, assim que uma massa considerável de dados for acumulada ao longo dos anos. É importante destacar que o método ou metodologia aqui relatado refere-se a um conjunto de procedimentos sistematizados e organizados que baseiam o processo de obtenção do conhecimento científico. A metodologia apresentada está fundamentada na estatística, ou seja, a ciência que se baseia nas teorias probabilísticas para explicar fenômenos observacionais ou experimentais.

6.5 ASPECTOS ATUARIAIS DO SEGURO AGRÍCOLA

Na literatura especializada um dos primeiros trabalhos que consideraram a análise probabilística aplicada a precificação do risco agrícola foi realizado por Botts & Boles

(1958), que utilizaram de forma bastante simples a distribuição Normal para parametrizar e quantificar o risco agrícola para uso no sistema de seguro agrícola norte-americano.

Alguns anos mais tarde Day (1965) estudou dados de produtividade resultantes de experimentos com diferentes níveis de nitrogênio para algodão e milho. A hipótese era de que a produtividade tinha assimetria positiva, na medida em que pequenas variações climáticas durante as fases mais sensíveis da cultura poderiam reduzir a produtividade. A conclusão apontou que as distribuições apresentaram, em geral, assimetrias e curtoses não-Normais.

Outro estudo conduzido por Gallagher (1987) procurou evidências de assimetrias nas distribuições de probabilidade em dados de produtividade potencial em nível nacional para a soja nos Estados Unidos. Os resultados mostraram uma assimetria negativa moderada e aumento da variabilidade correspondente a um patamar 20% maior na metade da década de 80 do que em 1972.

Na mesma linha de raciocínio Moss and Shonkwiler (1993) construíram um modelo de tendência estocástica para as séries de milho, agregadas em nível nacional, juntamente com a possibilidade de se transformar os dados para reduzir o efeito de outliers e induzir normalidade. Os resultados sugeriram que a produtividade do milho não segue uma distribuição normal apresentando, particularmente, assimetria negativa.

Taylor (1990) elabora um estudo para as séries de produtividade de milho, soja e trigo corrigidos para a tendência, por meio da construção de dois modelos distintos: método da transformação normal e método da distribuição condicional. Observou-se nos resultados indicativos de assimetria e curtose em todas as culturas.

Ao longo dos anos diversos especialistas analisaram o tema sem uma conclusão definitiva. Diversas abordagens foram apresentadas pelos atuários e econometristas na busca de distribuição de probabilidade que melhor se ajustasse aos dados de produtividade agrícola.

Por exemplo, a modelagem paramétrica, semi-paramétrica (Ker and Coble, 2003), não-paramétrica (Goodwin and Ker, 1998; Turvey and Zhao, 1999, Ozaki et al., 2008b), a abordagem Bayes empírico não-paramétrico (Ker and Goodwin, 2000) e a abordagem bayesiana espaço temporal (Ozaki et al., 2008a).

Dentro da abordagem paramétrica, diversos autores concluem que a produtividade agrícola segue uma distribuição normal (Just and Weninger, 1999). Entretanto, outros pesquisadores como Day (1965), Taylor (1990), Ramirez (1997), Ramirez et. al. (2003), encontraram evidências contra a normalidade na distribuição dos dados.

A adequação do uso da distribuição Normal facilitaria os cálculos atuariais, mas não há consenso quanto ao seu uso. Outras sugestões incluem o uso da distribuição Beta (Nelson and Preckel, 1989), Transformações Seno Hiperbólico Inversa (Moss and Shonkwiler, 1993) e Gama (Gallagher, 1987). Sherrick et al. (2004) ajustaram e compararam diversas distribuições paramétricas, como por exemplo, a Normal, Lognormal, Beta, Weibull e Logística utilizando a dados de produtividades por propriedades rurais nos EUA. Goodwin and Mahul (2004), corretamente apontaram que a distribuição Lognormal impõe assimetria positiva à distribuição, característica que pode ou não estar presentes nos dados.

Além dos testes realizados com diferentes distribuições de probabilidade, diversos estudos apontaram o pequeno número de observações de produtividade, em nível agregado e menor ainda em nível individual, como um empecilho para a modelagem atuarial.

A escassez de uma série temporal relativamente longa impossibilita a detecção de variações estruturais na produtividade e, conseqüentemente, verificar a ocorrência de eventos catastróficos. Quedas bruscas nas series de produtividade causadas por eventos climáticos adversos revelam situações de alto risco relativo.

Apesar de alguns métodos terem sido sugeridos para tentar superar tal problema, tais como, combinar dados de produtividade individuais e regionais ou utilizar dados agregados como proxies para os dados em nível de fazenda, ainda assim, isto dificulta sobremaneira o trabalho de modelar a estrutura de risco do produtor de maneira acurada. Por essa razão deve-se escolher métodos atuariais que reduzam tais problemas, ao mesmo tempo, que os dados sejam coletados e organizados para formar uma base consistida e relativamente longa.

Outra questão relacionada ao nível de agregação dos dados é com relação a alteração na variabilidade das séries. Os métodos de agregação dos dados geralmente induzem a um alisamento da série de dados reduzindo a variabilidade originalmente presente. Em outras palavras, séries municipais de produtividade revelam um perfil de oscilação menor do que os dados em nível de propriedade rural levando a erros na estimação das taxas de prêmio.

A tendência é outra característica importante que deve ser levado em conta na modelagem dos dados de produtividade agrícola. Em muitos casos, quando existem amostras relativamente grandes, percebe-se que o processo gerador dos dados de produtividade não é constante ou estável, mas varia de acordo com o tempo.

A incorporação de novas tecnologias permitiu ao produtor elevar sua produtividade com o passar do tempo. Assim, a produtividade observada em meados de 1960, em geral é menor do que a produtividade observada em 2018, por exemplo. Para que tais produtividades sejam comparáveis é preciso que a componente de tendência seja captada na modelagem. Para isso ajusta-se um modelo polinomial de ordem 1, utilizando o tempo como covariável.

Ao ajustar um modelo de regressão linear polinomial de ordem 1, com apenas uma variável (tempo) da forma $y_t = \alpha + \beta T + e_t$, $e_t \sim N(\mu, \sigma^2)$, capta-se a tendência que as séries possivelmente possam apresentar, em que y_t é o vetor de produtividade, T é o vetor de tempo, α e β os parâmetros de regressão. Utilizando os resíduos \hat{e}_t da regressão, a estimativa da

última observação através do modelo ajustado \hat{y}_{2017} , a tendência é retirada de acordo com a equação: $\tilde{y}_t = \hat{y}_{2017} \left(1 + \frac{\hat{\epsilon}_t}{\hat{y}_t}\right)$ (Ozaki et al, 2008, Tejada et al., 2008, Duarte et al., 2017). A estatística t foi utilizada para definir a significância dos parâmetros.

Para testar a autocorrelação nos resíduos utiliza-se o teste de Ljung and Box (1978) para tratar a autocorrelação serial da produtividade. Ljung e Box propuseram uma pequena alteração no teste de Box - Pierce, na qual torna o teste mais generalizado. Foi observado na literatura que o teste de Ljung - Box apresenta melhores resultados do que o teste de Box - Pierce. Na prática, ao se captar a tendência em uma série de produtividade, assume-se que a série resultante torna-se estacionária.

Outro ponto a se destacar refere-se a possível presença de correlação espacial entre dados de municípios vizinhos. Como os dados são tratados individualmente, ou seja, os modelos são para séries temporais univariadas e não modelos para dados em painel, a presença de correlação espacial não afeta os resultados. Exceto, na situação abordada por Ozaki et al. (2008), quando a amostra por município era relativamente pequena e optou-se pela expansão da amostra utilizando dados de municípios vizinhos ao município central e sua posterior estimação via abordagem não-paramétrica por meio do estimador Kernel.

Embora essa abordagem seja apropriada para a estimação da probabilidade de perda e, posteriormente, o cálculo da taxa, ela apresenta algumas limitações. A estabilidade do estimador Kernel em pequenas amostras é um fator que merece atenção, além disso, a taxa de convergência para a verdadeira densidade é lenta, embora o estimador produza estimativas consistentes de qualquer densidade. Ademais, percebe-se a grande importância na escolha da distância a ser escolhida para selecionar os municípios vizinhos que normalmente é realizada de forma ad hoc e, portanto, dependente da escolha do usuário.

Para o teste da heterogeneidade de variâncias utiliza-se o teste de Breusch and Pagan (1979). Baseado no teste multiplicador de Lagrange, o teste de Breusch-Pagan é bastante utilizado para testar a hipótese nula de que as variâncias dos erros são iguais (homoscedasticidade) versus a hipótese alternativa de que as variâncias dos erros são uma função multiplicativa de uma ou mais variáveis, sendo que esta(s) variável(eis) pode(m) pertencer ou não ao modelo em questão.

Inicialmente, ajusta-se um modelo de regressão linear (simples ou múltiplo) e obtêm-se os resíduos $e = (e_1, \dots, e_n)$ e os valores ajustados $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$. Em seguida, consideram-se os resíduos ao quadrado e estes são padronizados de modo que a média do vetor de resíduos padronizados u seja igual a 1. Esta padronização é feita dividindo cada resíduo ao quadrado pela SQE/n em que SQE é a Soma de Quadrados dos Resíduos do modelo ajustado e n é o número de observações. Desta forma, temos que cada resíduo padronizado é dado por

$$u_i = \frac{e_i^2}{SQE/n}, i = 1, \dots, n, \text{ em que } SQE = \sum_{i=1}^n e_i^2.$$

Por fim, faz-se a regressão entre $u = (u_1, \dots, u_n)$ (variável resposta) e o vetor \hat{y} (variável explicativa) e obtêm-se a estatística do teste χ_{BP}^2 calculando-se a Soma de Quadrados da Regressão de u sobre \hat{y} e dividindo-se o valor encontrado por 2. Sob a hipótese nula, esta estatística tem distribuição qui-quadrada com 1 grau de liberdade. Para a modelagem da produtividade foram testados modelos paramétricos simétricos e assimétricos, que incorporam a unimodalidade e bimodalidade dos dados. Todas estas análises foram realizadas no software R utilizando os pacotes “Optim” e “AdequacyModel”.

Testa-se também se os dados seguem uma distribuição Normal, por meio do teste de Kolmogorov-Smirnov (KS). O teste de Kolmogorov - Smirnov pode ser utilizado para avaliar as hipóteses: H_0 : os dados seguem uma distribuição Normal, contra H_1 : os dados não seguem uma distribuição Normal.

‘Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada dos dados, no caso a Normal, e a função de distribuição empírica dos dados. Como critério, comparamos esta diferença com um valor crítico, para um dado nível de significância. O valor-p resultante do no programa R pode ser interpretado como a medida do grau de concordância entre os dados e a hipótese nula. Quanto menor for o valor-p, menor é a consistência entre os dados e a hipótese nula. Então, a regra de decisão é rejeita-se H_0 , se o valor-p for menor ou igual a um certo valor arbitrado pelo analista.

Em seguida é preciso escolher entre o uso de distribuições paramétricas ou não-paramétricas para ajustar aos dados. Recomenda-se as seguintes distribuições paramétricas: Normal Assimétrica, Skew-T e a distribuição Odd Log-logística, por apresentarem robustez e flexibilidade na detecção de assimetrias e bimodalidades. É importante destacar que ambas as abordagens apresentam desvantagens em amostras relativamente pequenas.

6.6 ESTUDOS ATUARIAIS APLICADOS AO BRASIL

No Brasil, alguns trabalhos abrangeram a problemática da precificação utilizando diversas abordagens, que variaram em função do tipo de produto de seguros, ou seja, se o seguro agrícola é de produtividade ou de faturamento. Um dos primeiros trabalhos teve como objetivo de sugerir formas de calcular as taxas de prêmio de um hipotético seguro de renda, no formato dos seguros americanos Income Protection ou Revenue Assurance, para milho

nos municípios do Paraná/BR, por meio do uso de cópulas na distribuição conjunta entre preços e produtividade (Miqueleto, 2011).

Outro trabalho inovador, proposto por Brisolara (2013), considerou a metodologia para o cálculo da taxa de prêmio utilizando séries univariadas de receita agrícola e séries multivariadas de preço e produtividade, supondo independência entre as duas variáveis. A modelagem foi realizada por meio da Normal Multivariada relacionando a correlação entre preço e produtividade na matriz de variância e covariância da própria distribuição. A aplicação teve como foco o seguro agrícola de receita para soja no estado do Paraná.

Santos, Scalon e Ozaki (2014) sugeriram a distribuição normal-assimétrica como uma alternativa à distribuição normal para modelar a distribuição da produtividade agrícola no Brasil. Os pesquisadores destacam que a distribuição normal-assimétrica foi escolhida por apresentar a possibilidade de ajuste de uma série que contenha assimetria em sua distribuição, característica comum nos dados de produtividade. Porém, destacam que existem outras distribuições como a log-normal e beta que seriam possíveis também em dadas condições.

Um ponto a destacar refere-se ao fato de que as séries de produtividade utilizadas no trabalho apresentaram tendência, o que exigiu métodos de correção das mesmas. Sabe-se que as séries corrigidas podem apresentar valores negativos, fato que impossibilita o uso de algumas distribuições paramétricas, como é o caso da densidade log-normal.

Uma desvantagem da distribuição normal-assimétrica em relação à normal é a possibilidade de não ocorrer convergência no algoritmo que estima os parâmetros do modelo. Os pesquisadores ressaltaram que ambos os modelos apresentam uma limitação que é a impossibilidade de bimodalidade ou multimodalidade da distribuição de produtividade.

Os autores observaram também que a distribuição normal-assimétrica apresentou alguma assimetria positiva em relação à normal e, portanto, sugeriram que a normal estaria

subestimando o pagamento esperado em alguns municípios. O prêmio levou em conta o valor do pagamento esperado, como consequência, pode-se subestimá-lo e cobrar menos do que o necessário para compensar as perdas e ainda, em caso de sinistro, a seguradora deverá fazer um pagamento maior do que o previsto.

Duarte (2015), em sua tese de doutorado, apresentou métodos alternativos para precificação dos riscos de produtividade e faturamento agrícola. Na modelagem da distribuição de produtividade, utilizou-se distribuições paramétricas que capturam a simetria, a assimetria e a bimodalidade dos dados, características estas geralmente encontradas em produtividades brasileiras. Além disso, a metodologia de cópulas foi utilizada na análise multivariada entre produtividade e preço.

Esta metodologia apresenta um ganho significativo para estudos de problemas multivariados comparados à distribuição normal multivariada, pois pode-se levar em consideração estruturas de dependências linear, não linear e dependência somente nas caudas da distribuição multivariada. Além disso especifica qualquer tipo de distribuição para as distribuições marginais.

Polo (2016), em sua dissertação de mestrado, teve por objetivo aplicar redes Bayesianas (grafo acíclico direcionado, ou modelo hierárquico Bayesiano) para a estimação da distribuição de probabilidade de rendimento da soja em alguns municípios do Paraná, com foco na análise comparativa de riscos.

O pesquisador descreve que a modelagem baseada em redes bayesianas apresenta vantagens de flexibilidade de interpretação da estrutura do modelo. Em especial, a modulação do modelo a partir de um grafo da rede bayesiana permite dar uma interpretação física e agrônômica a um modelo empírico. Se a abordagem de ajuste é aceitável, ganha-se grande simplicidade no ajuste de um modelo complexo.

Por fim o autor relata que o modelo ainda permite a sua aplicação em momentos diferentes do ciclo de produção agrícola para obter informações diferentes (úteis para cada momento), e associar dados meteorológicos aos dados de sensoriamento remoto sem que haja necessidade de observações de todas as variáveis estarem disponíveis.

Adami e Ozaki (2016) tiveram como objetivo de analisar a viabilidade do seguro receita, considerando-se a receita bruta da soja da região oeste do Paraná, em nível de propriedade rural. Os autores relatam que tradicionalmente, o método usado pelo mercado para a precificação do risco, tanto de custeio quanto de receita, é baseado na relação entre as perdas esperadas e o valor garantido pela empresa seguradora, conhecido como "taxas empíricas" (ER). Uma das principais desvantagens do método ER é a dependência de um grande número de informações para refletir acuradamente a distribuição de probabilidade da produtividade ou da receita.

De modo a resolver esse problema, os autores assumiram que a receita da propriedade agrícola segue uma distribuição normal assimétrica, que incorpora um parâmetro flexível para captar, caso exista, a assimetria da distribuição de probabilidade. Optaram pela distribuição normal assimétrica em função de sua flexibilidade e devido a evidências de assimetria na distribuição da produtividade agrícola e da receita agrícola em estudos anteriores (SHERRICK et al., 2004; OZAKI e SILVA, 2009; SANTOS, SCALON e OZAKI, 2014).

6.7 A DISTRIBUICAO NORMAL ASSIMÉTRICA

Considerou-se a distribuição Normal Assimétrica (DNA) pela sua flexibilidade em acomodar assimetrias presentes nas distribuições dos dados. A DNA foi proposta por

Azzalini (1985) e é uma generalização da distribuição Normal. A vantagem está no fato de se captar assimetrias tanto negativas quanto positivas. Tais assimetrias tem impacto direto no calculo da probabilidade de perda e, conseqüentemente, no cálculo da taxa de prêmio.

Nesse contexto, considere f a função densidade de probabilidade (f.d.p.) simétrica ao redor de zero e G função distribuição cumulativa (f.d.c.) contínua e absoluta, tal que $g = G'$ é simétrica ao redor de zero (Ozaki e Silva, 2009). Então,

$$2 f(y) G(\gamma y), y \in R \quad (1)$$

É uma f.d.p. para qualquer $\gamma \times \in R$. Da equação (1), a DNA, com parâmetro de locação μ , escala σ e forma γ , é definida como:

$$\phi(y|\mu, \sigma, \gamma) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\frac{\gamma(y-\mu)}{\sigma}\right) \quad (2)$$

Onde ϕ e Φ são respectivamente a f.d.p. e a f.d.c. de uma variável aleatória Normal padrão. O grau de assimetria da DNA é dada pela equação

$$\varepsilon = \delta^3 \left[\frac{4}{\pi} - 1 \right] \left[1 - \frac{2\delta^2}{\pi} \right]^{-3/2} \sqrt{2/\pi} \quad (3)$$

onde $\delta = \gamma(1 + \gamma^2)^{-1/2}$ e $-0.99527 < \varepsilon < 0.99527$ com casos limites ($\gamma \rightarrow \pm\infty$) sendo uma distribuição Meio Normal. De fato, utilizou-se a abordagem de Henze (1986), representando a DNA como uma mistura de média variância de uma Normal e uma Meia Normal.

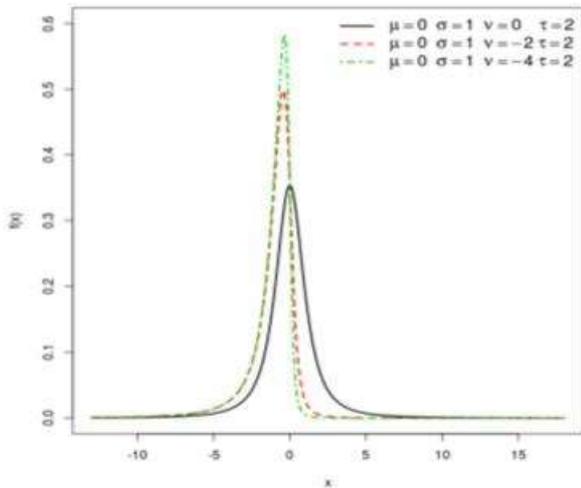
6.8 A DISTRIBUICAO SKEW-T

Quando se tem dados com um comportamento assimétrico e presença de valores extremos ou discrepantes, a distribuição Skew-Normal pode não se ajustar bem. Em tais situações, uma alternativa mais robusta ao modelo Skew-Normal é proposta por Azzalini e Capitanio (2003), a versão assimétrica da distribuição t-Student (ST), denominada t-Student Assimétrica.

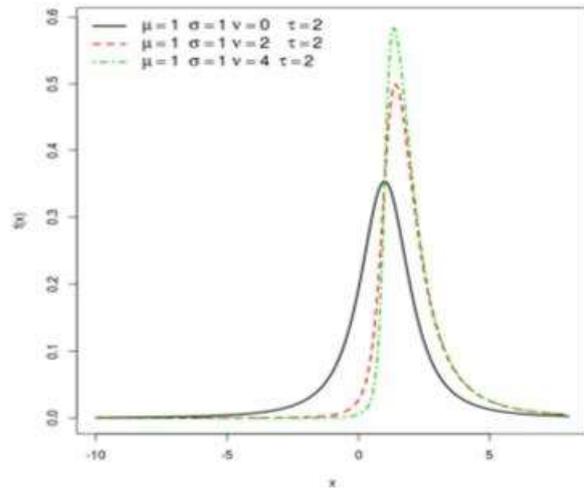
Uma variável X é chamada “Skew-t Student”, com parâmetros de locação μ , $\mu \in \mathbb{R}$, escala σ , $\sigma \in \mathbb{R}_+$, assimetria ν , $\nu \in \mathbb{R}$, e curtose τ , $\tau \in \mathbb{R}_+$, se sua f.d.p é dada por:

$$f(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} t\left(\frac{x-\mu}{\sigma}\right) T\left[\nu\sqrt{\lambda}\left(\frac{x-\mu}{\sigma}\right)\right], \quad (4)$$

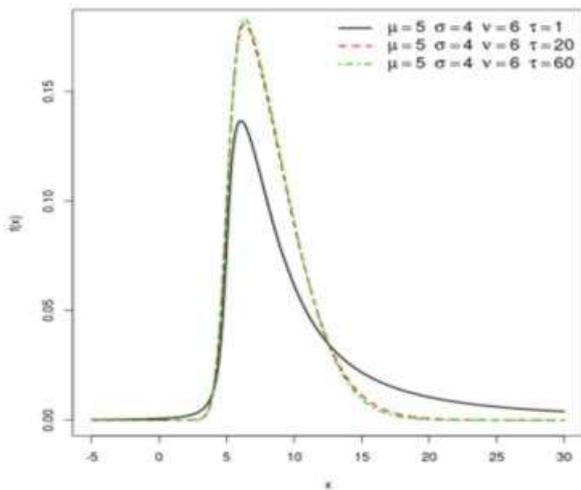
em que $z = \left(\frac{x-\mu}{\sigma}\right)$, $w = \nu\lambda^{1/2}z$, $\lambda = \frac{\tau+1}{\tau+z^2}$, $tZ1$ e $TZ1$ são respectivamente, a f.d.p da t-Student padrão com τ graus de liberdade e a fda da t-Student padrão com $\tau + 1$ graus de liberdade. Como notação, adota-se $Z \sim ST(\mu, \sigma, \nu, \tau)$. Note que, quando $\nu = 0$, a distribuição é equivalente a t-Student simétrica.



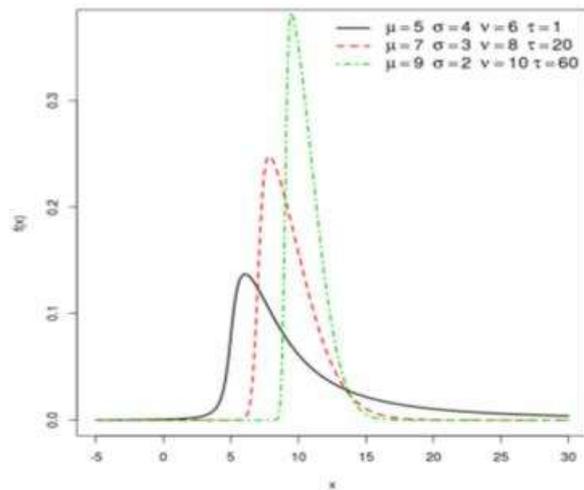
(a) Variando $\nu < 0$



(b) Variando $\nu > 0$



(c) Variando $\tau > 0$



(d) Variando μ e τ

Figura 1 : Gráfico da f.d.p da Skew-t para diferentes valores dos parâmetros.

A figura 1 ilustra o comportamento da f.d.p da ST. Nota-se nas figuras 1(a,b) a contribuição do parâmetro ν na forma da distribuição, com μ , σ , τ fixos. Os valores positivos de ν indicam assimetria positiva e os valores negativos de ν indicam a assimetria negativa.

Na figura 1(c), observa-se o efeito do parâmetro de curtose na distribuição quando os outros parâmetros são fixos. Na figura 1(d) observa-se o efeito de deslocamento da

distribuição pela variação do parâmetro μ e caudas mais “pesadas” com a variação do parâmetro τ .

3.5 A DISTRIBUICAO ODD LOG-LOGÍSTICA

Uma nova classe de família de distribuições contínuas recentemente introduzidas por Cruz (2016) é chamada odd log-logistic-F (OLLF). A inferência e estimação pelo método da máxima verossimilhança para grandes amostras são também abordados para dados de sobrevivência censurados. Esta nova classe de distribuições é interessante, pois através de suas simulações pode-se capturar a bimodalidade dos dados.

Dado uma função base de probabilidade acumulada qualquer com $F(x; \xi)$ um vetor de parâmetros ξ , a função distribuição cumulativa (f.d.c.) da distribuição odd log-logistic-F (OLL-F) com um parâmetro de forma $\alpha > 0$ é definido por:

$$G(x; \xi) = \int_0^{\frac{F(x;\xi)}{\bar{F}(x;\xi)}} \frac{\alpha x^{\alpha-1}}{(1+x^\alpha)^2} dx = \frac{F(x;\xi)^\alpha}{F(x;\xi)^\alpha + \bar{F}(x;\xi)^\alpha} \quad (5)$$

em que $\bar{F}(x, \xi) = 1 - F(x, \xi)$ e $\alpha = \frac{\log \frac{G(x,\xi)}{G(x,\xi)}}{\log \frac{F(x,\xi)}{\bar{F}(x,\xi)}}$.

O parâmetro α representa o quociente do log da razão de chances para a distribuição base. Se $F(x, \xi) = \frac{x}{1+x}$, a equação (5) então se reduz a função de probabilidade da distribuição log-logística. Várias distribuições podem ser geradas da equação acima, como os apresentados nos trabalhos de Alizadeh et al. (2015), Cruz et al. (2015) e Braga et al. (2016).

A nova família de distribuição OLLF permite uma maior flexibilidade nas caudas da distribuição. Se considerarmos a função base $G(x; \xi)$ como a distribuição acumulada da normal $N(\mu, \sigma)$ Assim, para a nova classe de família odd log logistic normal (OLL-N) a equação (5) se reduz a:

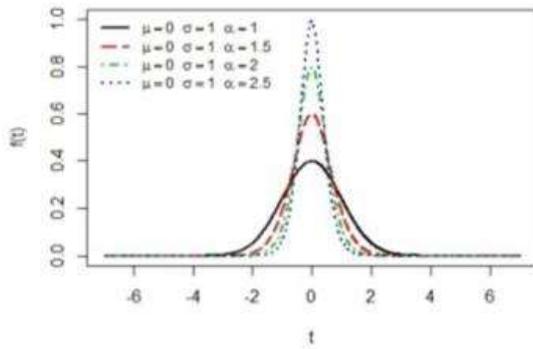
$$G(x; \mu, \sigma, \alpha) = \int_0^{\frac{\Phi(x; \mu, \sigma, \alpha)}{\bar{\Phi}(x; \mu, \sigma, \alpha)}} \frac{\alpha x^{\alpha-1}}{(1+x^\alpha)^2} dx = \frac{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right)}{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha} \quad (6)$$

em que $\bar{\Phi}(x; \mu, \sigma) = 1 - \Phi(x; \mu, \sigma)$.

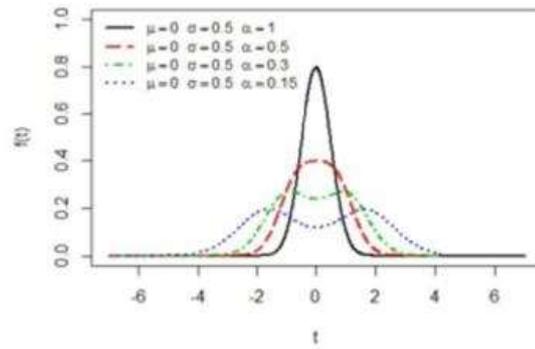
A função densidade de probabilidade (f.d.p) é dada por:

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right) \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}}{\sigma \left\{ \Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha \right\}^2} \quad (7)$$

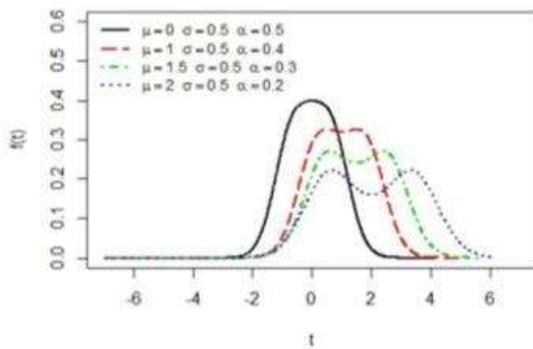
Note que $\alpha > 0$ é um parâmetro de forma. A partir de agora, assumimos que uma variável aleatória com função de densidade como da equação (7) é denotado por $X \sim \text{OLLN}(\alpha, \mu, \sigma)$.



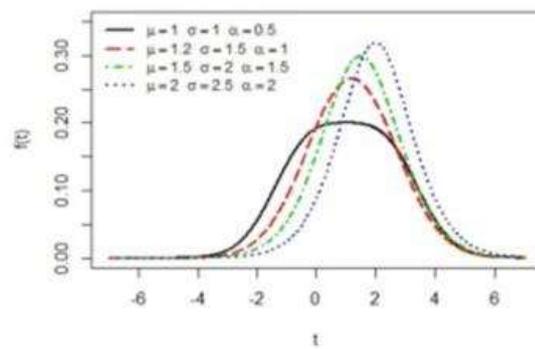
(a) Variando $\alpha > 1$



(b) Variando $\alpha < 1$



(c) Variando $\mu > 0$ e $0.2 < \alpha < 0.5$



(d) Variandos todos os parâmetros

Figura 2: Gráfico da f.d.p da OLLN para diferentes valores dos parâmetros.

A figura 2 apresenta algumas formas que a função densidade do modelo OLLN pode assumir para diferentes valores dos parâmetros μ , σ e α . Em simulações, pode-se observar a contribuição do parâmetro α sobre a unimodalidade e bimodalidade da distribuição, quando μ , σ são fixos, figura 2(a,b).

Quando o parâmetro α aproxima-se de zero, a f.d.p apresenta bimodalidade. Por outro lado, quando o valor de α aumenta, a função apresenta unimodalidade. Além disso, quando μ varia, os gráficos são transladados no eixo das abscissas, independentemente da forma como

mostra a Figura 1c. A figura 2d apresenta a f.d.p da OLLN com variação de todos os parâmetros.

6.9 A DISTRIBUICAO NÃO-PARAMÉTRICA

Diferentemente da abordagem paramétrica, a análise não-paramétrica não especifica à priori a distribuição dos dados. A sua vantagem é deixar que os dados revelem a forma da densidade. Ademais, certas características da distribuição podem ser captadas, tais como, assimetrias positivas e negativas e a bi ou multimodalidade.

Dentre os diversos tipos de estimadores da densidade, o histograma é o mais utilizado e conhecido nas análises exploratórias de dados. Dois aspectos merecem destaque na estimação dos histogramas, que são a origem e a largura das classes h ou “janelas”.

Em função da escolha da origem, diferentes histogramas podem ser construídos e, conseqüentemente, diferentes interpretações podem surgir. O outro aspecto, a largura das classes determina o alisamento da série. Apesar de sua simplicidade, este estimador apresenta desvantagens em relação a outros estimadores, como por exemplo, o estimador Kernel. Recomenda-se trabalhar com histogramas para realizar uma primeira análise dos dados.

No contexto de precificação de contratos de seguro agrícola, Goodwin e Ker (1998), Turvey e Zhao (1999) e Ozaki et al. (2008b) utilizaram o estimador Kernel para determinar a forma da densidade da produtividade condicional. O estimador kernel da densidade $\hat{f}(y)$ pode ser representada como uma convolução da distribuição amostral, utilizando-se uma função kernel K , tal que:

$$\hat{f}(y) = \int K_h(y-v) dF_n(v) \quad (8)$$

onde $K_h(v) = 1/hK(v/h)$ e $F_n(v)$ é a função de distribuição amostral (Goodwin e Ker, 2002).

O estimador kernel é a soma de “saltos” localizados em cada observação, de modo que a função kernel determina a forma destes saltos e a janela sua largura. Quanto maior o valor da janela, a série terá um alisamento maior e os detalhes tendem a desaparecer. Por outro lado, os saltos terão uma forma de pico quanto mais próxima de zero estiver a janela, tornando mais pronunciado os detalhes na densidade.

Algumas suposições devem ser feitas à respeito de K . A função kernel será não negativa, simétrica e constituirá uma função de densidade, tal que $\int K(v)dv = 1$. Além disso, $\int vK(v)dv = 0$ e $\int v^2K(v)dv = \theta_2 \neq 0$. Caso a função kernel seja representada pela densidade Normal, então θ_2 será a variância da distribuição.

Como uma medida de discrepância do estimador da densidade \hat{f} em relação a verdadeira densidade f , Silverman (1986) adota o erro quadrático integrado médio (MISE) dado por $\int E[\hat{f}(y) - f(y)]^2 dx$, que pode ser decomposta em dois componentes, o viés quadrático integrado e a variância integrada. Sob certas pressuposições, o viés pode ser aproximado por $\frac{1}{4}h^4\theta_2 \int f''(y)^2 dy$ e a variância por $(nh)^{-1} \int k(v)^2 dv$.

Ao se minimizar o MISE escolhe-se um alto valor para o parâmetro de alisamento, reduz-se a variação aleatória (variância) levando ao aumento do erro sistemático (viés). Caso

contrário, um baixo valor do parâmetro resulta em aumento da variância integrada e diminuição do viés quadrático integrado.

Na busca pelo parâmetro de alisamento ótimo e o kernel que minimize o MISE, o autor mostra que, se o h for ótimo, $h_{opt} = \theta_2^{-2/5} (\int K(v)^2 dv)^{1/5} (\int f''(y)^2 dy)^{-1/5} n^{-1/5}$, então o valor aproximado para o erro quadrático integrado médio será dado por $\frac{5}{4} W(K) (\int f''(y)^2 dy)^{1/5} n^{-4/5}$, onde $W(K) = \theta_2^{2/5} (\int K(v)^2 dv)^{4/5}$.

O problema de minimizar o MISE torna-se um problema de escolher o mínimo $W(K)$, dado o parâmetro de alisamento. Seguindo certas restrições, se a função kernel escolhida for o kernel de *Epanechnikov* $W(K_e)$, então o MISE será minimizado.

Para comparar outros kernel simétricos, procedeu-se na elaboração de um índice de eficiência igual a $[W(K_e) / W(K)]^{5/4}$. Utilizando as densidades biweight, triangular, gaussiana e retangular, observou-se que todos os kernel apresentaram eficiência próximas de um. Desta forma, a escolha do kernel resulta em pouca influência sobre o critério escolhido.

A escolha do parâmetro de alisamento ótimo, no caso da distribuição Gaussiana, será igual a $1.06\sigma n^{-1/5}$, onde σ é o desvio padrão da produtividade. Recomenda-se $\sigma = \min(\text{desvio padrão}, \text{interquartil}/1.34)$ para levar em conta desvios da normalidade. Além disso, se o fator 1,06 for reduzido para 0,9, obtêm-se melhores resultados empíricos. Desta forma, recomenda-se o uso desta estimativa para determinação do parâmetro de alisamento.

6.10 CRITÉRIOS PARA ESCOLHA DE DISTRIBUIÇÕES

Para a seleção da distribuição de probabilidade a ser utilizada na precificação de risco dois critérios são utilizados na literatura: o critério de informação Akaike (AIC), proposto por Akaike (1974) e o critério Bayesiano (BIC), proposto por Schwarz (1978).

Estes critérios são baseados no logaritmo da função de verossimilhança estimada. Assim, o critério proposto por Akaike e o critério Bayesiano são, respectivamente:

$$AIC = -2l(\hat{\theta}) + 2k,$$

$$BIC = -2l(\hat{\theta}) + \ln(n) k$$

em que $l(\hat{\theta})$ é o logaritmo da função de verossimilhança maximizada sob o modelo considerado, k o número de parâmetros e $\hat{\theta}$ o vetor de parâmetros estimado. O modelo com o menor valor utilizando um dos critérios deve ser escolhido.

1. CÁLCULO DA TAXA DE PRÊMIO

Uma vez escolhida a abordagem probabilística por meio do critério de informação de Akaike (AIC) ou critério de informação bayesiano (BIC), a ser descrita posteriormente, gera-se a estimativa da probabilidade de perda que é um dos parâmetros da taxa de prêmio e que pode ser calculada através da área sob a curva menor do que a produtividade garantida. Desta forma, se denominarmos o nível de cobertura por λ , tal que $0 < \lambda < 1$, e a produtividade esperada por y_e , então a probabilidade da produtividade ser menor do que λy_e será calculada através da área sob a densidade, numericamente estimada através da regra do trapezóide. A taxa de prêmio será dada por (Goodwin e Ker, 1998):

$$\text{Taxa de prêmio} = \frac{F_Y(\lambda y^e) E_Y[\lambda y^e - (Y | y < \lambda y^e)]}{\lambda y^e} \quad (9)$$

onde E é o operador de esperança e F a distribuição cumulativa da produtividade.

6.11 PRECIFICAÇÃO DO SEGURO DE FATURAMENTO: MODELANDO A PRODUTIVIDADE E O PREÇO CONJUNTAMENTE

Nota-se que, no caso anterior, a taxa de prêmio é calculada de forma univariada considerando apenas a produtividade agrícola. Outro produto que apresenta rápido crescimento e aceitação por parte dos produtores é o seguro agrícola de faturamento ou receita. Esse tipo de seguro é uma reivindicação relativamente antiga dos produtores rurais, mas atendida apenas em meados de 2010, quando o mercado ofertou o seguro de forma limitada em apenas 8 operações e consumindo apenas R\$ 63 mil de subvenção.

De 2010 até 2014 houve um aumento significativo do seguro de faturamento. Em 2014 foram aproximadamente 3.200 apólices e pouco mais de R\$ 35 milhões gastos em subvenção, para uma importância segurada de quase R\$ 1 bilhão. Em 2015, devido ao forte contingenciamento orçamentário sofrido pelo MAPA, o seguro de faturamento recuou drasticamente em razão de sua elevada dependência com o orçamento do PSR.

A partir de 2017, o MAPA priorizou o seguro de faturamento atrelando um maior percentual de subvenção nas operações do seguro de faturamento. Como resultado, em 2018, o seguro de faturamento alcançou quase 2.900 apólices, R\$ 34,4 milhões de subvenção gastos e uma importância segurada de quase R\$ 2 bilhões, praticamente o dobro de 2014 com praticamente o mesmo total gasto em subvenção. Atualmente, o seguro de faturamento

consome aproximadamente 10% dos recursos do PSR, equivalente a 16% de toda a importância segurada em 2018 e 4% das apólices realizadas, segundo o Atlas do Seguro Rural do MAPA.

No que tange a sua precificação, o seguro de faturamento é formado por duas ou três variáveis conforme as características do produto. A primeira delas é a produtividade e a segunda é o preço futuro, em reais. Caso o preço seja fixado em dólar, existe ainda uma terceira variável que é a taxa de câmbio.

Considerando o caso em que o preço seja fixado em reais, a taxa de prêmio será fixada levando-se em conta a relação entre as duas variáveis, de acordo com a seguinte equação:

$$\text{Taxa de Prêmio} = \frac{F_y(X < x; Y < y | x \cdot y < x_g \cdot y_g) [x_g \cdot y_g - E(XY | x \cdot y < x_g \cdot y_g)]}{x_g \cdot y_g} \quad (1)$$

onde E é o operador de esperança, F a distribuição cumulativa da produtividade, x_g representa o preço garantido, que em geral baseia-se em uma proporção dos preços futuros negociados em bolsas de mercadorias e futuros; y_g diz respeito à produtividade garantida, obtida através de uma proporção da produtividade média histórica do produtor; x representa o preço ao final da safra do produtor enquanto que y representa a produtividade obtida ao final da safra. Esse seguro paga, de acordo com $I = \max[(x_g \cdot y_g - x \cdot y); 0]$ caso haja perdas inesperadas na receita do produtor rural.

Nesse caso, a modelagem torna-se um pouco mais complexa. Em virtude da inclusão de mais uma variável aleatória, há necessidade de obter uma relação matemática entre distribuições multivariadas e suas distribuições marginais univariadas.

6.12 CÓPULAS

A dependência entre variáveis aleatórias X_1, X_2, \dots, X_n , é completamente descrita pela sua distribuição conjunta $H(X_1, X_2, \dots, X_n)$. Na maioria das situações aplicadas, apenas os marginais são conhecidas (estimadas ou fixadas a priori) e a distribuição conjunta pode ser desconhecida ou de difícil estimação. Na prática, a distribuição multivariada normal é muito utilizada em problemas multivariados. No entanto, essa suposição de normalidade restringe a tipo de associação entre as marginais, tornando-a linear e simétrica.

Por isso, a teoria de cópulas é uma ferramenta de modelagem multivariada muito utilizada em diferentes ramos da ciência, pois é possível identificar a estrutura de dependência linear, não linear, dependência em uma ou ambas as caudas da distribuição. A ideia de separar H em uma parte que descreve a estrutura de dependência e as partes que descrevem apenas o comportamento marginal levou ao conceito de cópulas, introduzido na literatura estatística por Sklar (Sklar, 1959).

De acordo com Joe (2014), a cópula é uma distribuição multivariada em que todas as distribuições unidimensionais marginais são uniformes $U(0, 1)$. Consequentemente, se C é uma cópula, então a distribuição de um vetor de variáveis aleatórias dependentes é $U(0, 1)$.

Definição de cópulas: Seja X_1, X_2, \dots, X_n variáveis aleatórias contínuas com função de distribuição conjunta $H(X_1, X_2, \dots, X_n)$ e distribuições marginais $F(X_1), F(X_2), \dots, F(X_n)$, respectivamente. Para todo $(x_1, \dots, x_n) \in [-\infty, \infty]^n$ considere o ponto em $[0, 1]^{n+1}$ com coordenadas $(F_{x_1}(x_1), \dots, F_{x_n}(x_n)), H(x_1, x_2, \dots, x_n)$. Este mapeamento de $[0, 1]^n$ a $[0, 1]$ é uma cópula n-dimensional.

A definição de cópula é conectada pelo teorema de Sklar (Sklar, 1959) que é o principal resultado na teoria da cópula.

Teorema de Sklar. Para uma distribuição d-dimensional $F \in \mathcal{F}(F_1, F_2, \dots, F_d)$ em que F_j é a j-ésima distribuição univariada marginal, a cópula associada a F é uma função de distribuição $C : [0, 1]^d \rightarrow [0, 1]$, $F(y) = C(F_1(y_1), F_2(y_2), \dots, F_d(y_d))$, $y \in \mathbb{R}^d$, com marginais $U(0, 1)$ que satisfazem:

1. Se F é uma distribuição d-dimensional contínua com marginais uniformes F_1, \dots, F_d e funções quantílicas F_{11}, \dots, F_{d1} , então: $C(u) = F_{11}(u_1), \dots, F_{d1}(u_d)$, $u \in [0, 1]^d$ é única; e,
2. Se F é uma distribuição d-dimensional de variáveis aleatórias discretas, então a cópula é única somente no conjunto $\text{Dom}(F_1) \times \dots \times \text{Dom}(F_d)$.

A cópula C é parametrizada por um vetor de parâmetros δ , chamado de parâmetro de dependência multivariado. As famílias de cópulas são geralmente dadas como função de distribuição acumulada (para a inferência via verossimilhança) e a cópula densidade é obtida

pela diferenciação. Se $C(u)$ é uma distribuição acumulada absolutamente contínua então sua

função densidade é dada por: $c(u) = c(u_1, \dots, u_d) = \frac{\partial^d C(u)}{\partial u_1 \dots \partial u_d}$, $u \in [0, 1]^d$

6.13 INFERÊNCIA DAS CÓPULAS

Suponha uma amostra independente e identicamente distribuída y_{i1}, \dots, y_{id} , $i = 1, \dots, n$, então a distribuição conjunta é $F(y; \eta_1, \dots, \eta_d, \delta) = C(F_1(y, \eta_1), \dots, F_d(y, \eta_d), \delta)$. Se todas as variáveis forem contínuas e F_1, \dots, F_d, C forem absolutamente contínuas com relação as densidades f_1, \dots, f_d, c , então a log-verossimilhança é dada por:

$$l(\theta) = \sum_{i=1}^n [\log c(F_1(y_{i1}; \eta_1), \dots, F_d(y_{id}; \eta_d); \delta) + \sum_{j=1}^d \log f_j(y_{ij}; \eta_j)], \quad (11)$$

em que $\theta = (\eta_1, \dots, \eta_d, \delta)^t$ é o vetor de todos os parâmetros das marginais e da cópula. Assim, o estimador de máxima verossimilhança (EMV) é obtido maximizando a expressão acima, $\hat{\theta}_{EMV} = \max_{\theta} l(\theta)$.

Sob condições usuais de regularidade, o estimador de máxima verossimilhança existe e é consistente e assintoticamente eficiente. Além disso, verifica-se a propriedade de assintoticamente normal: $\sqrt{n}(\hat{\theta}_{EMV} - \theta_0) \rightarrow N(0, \mathfrak{F}^{-1}(\theta_0))$, em que \mathfrak{F} é a matriz de informação de Fisher e θ_0 o valor verdadeiro usual.

Neste método é necessário estimar conjuntamente os parâmetros das distribuições marginais e os parâmetros da cópula, por isso, o método de máxima verossimilhança pode ser

computacionalmente dispendioso. Assim, a literatura sugere a estimação dos parâmetros em dois passos (Joe e Xu, 1996):

No primeiro passo, estima-se os parâmetros das marginais $\theta_1 = (\eta_1, \dots, \eta_d)$ ao realizar as estimativas das marginais $\hat{\theta}_1 = \max \sum_{i=1}^n \sum_{j=1}^d \log f_j(y_{i,j}; \eta_j)$.

No segundo passo, dado $\hat{\theta}_1$ estima-se os parâmetros da cópula θ_2 :

$$\hat{\theta}_2 = \max \sum_{i=1}^n [\log c(F_1(y_{i1}), \dots, F_d(y_{id}); \theta_2; \hat{\theta}_1)] \quad (12)$$

Este método é chamado de inferência por marginais (IFM). O estimador IFM é definido como o vetor $\hat{\theta}_{IFM} = (\hat{\theta}_1, \hat{\theta}_2)$. Joe e Xu (1996) mostra que o método IFM é altamente eficiente quando comparado com o método de MV.

6.14 MEDIDAS DE AJUSTE DOS MODELOS DE CÓPULAS

Para a seleção do modelo de cópulas os critérios mais utilizados são o critério de informação Akaike (AIC), proposto por Akaike (1974) e o critério Bayesiano (BIC), proposto por Schwarz (1978), conforme descrito na seção 3.6.

APLICAÇÃO DA METODOLOGIA PARA O CASO DA SOJA NO ESTADO DO PARANÁ: COMPARATIVO COM O PROAGRO

O problema da precificação mal especificada é conhecido no mercado segurador. Notoriamente, a causa está relacionada não apenas a falta de dados estatísticos adequados aos cálculos das taxas de prêmio, mas também ao desconhecimento de metodologias que acomodem satisfatoriamente as peculiaridades do setor agrícola.

A implicação direta deste fato é a definição de taxas que não refletem adequadamente o risco de cada unidade segurada, sejam eles, estados, municípios, propriedades agrícolas ou mesmo talhões dentro de uma propriedade. Em países em que o seguro agrícola está altamente desenvolvido é comum existirem produtos que cobrem talhões ao invés de propriedades. Nesse caso, a taxa deve ser fixada em nível de talhão ao invés de propriedade.

No Brasil, até 2016, o PROAGRO definia uma alíquota única para todas as culturas agrícolas em todo o território nacional repetindo o modelo de tarifação da CNSA, que juntamente com outros fatores, levaram ao fracasso da companhia e a sua dissolução em 1966.

A partir de 2017, houve uma mudança na tarifação do PROAGRO. As alíquotas passaram a se diferenciar por culturas, mas ainda assim, permaneceram as mesmas para todo o território nacional. O Manual do Crédito Rural define as seguintes alíquotas em empreendimentos em regime de sequeiro:

I - milho (verão) e soja: 4,5%;

II - milho safrinha (2ª safra): 6%;

III - ameixa, maçã, nectarina e pêsego, sem estrutura de proteção contra granizo: 6,5%;

IV - ameixa, maçã, nectarina e pêsego, com estrutura de proteção contra granizo: 3,5%;

V - cevada e trigo: 6,5%; e,

VI - demais culturas zoneadas: 4%.

Apesar da diferenciação por tipo de cultura, o problema da má especificação ainda persiste. O fato é que o PROAGRO não distingue entre os produtores de soja, por exemplo, aqueles de risco maior ou risco menor. A consequência desse problema é o aumento da antiseletividade, ou seja, a concentração de indivíduos de maior risco na carteira, o que pode levar a elevados prejuízos e déficit nos resultados da carteira.

Por definição, indivíduos de maior risco devem possuir taxas maiores, e vice-versa. Quando o PROAGRO define uma única taxa (alíquota) para todos os produtores de soja, dois graves problemas surgem: i) perda de receita, pois os produtores de alto risco deveriam pagar uma taxa bem maior do que aquela fixada no PROAGRO; e, ii) os produtores de baixo risco serão penalizados ao pagar uma taxa maior do que deveriam arcar, comprometendo sua margem operacional. Como a contratação do PROAGRO ou do seguro rural privado é compulsório para produtores que contratam crédito de custeio até R\$ 300 mil, a escolha muitas vezes é definida pelo preço. Possivelmente os produtores optarão pelo seguro privado, em função da subvenção ao prêmio.

Para que o exercício de cálculo das taxas de prêmio tenha efetividade e se aproxime à realidade do mercado usamos as mesmas fontes de dados das companhias credenciadas no PSR, ou seja, os dados de produtividade municipais do IBGE. A base de dados do IBGE é reconhecida pelo mercado segurador e, principalmente, pelos resseguradores como a principal fonte de dados para a precificação do risco agrícola, por pertencer a uma instituição governamental idônea no processo. Por esse motivo, a taxa de prêmio foi calculada utilizando os dados da produtividade média municipal do IBGE/SIDRA que pode ser acessada no link <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>

Com raras exceções, a taxa de prêmio é calculada com base nos dados das propriedades rurais pertencentes a determinadas cooperativas, principalmente do estado do Paraná. Nesses casos, é preciso um trabalho minucioso feito pelas seguradoras e aceito pelos resseguradores, no sentido de mostrar a confiabilidade na fonte supracitada.

Nesse contexto, para fins de comparação, calculamos as taxas de prêmio por municípios do estado do Paraná, para a cultura da soja, conforme metodologia descrita neste produto. Ao todo, 332 municípios foram selecionados para compor a amostra. Os dados de produtividade municipal utilizados correspondem ao período de 1974 a 2017 e são provenientes do IBGE no link <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>.

Nesse exercício é importante verificar o impacto quando se define as taxas por municípios ao invés de uma única taxa para todo o país. Observa-se uma grande variabilidade de taxas entre os municípios e que são completamente distintos da linha vermelha, que representa a taxa do PROAGRO (Fig. 3).

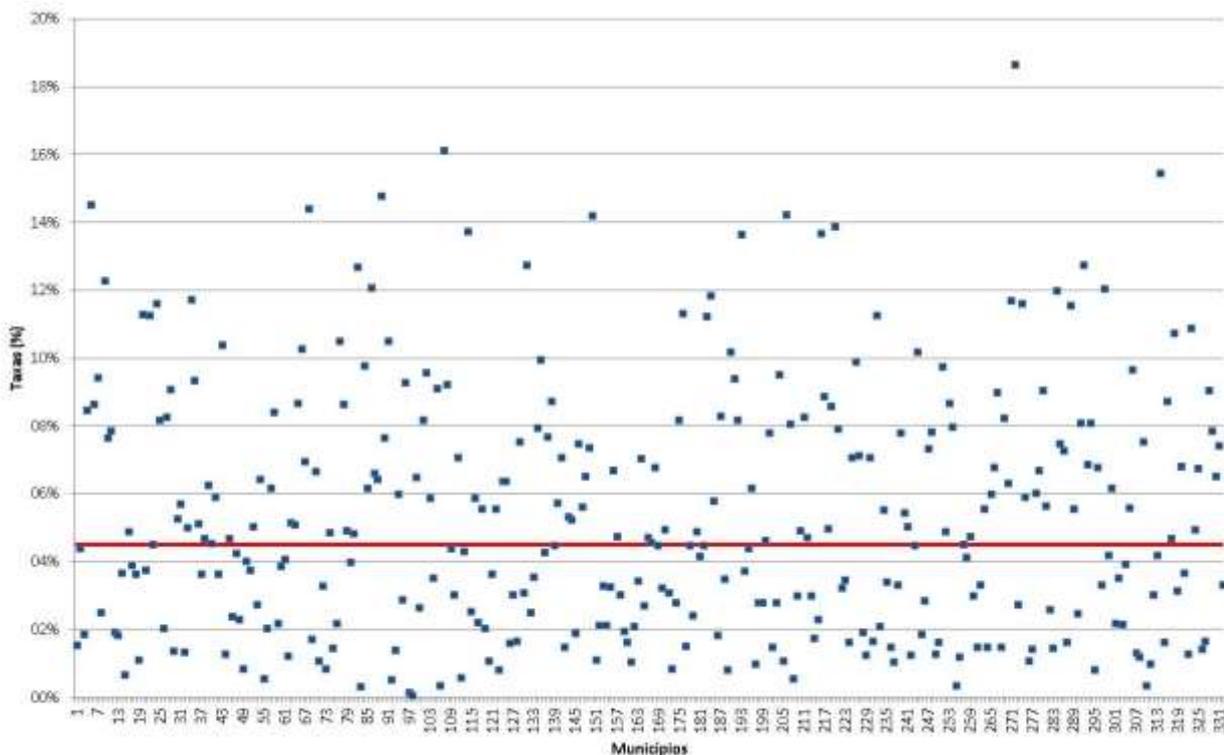


Figura 3. Comparativo entre a taxa do PROAGRO (4,5%) e as taxas municipalizadas calculadas pela consultoria.

Na figura, observa-se duas taxas próximas a 0% refletindo o risco muito baixo desses municípios. Aqui vale uma ressalva: as taxas calculadas neste exercício refletem apenas o risco puro da atividade. Normalmente, sobre o risco puro é acrescido uma taxa de carregamento, que corresponde ao custo tributário, administrativo, operacional e o lucro das seguradoras.

É importante ressaltar que os pontos acima da linha vermelha representam os municípios que possuem taxas maiores do que a atual taxa cobrada pelo programa (4,5%). Nesse contexto é possível ver claramente a receita perdida. Por outro lado, existe um conjunto de produtores que deveriam pagar menos por estarem em municípios com riscos menores. Ao todo, 53% dos municípios apresentaram taxas maiores do que o PROAGRO, e 47%, taxas menores.

Quando se avalia a diferença entre a taxa calculada e a taxa do PROAGRO (4,5%), nota-se uma variabilidade bastante significativa. As diferenças podem chegar a quase a um valor mínimo de -4,4% a um valor máximo de 14,2% (Fig. 4). Ademais, percebe-se a assimetria em direção as diferenças positivas.

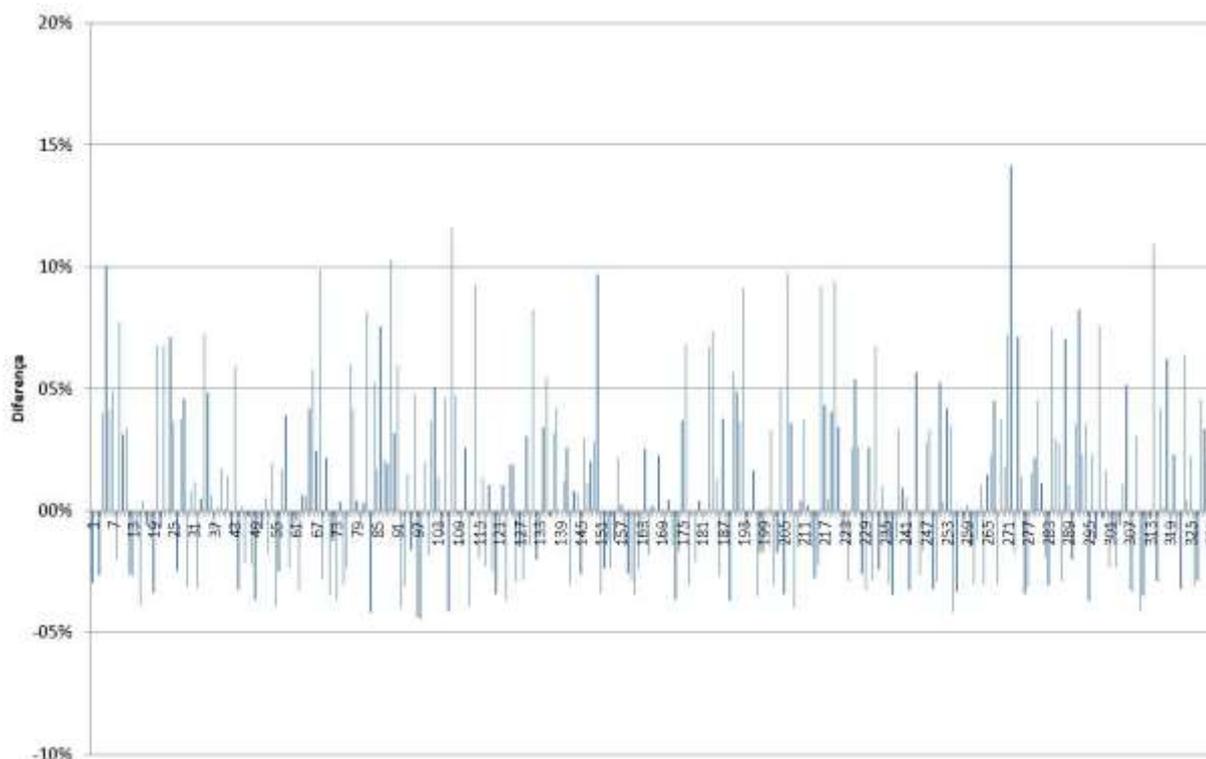


Figura 4. Diferença entre a taxa calculada usando a metodologia proposta neste projeto e a taxa de 4,5% do PROAGRO.

Para se ter uma ideia do ganho financeiro em precificar de acordo com a recomendação metodológica é preciso quantificar as perdas impostas pelo atual modelo de tarifação. Para isso é necessário levantar os dados do volume de crédito rural de custeio contratado no Paraná no ano safra 2017/18, em cada um dos municípios selecionados.

Utilizando a base do Sicor, levantamos esses dados. Para a soja e em todos os municípios selecionados, o valor total do custeio contratado foi de R\$ 4,7 bilhões na safra 2017/18. Assim, comparando a receita total utilizando a taxa única de 4,5% versus a receita total com a nova proposta, teremos uma boa noção do ganho em se aprimorar o modelo atual de precificação. Utilizando a taxa calculada para cada um dos municípios e multiplicando-os

pelos respectivos valores de crédito de custeio, obtemos a PERDA DE RECEITA DO PROAGRO, da ordem de **R\$ 88,3 milhões**.

Ademais, ressalta-se que além da perda de receita pela má especificação da taxa, há o fenômeno da migração dos produtores de baixo risco para o seguro privado, já que em função da subvenção ao prêmio, a taxa é muitas vezes menor do que no PROAGRO. Nesse caso, a PERDA DE RECEITA EM FUNÇÃO DA MIGRAÇÃO para o seguro privado é da ordem de **R\$ 53,8 milhões**. Essa é uma fonte de ineficiência no PROAGRO, na medida em que deixa na carteira apenas os produtores de risco maior, com maior probabilidade de perdas na lavoura, e conseqüentemente, maior a chance do PROAGRO ter que pagar as indenizações. O aumento das indenizações reflete-se em aumento da chance de déficit na carteira e possíveis dispêndios não programados pelo Governo.

Se considerarmos as duas perdas, atinge-se um valor máximo (teto) de perda da ordem de **R\$ 142,1 milhões**. Esse valor representa a perda de receita apenas para a soja em um único estado. Cabe ressaltar que o PROAGRO atua com muitas outras culturas em praticamente todo o território nacional, portanto a perda de receita é muito maior do que o valor supramencionado.

6.15 CONSIDERAÇÕES FINAIS

O estudo mostra uma ampla revisão de literatura sobre a precificação de risco agrícola e detalha a sua metodologia atuarial. A formulação apresentada segue os padrões internacionais utilizados, principalmente, pelo programa de seguro agrícola norte-americano que é o maior programa em nível mundial. No anexos deste relatório, estão incluídas um

sumário das etapas do processo de precificação, além de um script em R utilizado para geração das estimativas feitas no estudo de caso demonstradas nesta seção.

A metodologia presente neste estudo pode ser utilizada tanto para as apólices do mercado privado de seguros quanto para o PROAGRO. Uma vez definido o produto de seguros a ser precificado, depura-se os dados com o objetivo de verificar a presença de tendência e retirá-la da série de produtividade. Na sequência é preciso ajustar as distribuições aos dados, dentre as apresentadas neste estudo, ou seja, a Normal Assimétrica, t-Assimétrica, Odd Log-Logística e a modelagem não paramétrica.

As distribuições paramétricas supramencionadas são suficientes para captar certas idiosincrasias presentes nos dados, por exemplo, assimetrias e bimodalidades. Outras distribuições, por exemplo, Normal, Gama e Beta podem ser consideradas casos particulares das distribuições apresentadas.

Após o ajuste das distribuições é preciso escolher a distribuição que melhor representa os dados. Os critérios AIC e BIC podem ser utilizados para a escolha de modelos. Recomenda-se utilizar apenas um dos critérios apresentados e não os dois concomitantemente. Uma vez definida a distribuição de probabilidade o próximo passo consiste no cálculo das taxas de prêmio. Caso o produto escolhido seja o seguro de faturamento, há ainda um passo complementar que é a escolha da cópula que melhor se ajusta aos dados. Após a escolha da cópula o atuário deverá seguir os passos anteriores para realizar a precificação de risco.

Como exemplo prático, para mostrar os ganhos ao alterar a forma de precificação, vamos comparar as receitas obtidas com a forma atual e a receita com a nova proposta de precificação. Para isso, calculamos as taxas por municípios do Paraná, para a cultura da soja. Antes, porém, é importante destacar um ponto. As perdas financeiras no PROAGRO podem

ocorrer de duas formas distintas: i) perdas de receita pela má especificação da taxa, ou seja, não se distingue entre municípios de alto, médio e baixo risco, associando a todos uma única taxa. Dessa forma, perde-se receita, pois a taxa atual de 4,5% é menor do que aquela que deveria ser cobrada; e, ii) perdas de receita devido a migração de produtores ao seguro privado, pois não estão dispostos a pagar pelo PROAGRO e enxergam no seguro privado uma solução mais barata, devido a subvenção ao prêmio.

Utilizando a base do Sicor, levantamos esses dados. Para a soja e em todos os municípios selecionados, o valor total do custeio contratado foi de R\$ 4,7 bilhões na safra 2017/18. Assim, comparando a receita total utilizando a taxa única de 4,5% versus a receita total com a nova proposta, teremos uma boa noção do ganho em se aprimorar o modelo atual de precificação. Utilizando a taxa calculada para cada um dos municípios e multiplicando-os pelos respectivos valores de crédito de custeio, obtemos a PERDA DE RECEITA DO PROAGRO, da ordem de **R\$ 88,3 milhões**.

Ademais, ressalta-se que além da perda de receita pela má especificação da taxa, há o fenômeno da migração dos produtores de baixo risco para o seguro privado, já que em função da subvenção ao prêmio, a taxa é muitas vezes menor do que no PROAGRO. Nesse caso, a PERDA DE RECEITA EM FUNÇÃO DA MIGRAÇÃO para o seguro privado é da ordem de **R\$ 53,8 milhões**. Essa é uma fonte de ineficiência no PROAGRO, na medida em que deixa na carteira apenas os produtores de risco maior, com maior probabilidade de perdas na lavoura, e conseqüentemente, maior a chance de o PROAGRO ter que pagar as indenizações. O aumento das indenizações reflete-se em aumento da chance de déficit na carteira e possíveis dispêndios não programados pelo Governo.

Se considerarmos as duas perdas, atinge-se um valor máximo (teto) de perda da ordem de **R\$ 142,1 milhões**. Esse valor representa a perda de receita apenas para a soja em um único estado. Cabe ressaltar que o PROAGRO atua com muitas outras culturas em

praticamente todo o território nacional, portanto a perda de receita é muito maior do que o valor supramencionado.

6.16 REFERÊNCIAS

Adami, A. C. O., & Ozaki, V. A. (2016). Estudo Sobre a Viabilidade do Seguro Receita: o caso da região oeste do Paraná. *Revista de Economia e Sociologia Rural*, 54(2), 305-318.

Akaike, H. (1974) A new look at the statistical model identification, *Automatic Control, IEEE Transactions on* 19, 716–723.

Alizadeh, M; Emadi, M.; Doostparast, M.; Cordeiro, G.M.; Ortega, E.M.; Pescim, R.R. (2015) A new family of distributions: the kumaraswamy odd log-logistic, properties and applications, *Hacetatepe Journal of Mathematics and Statistics*, 44.

Azzalini, A. A class of distributions wich includes the normal ones, *Scand. J. Statist.* 12 (1985), pp. 171-178.

Azzalini, A.; Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 367–389.

Booth, P.; Chadburn, R.; Cooper, D.; Haberman, S.; James, D. (1999) *Modern actuarial theory and practice*. London: Chapman & Hall/CRC, 716p.

Botts, R.R.; Boles, J.N. (1958) Use of normal-curve theory in crop insurance ratemaking, *Journal of Farm Economics* 40, 733-740.

Braga, A. S.; Cordeiro, G.M.; Ortega, E.M.; Cruz, J.N. (2016) The odd log–logistic normal distribution: Theory and applications in analysis of experiments, *Journal of Statistical Theory and Practice*, 1–25.

Brisolara, C. S. (2013). Proposições para o desenvolvimento do seguro de receita agrícola no Brasil: do modelo teórico ao cálculo das taxas de prêmio. Tese de Doutorado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. doi:10.11606/T.11.2013.tde-02102013-141823. Recuperado em 2019-03-20, de www.teses.usp.br.

Cruz, J.N. (2016) A nova família de distribuições odd log-logística: teoria e aplicações, Ph.D. diss., Escola Superior de Agricultura “Luiz de Queiroz.

Cruz, J.N.; Ortega, E.M.; Cordeiro, G.M. (2015) The log-odd log-logistic weibull regression model: modelling, estimation, influence diagnostics and residual analysis, *Journal of Statistical Computation and Simulation*, 1–23.

Day, R.H. (1965) Probability distributions of field crop yields, *Journal of Farm Economics*, 47, 713-741.

Duarte, G. V. (2018). Precificação do seguro agrícola: novas abordagens utilizando distribuições de probabilidades alternativas e o uso de cópulas bidimensionais e tridimensionais. Tese de Doutorado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. Recuperado em 2019-03-20, de <http://www.teses.usp.br/teses/disponiveis/11/11134/tde-17012019-152335/>.

Gallagher, P. (1987) U.S. Soybean yields: estimation and forecasting with nonsymmetric disturbances, *American Journal of Agricultural Economics*, 69, 796-803.

Goodwin, B.K.; Ker, A.P. (1998) Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts, *American Journal of Agricultural Economics*, 80, 139-153.

Goodwin, B.K.; Mahul, O. (2004) Risk modeling concepts relating to the design and rating of agricultural insurance contracts, World Bank.

Hart, D. G.; Buchanan, R. A.; Howe, B. A. (1996) The actuarial practice of general insurance. 5 ed. Australia: The Institute of Actuaries of Australia, 592p.

Joe, H. (2014) Dependence modeling with copulas. CRC Press.

Joe, H.; Xu, J. J. (1996) The estimation method of inference functions for margins for multivariate models. Technical report.

Just, R.E.; Weninger, Q. (1999) Are crop yields normally distributed?, *American Journal of Agricultural Economics*, 81, 287-304.

Ker, A.P.; Goodwin, B.K. (2000) Nonparametric estimation of crop insurance rates revisited, *American Journal of Agricultural Economics*, 83, 463-478.

Ker, A.P.; Coble, K. (2003) Modeling conditional yield densities, *American Journal of Agricultural Economics*, 85, 291-304.

Moss, C.B.; Shonkwiler, J.S. (1993) Estimating yield distributions with a stochastic trend and nonNormal errors, *American Journal of Agricultural Economics*, 75, 1056-1062.

Miqueleto, G. J. (2011). Contribuições para o desenvolvimento do seguro agrícola de renda para o Brasil: evidências teóricas e empíricas. Tese de Doutorado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. doi:10.11606/T.11.2011.tde-12092011-163544. Recuperado em 2019-03-20, de www.teses.usp.br.

Nelson, C.H.; Preckel, P.V. (1989) The conditional beta distribution as a stochastic production function, *American Journal of Agricultural Economics*, 71, 370-378.

Ozaki, V.A.; Goodwin, B. K. ; Ghosh, S. ; Shirota, R. (2008a) Spatio-temporal modeling of agricultural yield data applied to pricing crop insurance contracts. *American Journal of Agricultural Economics*, 90, 951-961.

Ozaki, V.A.; Goodwin, B. K. ; Shirota, R. (2008b) Parametric and nonparametric statistical modeling of crop yield: implications for pricing crop insurance contracts. *Applied Economics*, 40, 1151-1164.

Ozaki, V.A.; Silva, R. S. (2009) Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *Journal of Applied Statistics*, 36, 443-452.

Polo, L. (2016). Redes Bayesianas aplicadas a estimação da taxa de prêmio de seguro agrícola de produtividade. Dissertação de Mestrado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. doi:10.11606/D.11.2016.tde-10082016-132524. Recuperado em 2019-03-20, de www.teses.usp.br.

Ramirez, O.A. (1997) Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, nonNormal random variables: the case of corn belt corn, soybean and wheat yields, *American Journal of Agricultural Economics*, 79, 191-205.

Ramirez, O.A.; Misra, S.; Field, J. (2003) Crop-yield distributions revisited, *American Journal of Agricultural Economics*, 85, 108-120.

Santos, C. O., Scalon, J. D., & Ozaki, V. A. (2014). A distribuição normal-assimétrica como modelo para produtividade de milho aplicada ao seguro agrícola. *Revista de Economia e Sociologia Rural*, 52(4), 725-742.

Schwarz, G. (1978) Estimating the dimension of a model, *The annals of statistics*, 6, 461–464.

Sherrick, B.J.; Zanini, F.C.; Schnitkey, G.D.; Irwin, S.H. (2004) Crop insurance valuation under alternative yield distributions, *American Journal of Agricultural Economics*, 86, 406-419.

Silverman, B.W. (1986) *Density estimation for statistics and data analysis*, Chapman and Hall, London, 175p.

Sklar, M. (1959) *Fonctions de repartition a n dimensions et leurs marges*. Universite Paris, 8.

Taylor, C.R. (1990) Two practical procedures for estimating multivariate nonNormal probability density functions, *American Journal of Agricultural Economics*, 72, 210-217.

Turvey, C.; Zhao, C. (1999) Parametric and nonparametric crop yield distributions and their effects on all-risk crop insurance premiums, Working paper, Dep. of Agricultural Economics and Business, University of Guelph.

7. SEÇÃO 6 - ANÁLISE CONTRAFACTUAL – CONSIDERANDO O EFEITO DA MUDANÇA CLIMÁTICA

7.1 INTRODUÇÃO

De acordo com os Termos de Referência do Projeto intitulado “Investimento Público e Adaptação à Mudança do Clima na América Latina (IPACC II)”, PN: 2014.9049.9-002: “No marco de implementação do projeto IPACC II, as contrapartes principais do Brasil, o Ministério de Planejamento, Desenvolvimento e Gestão (MP), o Ministério da Fazenda (MF) e o Ministério do Meio Ambiente (MMA), identificaram o setor agrícola como sendo um dos setores que aloca altos níveis de investimento público e privado, e ao mesmo tempo sendo um dos setores de maior vulnerabilidade aos efeitos da mudança do clima. Tais condições motivaram a execução de uma iniciativa piloto, conjuntamente com o Ministério de Agricultura, Pecuária e Abastecimento (MAPA), com a finalidade de inserir a gestão do risco climático nos instrumentos de análise de investimentos no setor e, desta forma, complementar o instrumento atualmente utilizado Zoneamento Agrícola de Risco Climático (ZARC) buscando aprimorá-lo a partir da criação de recomendações de novas metodologias científicas, como suporte à política de gestão agrícola.”

O objetivo geral deste projeto é desenvolver um conjunto de recomendações visando reduzir as perdas agrícolas associadas aos efeitos da mudança do clima e que contribuam para uma alocação eficiente dos investimentos públicos e privados na agricultura. Como objetivo específico encontra-se o produto 7: “*Análise contrafactual em relação ao resultado encontrado no produto 5.2, levando em conta os cenários de mudanças do clima – RCP 4.5 e RCP 8.5*”.

7.2 NOTA METODOLÓGICA

Este produto revela qual foi o “custo” em termos de aumento ou redução da produtividade estimada, probabilidade de perda e índice de perda, em média, tomando-se por base o cenário “baseline” e os cenários futuros. As entradas utilizadas foram os resultados dos produtos 2, 4 e 5. Em outras palavras, visando a medição do impacto de cada um dos cenários RCP (Representative Concentration Pathways), desenvolvidos pelo Intergovernmental Panel on Climate Change (IPCC), na probabilidade de perda (PP) e índice de perda (IP) foi calculado o percentual de variação entre cada cenário e o nível atual (baseline).

A variação da produtividade ponderada (Ypon) no clima futuro estimada pelos dois modelos pode ser observada nos mapas apresentadas na próxima seção deste produto. Para a combinação de cenários (baseline e RCP's) para o caso da produtividade e, cenários e níveis de cobertura, no caso de índices de perda (IP) e probabilidade de perda (PP), os valores finais foram obtidos pela média ponderada de todas as variáveis utilizadas para compor as variações tanto da produtividade quanto dos índices e probabilidades de perda, quais sejam: o tipo de solo (1, 2 e 3), o mês de semeadura (decêndios de agosto a dezembro), material genético (precoce, médio e tardio) e dois modelos Dssat e Apsim, conforme os pesos relativos apresentados na Tabela 1 e obtidos junto a literatura, especialistas regionais e ao Departamento de Economia Rural (DERAL), da Secretaria da Agricultura do Paraná. O cálculo das proporções de cada solo foi feito considerando-se o mapa de solos do projeto RADAMBRASIL, escala 1:1000.000. Assim, ao final deste processo, obteve-se um conjunto de 88 mapas, em nível de microrregião, detalhando a variação de Ypon, PP e IP para cada cenário. Ainda, vale mencionar, que a Tabela 1 não representa os valores de IP e PP

resultantes, mas que estão disponíveis na base de dados já compartilhada com a coordenação do produto.

Este produto tem uma configuração espacial diferente da escala espacial anteriormente utilizada por conta das especificidades técnicas. A principal é que ele é apresentado na escala de microrregião, segundo a definição do IBGE, evidentemente após ponderação da área cultivada em cada microrregião, e não na escala municipal, como os anteriores e isso se deve ao fato de que os cálculos de IP e PP foram efetivamente realizados apenas para os 142 municípios para os quais foram constituídas as series meteorológicas. Nos mapas de produtividade, por exemplo, adotou-se a estratégia de interpolação espacial para cobertura integral do Estado do Paraná, mas no caso dos índices PP e IP esta opção não se mostrou tecnicamente viável, uma vez que a variabilidade espacial de PP e IP não segue o mesmo padrão daquele observado pela produtividade e sua interpolação poderia gerar estimativas com excessiva incerteza, levando a erros na tomada de decisão. Por esse motivo, optamos por realizar a média dos valores do PP e IP ponderada pelas respectivas áreas de cada município pertencente a uma microrregião. Segundo o IBGE, vale destacar, microrregião é uma área individualizada dentro de uma mesorregião em uma Unidade da Federação, que apresenta formas de organização do espaço geográfico definidas pelas seguintes dimensões: o processo social, como determinante, o quadro natural, como condicionante e, a rede de comunicação e de lugares, como elemento da articulação espacial. Ainda segundo o IBGE, “as microrregiões foram definidas como parte das mesorregiões que apresentam especificidades quanto à organização do espaço. Essas especificidades se referem à estrutura de produção: agropecuária, industrial, extrativismo mineral ou pesca. Essas estruturas de produção diferenciadas podem resultar da presença de elementos do quadro natural ou de relações sociais e econômicas particulares.”

Assim, a escala adotada é condizente com uma representação formal do IBGE e com a estrutura socioeconômica regional e, caso se pretenda utilizar os dados na escala municipal para a formulação de políticas públicas, por exemplo, pode-se admitir que, dentro de cada microrregião, os municípios ali contidos tenham os mesmos valores de PP e IP. Ainda, esta estratégia foi na tentativa de se evitar a recomendação incorreta de valores de PP e IP para municípios para os quais os valores de produtividade não foram efetivamente estimados e, por conseguinte, dos cálculos atuariais decorrentes desses índices.

Tabela 1. Pesos utilizados, por nível de cada variável, para o cálculo da média ponderada da probabilidade de perda e índice de perda.

Geocódigo	Solo			Mês de Semeadura				Material Genético			Modelo		
	Tipo 1	Tipo 2	Tipo 3	Ago	Set	Out	Nov	Dez	Precoce	Médio	Tardio	DSSAT	APSIM
4100459	0,05	0,05	0,90	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4100509	0,90	0,05	0,05	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4100608	0,90	0,05	0,05	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4100707	0,90	0,05	0,05	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4100905	0,90	0,05	0,05	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4101002	0,05	0,90	0,05	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4101051	0,47	0,00	0,53	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4101150	0,90	0,05	0,05	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4101655	0,37	0,00	0,63	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4101705	0,90	0,05	0,05	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50
4101853	0,51	0,00	0,49	0,00	0,03	0,42	0,46	0,08	0,33	0,33	0,33	0,50	0,50

Fonte: Elaborado pelos autores.

7.3 SOBRE OS MAPAS

A variação relativa da produtividade é apresentada nos oito mapas da Figura 1, elaborados a partir da ponderação de todas as combinações simuladas e descritas na seção anterior deste produto.

Neles, pode-se observar que os cenários que consideram variações mais extremas nas condições de temperatura (Figs. 1C e D) e no modelo HadGen2 (Figs. 1A e 1B) uma tendência de queda na produtividade especialmente na região noroeste do Paraná, com zonas de variação que podem passar de 20% de perda em relação ao clima atual. Os cenários climáticos futuros que apontam condição de variação menos intensa nos cenários futuros (Figs. 1E a 1H), contudo, apontam para uma possibilidade de estabilidade produtividade ou até mesmo pequenos ganhos de até 5% em relação ao clima atual.

Para todas os cenários climáticos simulados, contudo, observa-se que a região Sudeste do Paraná deve ser beneficiada com o clima futuro, o que é uma informação bastante relevante para a estrutura social e econômica do Paraná, uma vez que esta região contém uma importante estrutura produtiva que envolve cooperativas de relevância para o Estado e para o Brasil.

Tais variações produtivas são complexas de serem entendidas sem um modelo baseado em processos, pois a variação da temperatura, da chuva e da concentração do CO₂ interferem em diferentes processos simulados pelos modelos, de modo não linear, e com efeitos nem sempre de mesma magnitude e tendência de variação. A variação no volume de chuva é, talvez, a mais simples de ser interpretada, uma vez que a disponibilidade de água no solo condiciona o potencial da água na copa da vegetação e, portanto, o nível de abertura estomática e das taxas de troca de água e CO₂ entre a copa e a atmosfera. A temperatura, por

sua vez, condiciona a taxa de desenvolvimento da cultura da soja, a taxa de evapotranspiração, tornando assim a cultura mais dependente da chuva regular e em bons volumes. Assim, mantendo-se o volume atual de chuvas, maior temperatura no futuro implicaria em maior demanda hídrica por parte da cultura, o que pode resultar em maior estresse por deficiência hídrica. Nas regiões onde a chuva suplanta as necessidades atuais da cultura, contudo, pode ser que mesmo com aumento da necessidade de água da cultura, não haveria estresse e a produtividade se manteria ou poderia até mesmo elevar-se.

O CO₂ talvez seja o elemento com maior probabilidade de variação positivo no futuro, e a elevação da concentração de CO₂ [CO₂] em plantas C3 tem como consequência o aumento na taxa de assimilação via aumento da pressão parcial de CO₂ intercelular, melhora nas relações hídricas, devido à redução da condutividade estomática e consequente redução dos efeitos negativos do estresse hídrico no crescimento das folhas, e o aumento da temperatura foliar, estimulando a divisão e expansão celular, elevando assim o acúmulo de biomassa (Ghannoum et al., 2000).

Entre os elementos meteorológicos com projeção de mudança nas próximas décadas, a [CO₂] é a que apresenta a menor incerteza em relação a expectativa de elevação. Análises de sensibilidade utilizando os modelos DSSAT/CANEGRO e APSIM-Sugar mostraram que o [CO₂] tem papel fundamental para as projeções de cenários futuros (Marin et al., 2014). Os modelos atualmente disponíveis para simulação de crescimento e desenvolvimento da soja podem fornecer subsídio muito útil para a geração desse tipo de cenário, mas atualmente sabe-se muito mais sobre a cinética da Rubisco, a resposta fotossintética e estomática da cultura à mudança do [CO₂], e sobre os mecanismos de resposta retroalimentadora ao nível foliar na soja (Silva, 2018). Estes avanços já foram quase que integralmente incorporados aos modelos utilizados neste projeto e isso trazer maior confiança às simulações aqui realizadas.

Considerando-se os dois modelos utilizados neste projeto quanto ao modo como as respostas a elevação do [CO₂] foram incorporadas ao modelo, pode-se verificar que o modelo APSIM-Sugar, a elevação na [CO₂] é considerada pela multiplicação dos coeficientes de referência da eficiência de transpiração (TUE) e da eficiência de uso da radiação (RUE) por valores modificadores (“TUE e RUE CO₂ modifiers”), como utilizado por Howden et al. (2007), Webster et al. (2009) e Biggs et al. (2013):

$$\text{TUE-CO}_2\text{-modifier} = 0,0008 \cdot [\text{CO}_2] \quad (1)$$

$$\text{RUE-CO}_2\text{-modifier} = \left(\frac{1}{350} \cdot 0,05\right) \cdot [\text{CO}_2] + 0,95 \quad (2)$$

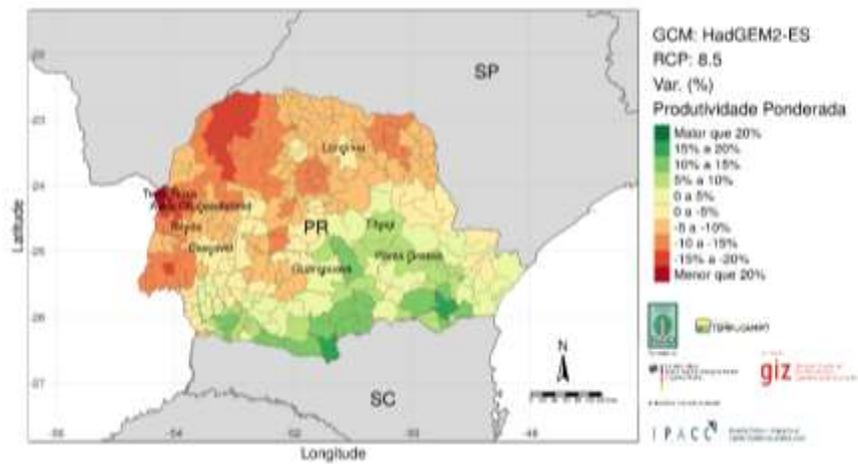
em que TUE-CO₂-modifier é um fator modificador da produtividade de água pela transpiração e RUE-CO₂-modifier é um fator modificador da “eficiência” de uso da radiação solar.

O modelo DSSAT/CROPGRO, por sua vez, computa uma baixa resposta direta da fotossíntese e uma resposta mais pronunciada da transpiração em relação a [CO₂] elevada, principalmente por causa do forte aumento da resistência foliar à difusão de vapor (Allen et al., 1985), conduzindo a uma maior produtividade de água e redução do estresse hídrico. O modelo DSSAT/CROPGRO também utiliza um fator de ajuste para a evapotranspiração de referência derivado de razão teórica entre a evapotranspiração de referência de uma superfície hipotética (representando um gramado) em atmosfera [CO₂]=330 ppm e sob [CO₂] elevada.

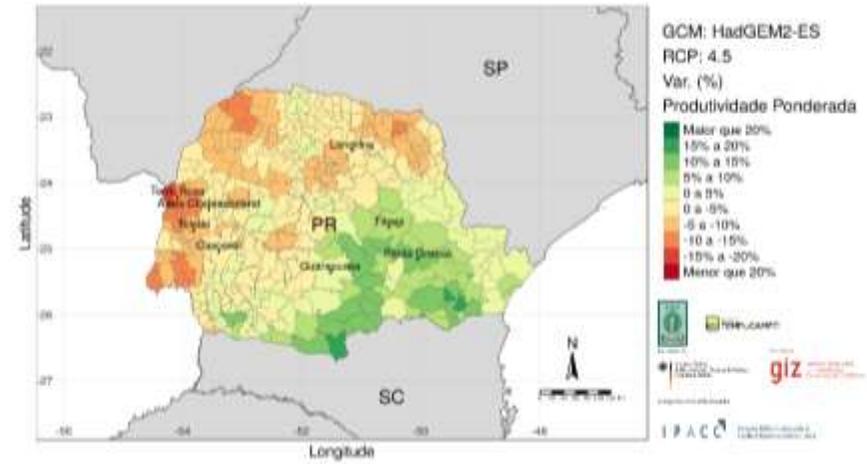
Quanto a fotossíntese, o modelo DSSAT/CROPGRO calcula o incremento diário de biomassa (P_G , $g\ m^{-2}\ d^{-1}$) baseada na abordagem da eficiência de uso da radiação e um algoritmo para computar a fertilização decorrente da $[CO_2]$ elevada, como segue:

$$P_G = FI \cdot PAR \cdot RUE \cdot R_{PG} \quad (5)$$

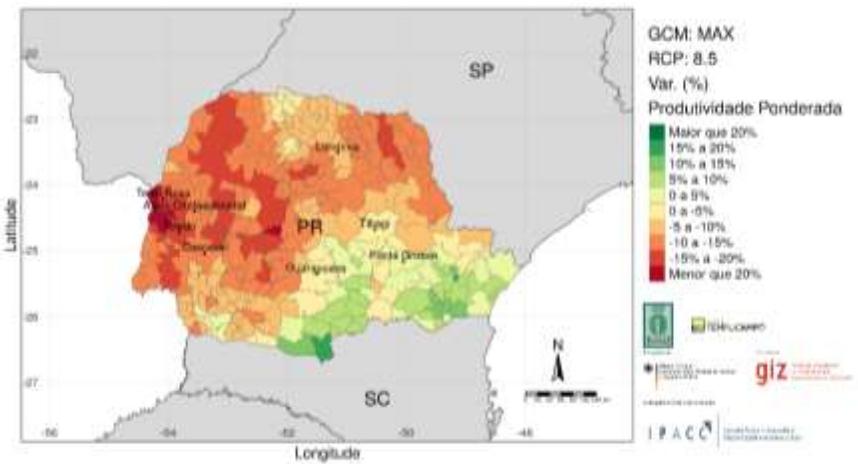
em que FI é a fração de interceptação da densidade de fluxo de fótons fotossinteticamente ativos (PAR , $MJ\ m^{-2}$) e RUE é a eficiência de uso da radiação ($g\ MJ^{-1}$). O valor padrão para P_G em $[CO_2]=330\ ppm$ é então ajustado pelo fator de ajuste (R_{PG}) que depende da $[CO_2]$. O valor de R_{PG} é calculado pela interpolação de um conjunto de pontos coordenados.



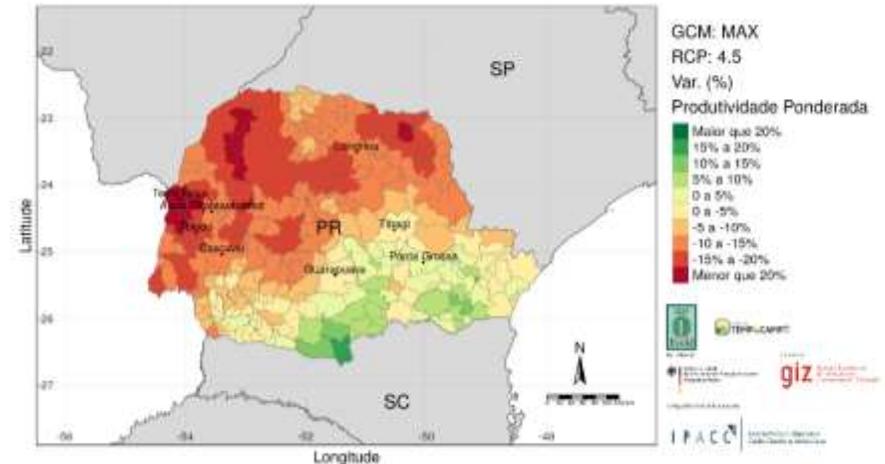
A



B

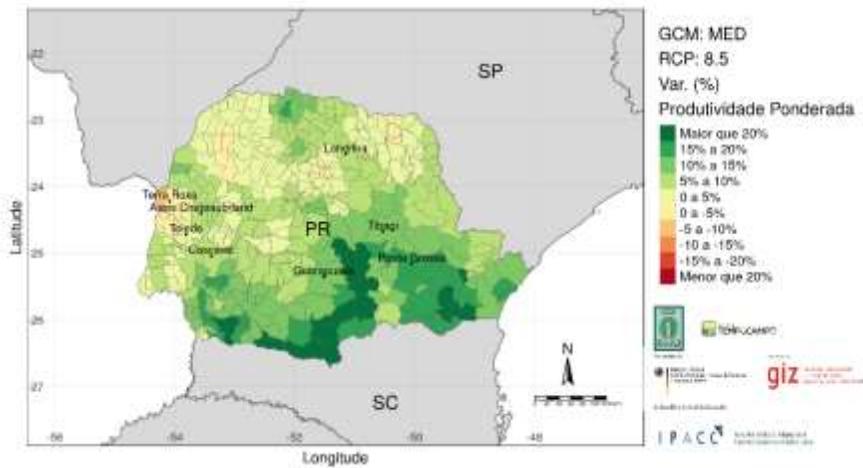


C

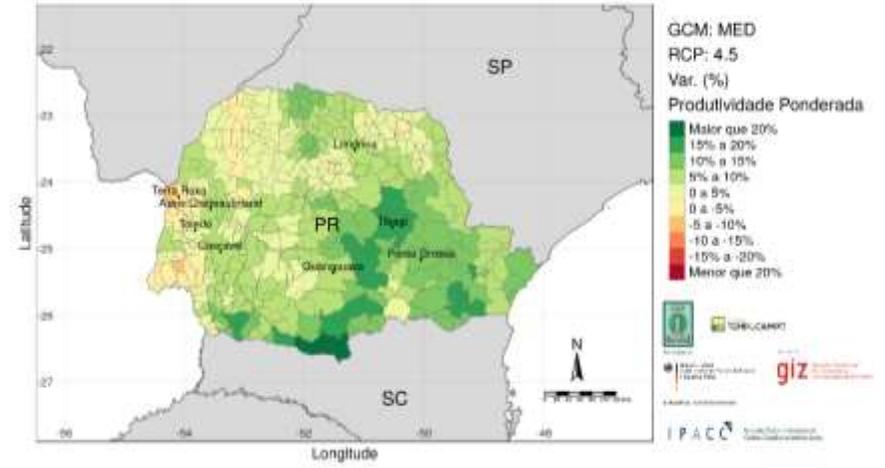


D

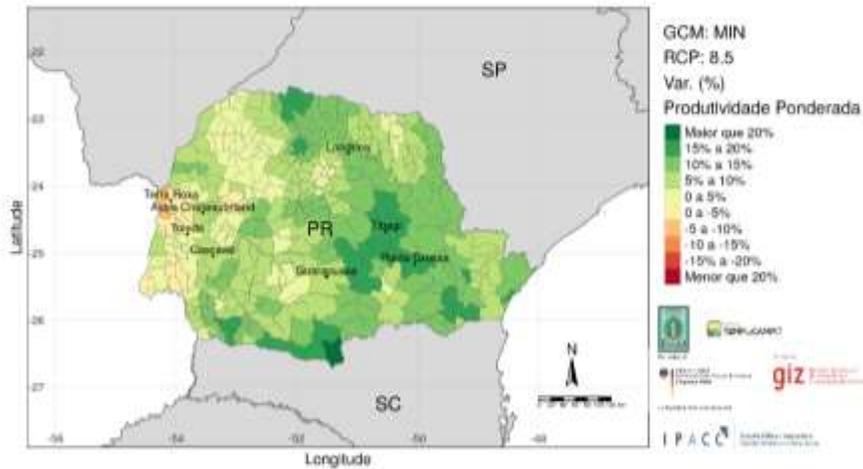
Figura 1. Variação relativa da produtividade ponderada simulada para oito cenários climáticos futuros, considerando como referência a série climática atual (linha de base).



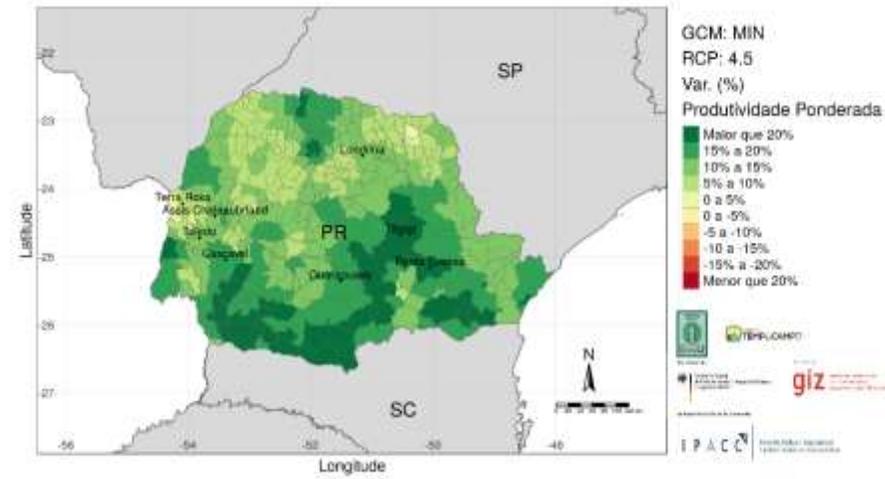
E



F



G



H

Figura 1. Variação relativa da produtividade ponderada simulada para oito cenários climáticos futuros, considerando como referência a série climática atual (linha de base) (continuação).

Com relação a variação relativa de IP e PP nos cenários de clima futuro, a seguir estão apresentados 2 mapas (Figuras 2 e 3) dentre os 80 mapas existentes tratando deste assunto e que ilustram a sequência lógica de todos os mapas. Os mapas estão variando em cinco faixas de níveis de cobertura, de 60 a 80%, 4 GCMs e 2 RCPs, totalizando 40 mapas de probabilidade de perda e 40 mapas de índice de perda.

A Figura 2 mostra os percentuais de variação entre o cenário base e o cenário futuro, levando-se em conta as mudanças do clima. A probabilidade de perda aumenta em cada microrregião conforme as cores passam do verde mais escuro para o vermelho mais escuro. Na Figura 2, nota-se que em todas as regiões, para o nível de cobertura de 60%, a probabilidade de perda tende a se reduzir (valores negativos na escala).

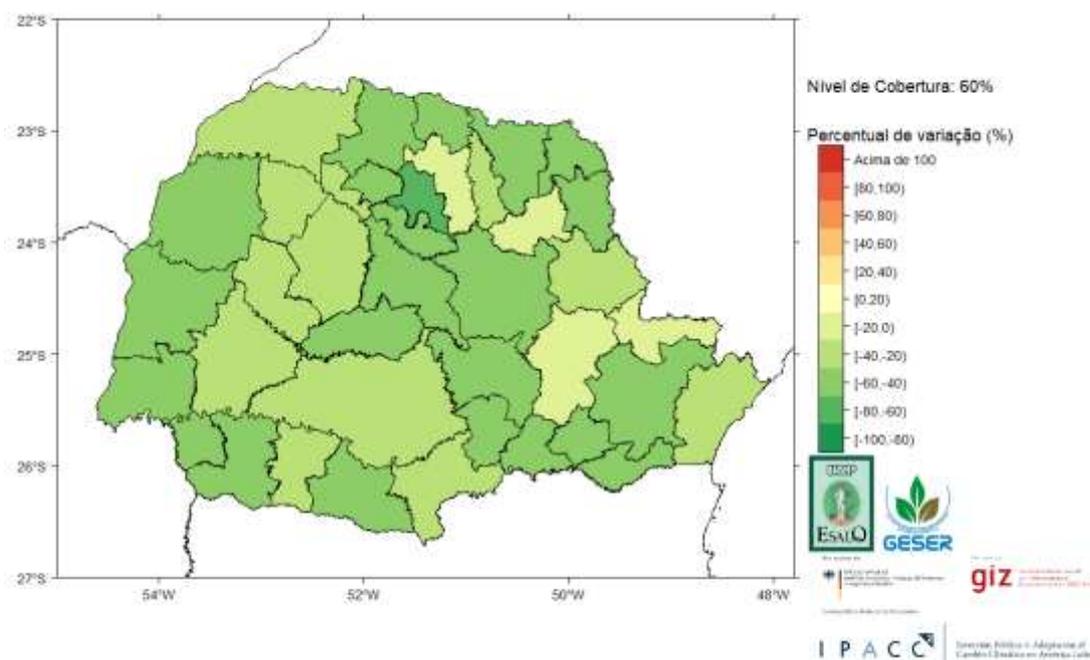


Figura 2. Percentual de variação da probabilidade de perda, ao nível de cobertura de 60%, no cenário GCM HadGem2 e RCP 4.5.

Por outro lado, atentando-se ao índice de perda, nota-se que nas microrregiões em laranja e vermelho, o índice de perda, ao nível de cobertura de 60%, tende a aumentar. Em todas as outras microrregiões, a tendência é de redução do índice de perda, no cenário no cenário GCM HadGem2 e RCP 4.5.

Uma interpretação interessante dos resultados do índice de perda baseia-se no fato de que podemos pensar diretamente em perdas financeiras. Em outras palavras, multiplicando o índice de perda pela respectiva unidade monetária, temos o prejuízo em termos monetários. Por exemplo, se o índice de perda foi de 0,40, para cada 1 real segurado temos um prejuízo de R\$ 0,40. No caso da taxa de variação, um aumento de 100% significa que o prejuízo poderá chegar a R\$ 0,80 considerando um cenário futuro.

A partir dessa constatação, os valores negativos nos mapas e na tabela anexa mostram um cenário positivo, pois a produtividade futura será maior do que a produtividade atual (base). Nessa situação as tabelas e os mapas mostram as variações percentuais dos índices de perda para cada uma das microrregiões.

Em particular, quando se observa a primeira linha na tabela anexa os seguintes percentuais de variação -97,6%, -97,8%, -97,7%, -97,6% e -97,4% mostram a variação percentual entre o cenário base e o cenário futuro. Em outras palavras, nesse cenário futuro, os resultados indicam uma redução drástica no índice de perda, equivalentes a 2,4%, 2,2%, 2,3%, 2,4% e 2,6% do índice de perda base, para cada nível de cobertura.

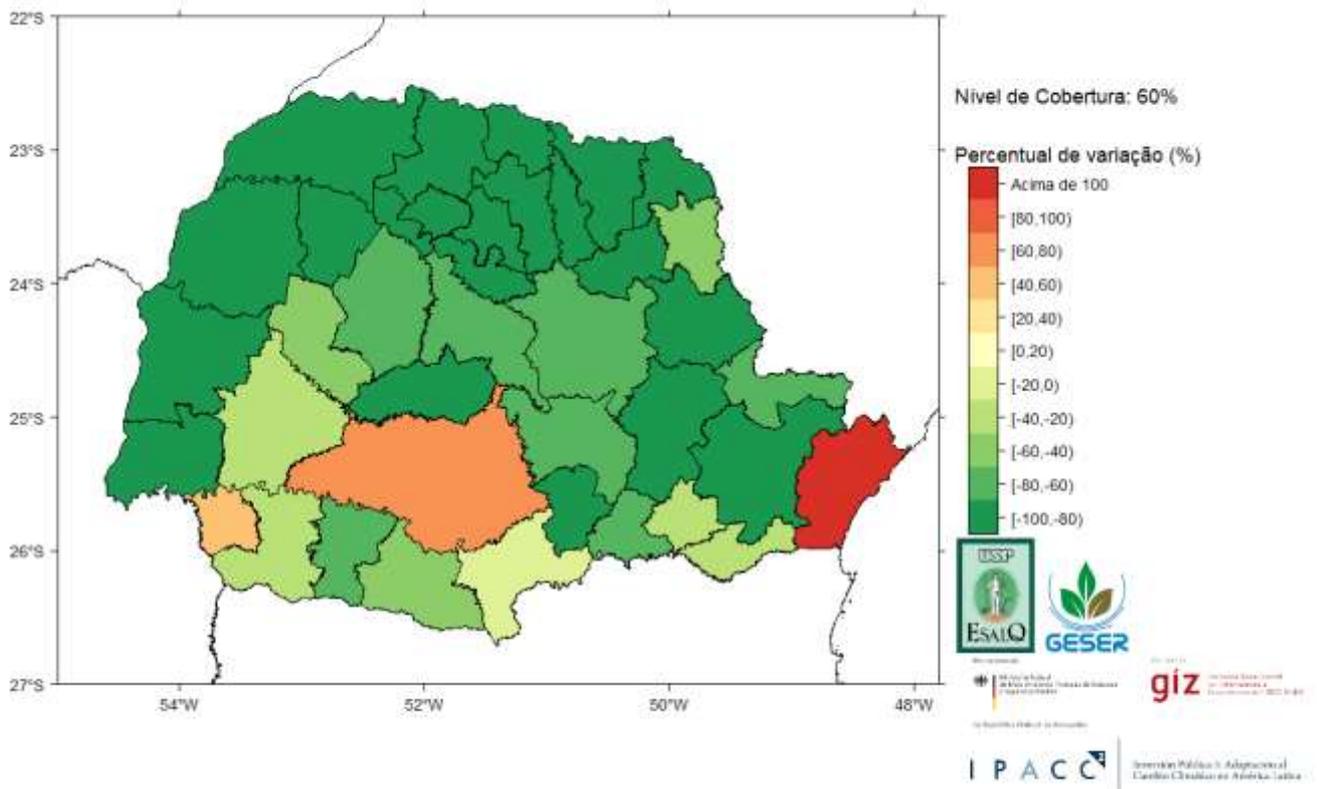


Figura 3. Percentual de variação do índice de perda, ao nível de cobertura de 60%, no cenário GCM HadGem2 e RCP 4.5.

7.4 REPRESENTAÇÃO TABULAR DO ÍNDICE DE PERDA

Nesta representação tabular, as linhas referem-se as microrregiões e os códigos obedecem a seguinte ordem na codificação GCM/RCP/NC, assim, HA45N1, significa HadGem2, RCP4.5, nível de cobertura 60%. Assim, cada coluna representa uma sigla composta das seguintes variáveis:

1. GCM de variação mínima em relação a LB: MIROC5 (MI);
2. GCM de variação média em relação a LB: BNU-ESM (ME);
3. GCM de variação máxima em relação a LB: CanESM2 (MA);
4. GCM adicional: HadGem2 (HA);
5. Concentrações de CO₂ de 526 ppm para o RCP 4.5 (45);
6. Concentrações de CO₂ de 628 ppm para o RCP 8.5 (85);
7. Níveis de cobertura 60, 65, 70, 75, 80% (N1, N2, N3, N4, N5).

HA45N1	HA45N2	HA45N3	HA45N4	HA45N5
-97.6	-97.8	-97.7	-97.6	-97.4
-94.7	-94.8	-94.6	-94.2	-93.5
-95.5	-95.5	-95.3	-94.8	-93.9
-40.7	-43.2	-44.4	-45.1	-43.4
-62.1	-37.3	4.6	40.9	52.6
-98.8	-99.1	-99.2	-99.3	-99.3
-98.0	-98.1	-97.9	-97.0	-95.3
-91.2	-91.6	-91.7	-90.7	-90.0
-93.7	-93.6	-93.3	-92.9	-91.9
-98.7	-99.0	-99.2	-99.3	-99.4
-82.1	-83.9	-84.9	-85.3	-85.2
-82.0	-82.5	-82.2	-82.0	-81.8
-69.2	-71.4	-72.3	-72.5	-71.6
-97.8	-98.3	-98.5	-98.7	-98.7
-91.4	-91.7	-91.3	-90.5	-89.8
-88.8	-87.9	-86.5	-85.4	-83.9
-92.7	-93.0	-93.1	-93.0	-92.8
-58.2	-60.1	-60.8	-60.4	-58.0
-72.4	-72.5	-71.8	-71.2	-70.6
-98.6	-98.3	-97.8	-97.2	-96.0
-83.1	-83.8	-83.6	-83.2	-82.7
-95.2	-95.2	-95.0	-94.7	-94.2

-31.0	-39.3	-43.8	-47.0	-47.8
-95.9	-95.9	-95.7	-95.4	-95.1
55.9	74.9	90.0	99.7	107.1
-29.8	-22.7	-16.7	-13.9	-12.4
-79.2	-79.7	-79.3	-78.5	-77.6
-94.2	-91.6	-81.9	-72.3	-65.0
62.1	92.4	120.9	153.6	181.5
-50.0	-44.6	-33.5	-24.7	-16.7
-69.5	-70.8	-71.4	-71.7	-70.6
-85.4	-84.5	-82.8	-81.0	-78.6
-18.3	-14.7	-10.9	-7.0	-2.2
-74.7	-75.2	-75.4	-75.5	-75.5
-74.2	-76.8	-78.1	-78.4	-78.2
-38.7	-40.0	-38.8	-34.9	-29.1
-91.7	-91.5	-90.7	-89.7	-88.4
-32.8	-26.7	-20.5	-16.3	-13.7
HA85N1	HA85N2	HA85N3	HA85N4	HA85N5
-97.5	-97.5	-97.5	-97.3	-96.9
-94.6	-94.8	-94.5	-94.1	-93.1
-95.5	-95.4	-94.5	-93.5	-91.9
-42.6	-44.9	-46.0	-44.6	-42.7
92.4	124.6	155.6	167.7	137.6
-98.8	-99.1	-99.2	-99.3	-99.0
-98.5	-98.0	-95.7	-93.8	-91.2
-91.1	-91.7	-91.2	-90.5	-89.9
-92.1	-92.6	-92.7	-92.1	-91.3
-98.7	-99.0	-99.2	-99.3	-99.1
-82.6	-84.3	-85.3	-85.3	-85.3
-86.8	-87.7	-86.9	-86.2	-85.7
-69.8	-71.8	-72.8	-73.1	-72.8
-97.8	-97.8	-96.9	-96.4	-94.5
-91.5	-91.5	-91.2	-90.6	-89.5
-88.5	-89.0	-87.9	-86.4	-83.7
-92.6	-93.1	-93.3	-93.3	-92.6
-63.3	-64.2	-64.2	-63.1	-61.3
-81.3	-80.7	-79.3	-77.7	-76.6
-96.7	-96.0	-94.7	-93.6	-92.4
-83.3	-84.2	-84.3	-83.9	-83.3
-94.9	-94.9	-94.9	-94.6	-94.0
-33.5	-41.3	-45.6	-47.5	-48.6
-95.7	-95.8	-95.8	-95.7	-95.0
26.1	40.4	59.1	73.4	83.1
-43.7	-39.5	-32.9	-28.7	-24.7
-78.5	-79.1	-78.8	-78.1	-77.0
-82.7	-73.5	-63.0	-55.4	-48.5
49.3	80.2	126.0	167.7	207.4
-48.8	-45.6	-34.4	-22.6	-12.5
-69.2	-70.8	-71.7	-71.9	-71.4
-84.9	-83.8	-82.0	-80.0	-77.9
-22.2	-20.4	-16.4	-11.4	-4.6
-81.3	-80.8	-80.3	-80.0	-79.7

-75.3	-77.6	-78.3	-78.5	-78.6
-41.0	-43.0	-42.1	-38.8	-34.6
-92.7	-92.7	-91.8	-90.4	-88.1
-43.3	-39.8	-32.8	-27.1	-20.3
MA45N1	MA45N2	MA45N3	MA45N4	MA45N5
-97.2	-96.9	-96.5	-95.7	-94.8
-92.3	-92.2	-91.8	-91.1	-90.1
-90.5	-90.3	-89.7	-88.3	-87.0
-35.5	-38.3	-38.6	-36.3	-32.5
-9.1	20.3	59.6	82.2	83.9
-98.7	-98.9	-98.5	-97.5	-96.5
-85.8	-85.4	-84.7	-82.5	-78.9
-89.5	-90.0	-89.1	-88.5	-88.0
-89.7	-90.2	-89.3	-88.7	-88.2
-98.7	-99.0	-99.2	-99.3	-99.4
-79.1	-81.5	-82.3	-82.7	-82.7
-84.4	-84.0	-83.5	-83.1	-81.8
-70.5	-70.6	-69.9	-68.4	-66.7
-97.2	-96.2	-95.2	-94.6	-93.7
-87.7	-87.3	-86.3	-85.3	-84.1
-80.1	-80.0	-78.5	-77.0	-74.0
-91.3	-91.8	-91.9	-91.1	-90.3
-50.9	-52.8	-53.5	-52.2	-50.5
-78.2	-77.5	-75.3	-73.3	-71.4
-96.2	-95.0	-93.1	-90.9	-89.2
-80.3	-80.9	-80.2	-79.5	-78.4
-94.7	-94.0	-93.0	-92.0	-90.9
-25.7	-34.2	-38.5	-41.2	-41.5
-93.9	-93.4	-92.5	-91.2	-89.9
68.9	91.5	121.5	153.7	183.2
-27.2	-18.5	-6.3	6.8	20.8
-72.6	-73.4	-73.2	-72.2	-70.5
-88.3	-82.6	-71.2	-61.5	-52.3
66.3	102.3	147.6	188.4	229.6
-34.7	-33.2	-22.8	-11.9	-0.5
-70.7	-70.9	-70.4	-69.5	-68.2
-85.3	-84.0	-81.4	-78.6	-75.9
-6.3	-5.0	0.7	6.7	16.5
-76.1	-76.0	-76.1	-76.1	-75.4
-71.0	-73.9	-75.0	-75.4	-75.5
-30.3	-32.5	-29.0	-23.6	-17.5
-89.7	-89.1	-87.4	-85.3	-81.8
-27.9	-19.0	-10.1	-2.0	7.4
MA85N1	MA85N2	MA85N3	MA85N4	MA85N5
-97.3	-97.2	-97.0	-96.4	-95.6
-92.7	-92.9	-92.2	-91.6	-90.8
-94.5	-94.7	-94.7	-93.7	-92.7
-41.0	-42.6	-41.4	-39.7	-37.2
-61.4	-55.3	-27.5	10.3	20.3
-98.5	-98.7	-98.8	-97.8	-96.7
-79.9	-81.4	-82.1	-82.6	-81.9

-89.2	-89.8	-88.9	-88.3	-87.3
-89.3	-89.6	-88.8	-88.2	-87.7
-98.7	-99.0	-99.2	-99.3	-99.3
-85.8	-86.7	-87.1	-87.2	-86.7
-78.8	-79.5	-79.7	-78.5	-77.3
-72.0	-72.4	-71.5	-70.6	-69.1
-97.5	-97.7	-97.7	-97.6	-97.6
-89.2	-88.4	-87.2	-86.2	-85.4
-82.8	-83.5	-81.7	-80.2	-78.5
-90.5	-90.9	-90.5	-89.6	-88.8
-53.9	-55.7	-56.1	-56.1	-54.5
-77.2	-77.2	-75.4	-74.0	-72.8
-97.7	-97.4	-97.0	-96.2	-94.9
-84.4	-84.9	-84.6	-83.8	-82.7
-94.4	-94.1	-93.6	-92.7	-91.6
-32.4	-39.2	-43.0	-44.5	-45.1
-94.4	-94.4	-93.8	-93.0	-92.0
99.0	118.3	132.1	149.5	173.4
-15.9	-10.8	-6.2	0.2	12.1
-74.9	-75.7	-75.4	-74.6	-72.9
-90.1	-80.6	-73.5	-65.4	-58.8
102.3	149.7	198.4	233.5	257.3
-38.4	-25.7	-11.2	1.3	10.3
-69.2	-70.0	-70.3	-70.2	-69.5
-85.2	-84.1	-82.0	-79.9	-77.8
-5.0	-1.4	8.1	15.7	22.5
-71.3	-72.6	-72.9	-71.5	-70.3
-79.5	-80.8	-81.5	-81.8	-81.3
-22.7	-22.8	-19.0	-13.2	-8.6
-90.1	-89.3	-88.0	-86.8	-84.8
-20.0	-12.9	-5.0	4.4	14.6
ME45N1	ME45N2	ME45N3	ME45N4	ME45N5
-97.1	-97.3	-97.4	-97.2	-96.9
-94.3	-94.3	-93.9	-93.4	-92.7
-95.8	-95.8	-94.9	-93.4	-92.3
-37.4	-41.1	-42.2	-42.1	-37.8
-38.2	-14.6	23.2	53.4	53.2
-98.8	-99.1	-99.2	-99.2	-99.2
-97.7	-96.3	-94.8	-93.2	-91.6
-89.8	-89.1	-88.5	-88.0	-87.5
-91.0	-90.8	-89.8	-89.2	-88.6
-98.7	-99.0	-99.2	-99.3	-99.4
-82.5	-84.3	-85.2	-85.5	-85.4
-80.7	-81.2	-81.5	-81.5	-81.5
-63.5	-66.9	-68.7	-69.3	-67.5
-97.8	-98.3	-98.5	-98.7	-98.8
-88.8	-88.7	-88.6	-88.4	-87.9
-83.4	-83.3	-83.2	-83.0	-81.7
-92.4	-92.8	-92.9	-92.8	-92.5
-52.4	-55.3	-56.9	-57.0	-53.6
-70.9	-70.4	-69.9	-69.3	-68.0

-98.5	-98.3	-97.7	-96.5	-95.3
-83.4	-84.1	-83.8	-83.3	-82.4
-94.5	-94.6	-94.4	-94.0	-93.4
-33.0	-40.7	-44.2	-47.3	-47.5
-96.0	-96.0	-95.6	-94.9	-94.2
10.2	27.8	48.3	66.6	78.8
-46.4	-40.0	-33.9	-29.8	-26.5
-76.8	-77.6	-77.2	-76.5	-75.3
-94.3	-90.3	-79.0	-69.4	-62.0
61.3	92.6	121.3	145.2	172.7
-45.3	-36.2	-27.2	-19.6	-13.4
-60.3	-64.1	-66.0	-66.9	-64.6
-88.9	-87.7	-86.5	-84.3	-81.6
-18.0	-14.8	-9.4	-4.4	0.3
-73.6	-74.5	-75.0	-75.5	-75.8
-74.9	-77.2	-78.3	-78.8	-78.6
-38.5	-39.8	-38.5	-34.8	-30.8
-90.4	-90.2	-89.7	-88.8	-87.5
-51.9	-48.5	-40.0	-33.2	-28.2
ME85N1	ME85N2	ME85N3	ME85N4	ME85N5
-97.4	-97.5	-97.4	-97.2	-96.9
-94.5	-94.6	-94.3	-94.0	-93.5
-95.4	-95.5	-95.3	-94.9	-93.8
-39.0	-41.9	-43.3	-43.8	-41.3
-31.6	-5.7	25.8	52.5	56.0
-98.8	-99.1	-99.2	-99.3	-99.3
-76.3	-78.7	-79.8	-79.9	-78.9
-90.1	-90.5	-89.6	-88.8	-88.3
-92.0	-92.1	-92.1	-91.3	-90.5
-98.7	-99.0	-99.2	-99.3	-99.4
-81.4	-83.4	-84.5	-84.6	-84.6
-80.6	-81.1	-81.3	-81.2	-80.9
-67.3	-69.5	-70.6	-71.4	-70.5
-97.8	-98.3	-98.5	-98.7	-98.8
-90.0	-90.0	-89.6	-89.1	-88.6
-87.1	-86.8	-85.9	-85.1	-83.7
-92.7	-92.9	-92.9	-92.9	-92.8
-54.3	-56.7	-57.9	-58.6	-57.0
-77.9	-76.3	-74.8	-73.7	-72.5
-98.6	-98.2	-97.5	-96.5	-95.2
-82.8	-84.0	-83.8	-83.3	-82.8
-94.8	-94.8	-94.7	-94.4	-93.9
-33.9	-41.4	-45.2	-48.6	-50.7
-95.7	-95.7	-95.6	-95.4	-94.8
13.9	29.9	50.6	68.3	80.6
-50.3	-44.6	-35.4	-30.1	-26.1
-78.1	-79.0	-78.7	-78.2	-77.2
-94.9	-92.6	-81.4	-72.6	-66.0
62.8	93.9	123.3	145.9	170.7
-45.2	-40.0	-31.1	-22.9	-16.4
-63.7	-66.4	-67.9	-68.8	-68.2

-88.0	-87.6	-86.6	-84.8	-82.9
-22.4	-19.6	-15.8	-10.3	-5.2
-72.9	-73.9	-74.6	-75.1	-75.4
-73.3	-75.9	-77.3	-77.6	-77.6
-41.1	-41.8	-40.2	-36.9	-32.1
-91.9	-91.7	-90.9	-89.9	-88.8
-48.9	-46.2	-37.8	-31.0	-26.2
MI45N1	MI45N2	MI45N3	MI45N4	MI45N5
-96.0	-96.2	-96.0	-95.5	-95.1
-92.0	-92.2	-92.0	-91.7	-91.3
-88.6	-88.3	-88.0	-87.3	-85.9
-21.7	-27.6	-29.3	-28.3	-25.1
-34.1	-0.2	38.3	65.3	62.0
-98.8	-99.1	-99.2	-99.2	-99.2
-92.6	-90.6	-89.3	-88.6	-86.0
-70.1	-73.6	-74.7	-75.4	-75.8
-88.2	-87.6	-87.3	-86.9	-86.6
-98.7	-99.0	-99.2	-99.3	-99.4
-78.2	-80.0	-81.0	-81.3	-81.0
-78.9	-79.8	-80.2	-80.5	-80.4
-56.4	-61.7	-64.0	-64.0	-61.8
-97.8	-98.3	-98.5	-98.7	-98.8
-79.9	-81.8	-82.6	-82.6	-82.4
-70.8	-73.8	-74.3	-73.9	-73.3
-91.5	-92.0	-92.1	-92.2	-91.7
-28.8	-37.4	-41.8	-41.4	-37.3
-54.9	-57.6	-58.4	-57.6	-56.0
-97.0	-95.9	-94.5	-93.1	-91.5
-79.8	-80.1	-79.9	-79.4	-78.4
-90.0	-90.9	-90.6	-89.9	-89.2
-16.9	-28.8	-33.8	-35.1	-34.9
-93.2	-93.1	-92.9	-92.4	-91.9
12.1	34.6	56.6	74.6	86.0
-19.4	-15.4	-11.0	-10.0	-8.9
-73.1	-74.2	-74.7	-74.1	-72.4
-95.2	-89.0	-75.7	-65.0	-57.1
69.1	101.8	130.9	153.3	184.4
-34.9	-25.9	-17.6	-10.7	-4.9
-52.7	-58.3	-60.7	-60.5	-58.9
-88.5	-87.1	-85.5	-83.1	-79.9
1.6	4.3	7.8	11.3	16.3
-72.5	-73.7	-74.4	-74.8	-74.8
-68.1	-70.7	-72.2	-72.8	-72.4
-29.1	-30.4	-26.0	-20.9	-17.6
-86.3	-86.8	-86.6	-85.9	-84.7
-48.2	-44.8	-35.0	-28.7	-24.1
MI85N1	MI85N2	MI85N3	MI85N4	MI85N5
-95.2	-95.1	-94.7	-93.9	-93.1
-89.0	-89.6	-89.3	-89.1	-88.2
-89.3	-89.7	-89.9	-89.1	-87.6
10.7	1.4	-3.3	-5.4	-1.3

-8.2	32.3	101.3	143.1	131.8
-98.8	-98.9	-98.8	-98.6	-97.6
-98.8	-98.9	-98.9	-98.4	-95.6
-68.9	-71.5	-73.0	-74.0	-74.6
-84.9	-83.9	-82.7	-82.0	-81.5
-98.7	-99.0	-99.2	-99.3	-99.3
-68.4	-72.1	-72.9	-73.5	-73.9
-79.4	-80.1	-78.4	-77.5	-76.8
-32.6	-40.6	-44.6	-46.6	-46.5
-97.8	-98.3	-98.5	-98.7	-98.9
-74.3	-76.4	-77.5	-78.2	-78.1
-67.4	-68.8	-69.5	-69.5	-66.6
-86.6	-88.0	-88.8	-89.2	-89.1
-10.7	-18.7	-22.9	-24.8	-22.9
-50.6	-50.5	-50.0	-48.0	-46.6
-94.6	-94.2	-92.0	-89.3	-87.4
-70.6	-71.4	-71.0	-70.3	-69.8
-89.6	-90.0	-90.2	-90.1	-89.3
10.9	-4.0	-12.3	-18.7	-22.9
-90.7	-90.7	-90.5	-89.4	-88.3
118.8	134.4	165.3	189.7	205.2
-5.7	2.6	15.5	21.3	26.1
-61.7	-62.9	-63.3	-62.9	-61.6
-93.9	-93.8	-85.1	-66.9	-52.5
169.6	215.5	255.6	283.5	306.1
-21.0	-4.7	12.2	24.9	35.0
-28.1	-36.6	-41.5	-43.9	-42.7
-77.7	-77.3	-74.6	-72.2	-69.5
37.6	37.2	37.2	41.9	48.4
-76.5	-75.4	-73.6	-72.1	-71.3
-57.5	-61.5	-63.0	-63.5	-63.7
11.9	4.6	1.7	0.4	5.4
-84.0	-83.8	-83.5	-82.2	-80.3
-10.4	-6.8	5.9	14.8	20.3

8. SEÇÃO 7 - DESCRIÇÃO METODOLÓGICA PARA OBTENÇÃO DOS PRODUTOS 2 A 7 (SEÇÕES 1 A 6 DESTE RELATÓRIO)

8.1 INTRODUÇÃO

De acordo com os Termos de Referência do Projeto intitulado “Investimento Público e Adaptação à Mudança do Clima na América Latina (IPACC II)”, PN: 2014.9049.9-002:

“No marco de implementação do projeto IPACC II, as contrapartes principais do Brasil, o Ministério de Planejamento, Desenvolvimento e Gestão (MP), o Ministério da Fazenda (MF) e o Ministério do Meio Ambiente (MMA), identificaram o setor agrícola como sendo um dos setores que aloca altos níveis de investimento público e privado, e ao mesmo tempo sendo um dos setores de maior vulnerabilidade aos efeitos da mudança do clima. Tais condições motivaram a execução de uma iniciativa piloto, conjuntamente com o Ministério de Agricultura, Pecuária e Abastecimento (MAPA), com a finalidade de inserir a gestão do risco climático nos instrumentos de análise de investimentos no setor e, desta forma, complementar o instrumento atualmente utilizado Zoneamento Agrícola de Risco Climático (ZARC) buscando aprimorá-lo a partir da criação de recomendações de novas metodologias científicas, como suporte à política de gestão agrícola.”

O objetivo geral deste projeto é desenvolver um conjunto de recomendações visando reduzir as perdas agrícolas associadas aos efeitos da mudança do clima e que contribuam para uma alocação eficiente dos investimentos públicos e privados na agricultura. Como objetivo específico encontra-se o produto 8: “*Descrição metodológica do passo a passo para obtenção dos produtos 2 a 7, a fim de servir como guia para sua replicação em outras regiões e para outras culturas*”.

8.2 NOTA EXPLICATIVA

Este produto mostra a descrição metodológica e o passo a passo para obtenção dos resultados 2 a 7. O documento está dividido em 9 seções. A primeira detalha a descrição do produto 2. A segunda seção mostra como se obtém o produto 3, em seguida, a seção 3 e 4, detalham os produtos 4 e 5. A seção 5 detalha o passo a passo para calcular a taxa de prêmio e a seção 6 mostra a forma de obtenção do produto 7.

8.3 NOTA METODOLÓGICA DO PRODUTO 2

A simulação com os modelos baseados em processos para a produtividade da cultura da soja no Paraná foi feita com dois dos sistemas de simulação mais reconhecidos científicas no mundo atualmente. São eles o Sistema de Suporte à Decisão e Transferência de Agrotecnologia – DSSAT, dentro do qual utilizou-se o modelo CropGro-Soybean e o Simulador de sistemas produtivos agrícolas – APSIM, sendo utilizado o APSIM-Soybean como modelo de simulação da cultura da soja. O Sistema de Suporte à Decisão e Transferência de Agrotecnologia (DSSAT do inglês Decision Support System Agrotechnology Transfer) simula o crescimento e desenvolvimento de 16 culturas ao longo do tempo, levando em conta o clima, o solo, o manejo e aspectos genéticos da cultura e vem sendo utilizado nos últimos 30 anos por usuários de todo mundo. O DSSAT contém também uma coleção de programas independentes (Figura 1) para organização dos dados de clima, solo, observações de campo, condições experimentais e informações genotípicas (JONES et al, 2003).

Este sistema incorpora diferentes modelos em um software que facilita a avaliação e aplicação dos modelos de culturas para diferentes propósitos (JONES et al., 2003). James & Cutforth (1996) descrevem o DSSAT como um conjunto de programas de computador

projetados para acomodar modelos padronizados de culturas, permitindo ao usuário inserir, organizar e armazenar dados de cultura, solo e climáticos, permite a calibração e validação de modelos de crescimento de culturas e, também, permite a avaliação de diferentes sistemas de manejo em um local.

Cada modelo é estruturado utilizando abordagem modular, como descrito por Jones et al. (2001) Suas características mais importantes são: a separação dos módulos por linhas disciplinares; definição clara das interfaces para cada módulo; permite que componentes individuais possam ser inseridos ou retirados com pouco impacto no programa principal; facilitação na documentação e manutenção dos códigos; permite que programações em diferentes linguagens possam ser ligados; permite uma fácil integração de diferentes tipos de aplicações; aceita para evolução do programa a integração de outros componentes como pecuária e consórcio de culturas; facilita a cooperação entre diferentes grupos de modelos onde cada um pode ser concentrado em módulos específicos.

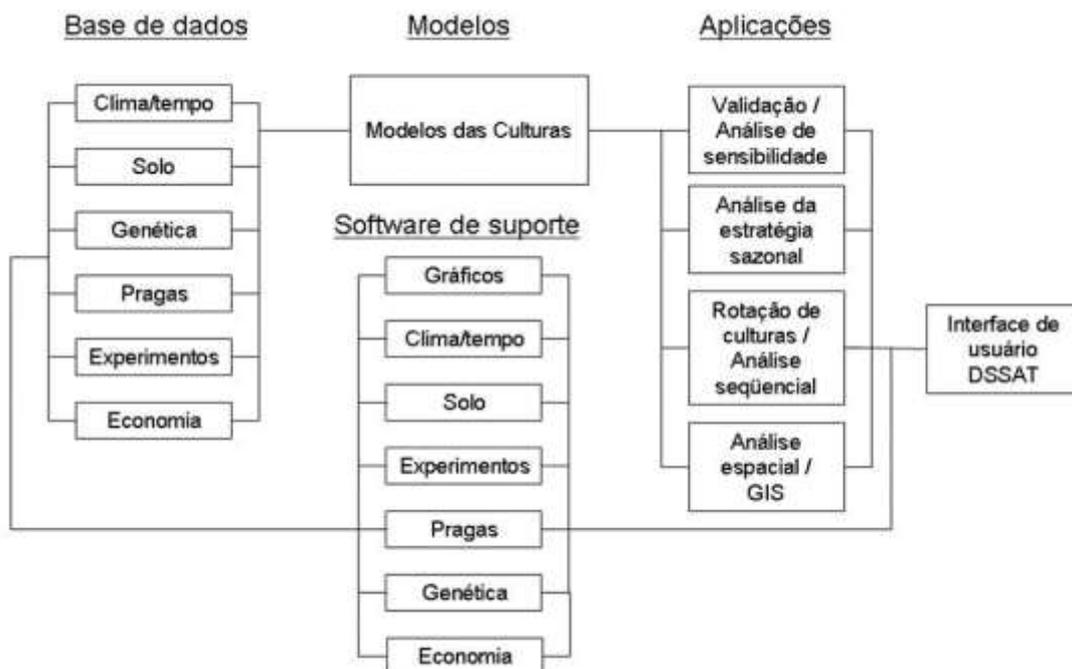


Figura 1 - Representação esquemática da base de dados, aplicação, e componentes de software e seus usos com modelos de culturas para aplicações no sistema DSSAT (adaptado de JONES et al., 2003).

De modo a otimizar o sistema, o DSSAT trabalha com alguns módulos comuns a diversos modelos, como aqueles que gerenciam o clima/tempo, o solo, o balanço de carbono e nitrogênio. No módulo de solo, por exemplo, a maioria das culturas utiliza uma abordagem com perfil unidimensional (horizontalmente homogêneo) e constituído por diferentes camadas verticais (JONES et al, 2003). Atualmente, contudo, ele já conta uma abordagem que simula o balanço hídrico de modo bidimensional, considerando efeitos de cobertura de solo e interseção de sistemas agrícolas. O módulo solo integra informações de quatro módulos componentes: temperatura do solo, carbono e nitrogênio no solo e dinâmica da água. Para estimativa de evapotranspiração, o DSSAT possui o módulo solo-planta-atmosfera, reunindo dados de entrada de solo, planta e atmosfera, computando a interceptação da radiação solar pela cultura e calculando a transpiração máxima da copa em conjunto com a evaporação do solo. No cálculo de evapotranspiração, utilizada para estimar a transpiração foliar, o modelo pode utilizar uma ou duas opções. O padrão é o cálculo pelo método de Priestley & Taylor (1972), que exige somente dados de radiação solar e temperatura, descrito por Jones & Ritchie (1991). O método de Penman-FAO (DOORENBOS & PRUITT, 1977) para o cálculo da evapotranspiração potencial pode ser utilizado opcionalmente para melhor representação em condições áridas e de muito vento, sendo necessário dados de umidade do ar e velocidade do vento na entrada do DSSAT.

O Simulador de Sistemas de Produção Agrícola (APSIM - Agricultural Production Systems Simulator) é utilizado para simulação de sistemas agrícolas e que foi desenvolvido pela Unidade de Pesquisa em Sistemas de Produção Agrícola (APSRU), grupo colaborativo estabelecido sob liderança do CSIRO da Austrália. Desde sua criação, o APSIM foi projetado

como um simulador de sistemas agrícolas que possui como característica combinar a estimativa de rendimento das culturas, de modo acurado e inerente ao sistema de gestão, com a previsão das possíveis consequências, em longo prazo, das práticas agrícolas sobre os recursos do solo (KEATING et al., 2003).

O sistema possui a finalidade de simular processos biofísicos em sistemas agrícolas, principalmente para os que têm interesse econômico e ecológico nas práticas de gestão, tomadas de decisão e análise dos riscos envolvidos na agricultura, como os de clima, solo e manejo. O sistema tem sido largamente utilizado no mundo em aplicações diversas, como por exemplo, no suporte para a tomada de decisão no manejo nutricional de culturas, planejamento de sistemas agrícolas envolvendo sequência de cultivos e seu manejo, avaliação da previsão climática, previsão de safras, gestão e análise da cadeia de suprimentos em atividades do agronegócio, desenvolvimento de diretrizes para gestão de resíduos, avaliação dos riscos nas decisões político-governamentais e como um guia para atividades de investigação e educação (KEATING et al., 2003).

A engenharia de software que suporta o APSIM proporciona uma estrutura flexível para simulações climáticas, de manejo dos solos, no crescimento das culturas em diferentes sistemas agrícolas e para prever as possíveis alterações nos recursos do solo (PROBERT et al., 1998). O modelo também é capaz de simular o crescimento vegetativo, o acúmulo de nitrogênio e de açúcares, o consumo de água, a massa seca e fresca, com diferenciação entre estádios fenológicos. Para cana-de-açúcar também há diferenciação entre cana-planta e cana-soca e simula-se do efeito do manejo da palhada na cultura (Marin et al., 2013; Costa et al., 2014), da irrigação e de doses de nitrogênio (Thorburn et al. 2012). Os fatores avaliados são respostas ao clima, manejo, solo e aos fatores genéticos (KEATING et al.; 1999).

A estrutura do sistema no APSIM é constituída por:

- Um conjunto de módulos que simulam os processos biológicos e físicos em sistemas agrícolas;
- Módulos de gerenciamento que permitem ao usuário especificar as regras de manejo pretendidas e caracterizam o cenário simulado;
- Vários módulos para facilitar a entrada e saída de dados;
- Um mecanismo de simulação que controla todos os processos e mensagens entre os módulos independentes;

Estes elementos podem ser representados por um comando central e seus periféricos, como mostra a Figura 2.

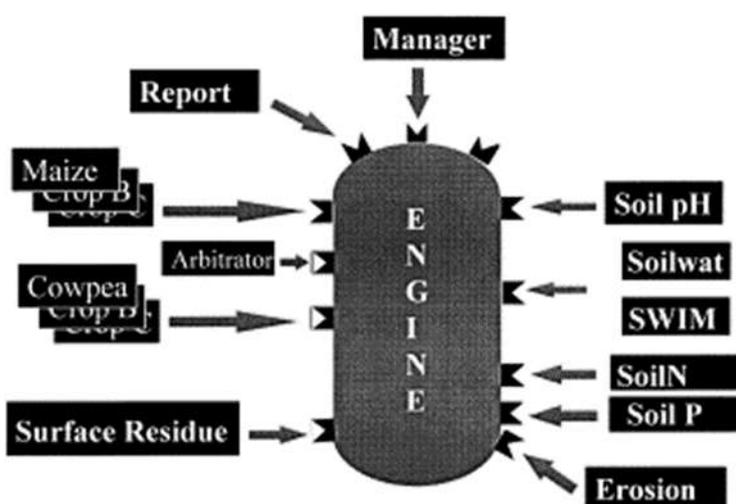


Figura 2. Representação esquemática dos componentes e modo de operação do sistema APSIM. Fonte: Keating et al., (2003).

A abordagem utilizada para seleção dos Modelos de Circulação Global (GCMs) a serem considerados neste projeto baseou-se no protocolo proposto pelo Projeto AgMIP que, por sua vez, vale-se do conceito de *Representative Climate Futures* (RCF) proposto por Whelton et al, 2012)⁷. A justificativa para o uso da abordagem RCFs está em se reduzir o tempo e a capacidade computacional necessária para as simulações e assim produção de uma

⁷ Whetton, P., Hennessy, K., Clarke, J., McInnes, K., Kent, D. 2012. Use of Representative Climate Futures in impact and adaptation assessment. *Climatic Change*. 115(3-4), 433-442.

grande quantidade de dados similares entre si. Além disso, dado que as projeções climáticas de longo prazo carregam incertezas e dado que a finalidade de estudos de vulnerabilidade climática como no presente projeto é a geração de cenários agrícolas futuros, a abordagem RCF é suficientemente robusta para permitir a avaliação de cenários futuros para a cultura da soja para o Estado do Paraná, assim como largamente utilizado na literatura científica⁸ e proposto do termo de referência do presente projeto. Com base nesta abordagem, foram gerados um cenário otimista (com o GCM de menor variação em relação a linha de base), intermediários e pessimistas (com base no GCM de maior variação em relação a linha de base) de produtividade futura da cultura da soja.

A variável de verificação utilizada como referência para a geração dos RCFs foi a temperatura do ar, por se tratar de variável climática que expressa a quantidade de energia no ambiente para a realização dos processos atmosféricos e que está fortemente correlacionada com a chuva e com a própria irradiância solar global. Vale lembrar que a temperatura, radiação solar e chuva são as variáveis de entrada para os modelos agrícolas utilizados neste projeto (DSSAT e APSIM, descritos mais a frente neste documento). Assim, com base na abordagem RCF e tendo a temperatura do ar como variável-critério, a partir de uma análise das base de dados do Paraná foram selecionados três GCMs, representado, como já mencionado, o modelo com a maior variação de temperatura em relação a linha de base, o modelo com a menor variação e, por fim, um modelo utilizando uma variação intermediária

⁸ Diversos estudos científicos publicados em periódicos internacionais de alto impacto valem-se da mesma abordagem usada neste projeto, citando-se como exemplos os seguintes:

- Marin, F.R., J.W. Jones, A. Singels, F. Royce, E.D. Assad, G.O. Pellegrino, and F. Justino. 2013. Climate change impacts on sugarcane attainable yield in Southern Brazil. *Climatic Change* 117: 227–239.
- Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., ... Winter, J. M. (2013). The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170, 166–182. <https://doi.org/10.1016/j.agrformet.2012.09.011>
- Singels, A., Jones, M., Marin, F., Ruane, A., & Thorburn, P. (2014). Predicting Climate Change Impacts on Sugarcane Production at Sites in Australia, Brazil and South Africa Using the Canegro Model. *Sugar Tech*, 16(4). <https://doi.org/10.1007/s12355-013-0274-1>
- Wilby, R.L., S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns. 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. Intergovernmental Panel on Climate Change (IPCC) Supporting Material, available from the Data Distribution Centre of the IPCC Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGCI), http://www.ipcc-data.org/guidelines/dgm_no2_v1_09_2004.pdf.

entre dois anteriores, dentre os 20 disponíveis AGMIP. Além desses três GCMs, incluiu-se um quarto modelo (HadGem2), por solicitação da equipe gestora do por ocasião da reunião de março de 2018. Assim, trabalhou-se com quatro bases de dados climáticos originados dos seguintes GCM's:

5. GCM de variação mínima em relação a LB: MIROC5;
6. GCM de variação média em relação a LB: BNU-ESM;
7. GCM de variação máxima em relação a LB: CanESM2.
8. GCM adicional: HadGem2

Para cada um dos 4 GCMs selecionados, considerou-se 2 níveis de concentração de CO₂ atmosférico, de modo a representar as forçantes radioativas descritas nos *Representative Concentration Pathways* (RCPs) 4.5 e 8.5. Assim, para efeito de simulação, foram assumidos os níveis de concentrações de CO₂ de 526 ppm (RCP 4.5) e 628 ppm (RCP 8.5), constituindo-se dessa forma, para cada local selecionado, 9 bases climáticas (1 atual e 8 futuros), como mostrado na Tabela 1. A Tabela 2 apresenta a variações médias de temperatura e chuva associadas a cada um dos GCM's selecionados.

Tabela 1. Relação de GCM's e RCP's utilizados nas simulações com os modelos agrícolas.

Base climática	Clima Atual	Cenário Futuro/GCM	RCP
1	Linha de base	--	--
2	--	MIROC5	4.5
3	--	MIROC5	8.5
4	--	BNU-ESM	4.5
5	--	BNU-ESM	8.5
6	--	CanESM2	4.5
7	--	CanESM2	8.5
8	--	HadGem2	4.5
9	--	HadGem2	8.5

Tabela 2. Variação em relação a linha de base de temperatura do ar e chuva total para o Estado do Paraná associada a cada CGM selecionados a partir da metodologia proposta por Hudson & Ruane (2013)⁹.

Base climática	Variação de temp. (°C)	Variação de chuva (%)
2	0.9752	8.653%
3	1.6816	3.780%
4	1.7951	-5.737%
5	2.2237	-5.482%
6	2.8691	0.136%
7	3.9698	4.974%
8	2.3108	18.774%
9	3.1621	6.821%

Referente a resolução espacial dos mapas finais, considerando-se que a base de dados climáticos obtidos para a simulação da linha de base (densidade de 142 pontos de dados distribuídos no Estado do Paraná), isso permitiu gerar mapas com resolução de até 1,6 km de resolução¹⁰, portanto com maior nível de informação espacial do que o proposto inicialmente no termo de referência, que era de 5 km. Os métodos de downscaling podem ser de natureza dinâmica (e.g. valendo-se de modelo climáticos regionais) ou estatísticos (i.e. utilizando estatísticas climáticas históricas e informações sobre mudanças climáticas projetadas) ou ainda uma combinação de ambos. Assim, em acordo com os protocolos da principal referência científica mundial sobre estudos no campo da modelagem agrícola aplicada à mudança do clima - o programa AgMIP (www.agmip.org) e conforme descrito em Hudson &

⁹ Hudson, N; Ruane, A. 2013. Guide for Running AgMIP Climate Scenario Generation Tools with R in Windows. Disponível em <http://www.agmip.org/wp-content/uploads/2013/10/Guide-for-Running-AgMIP-Climate-Scenario-Generation-with-R-v2.3.pdf>

¹⁰ Como disponível nas ferramentas de cartografia do software ArcGIS e detalhado em <https://www.esri.com/arcgis-blog/products/product/mapping/choosing-an-appropriate-cell-size-when-interpolating-raster-data/>

Ruane (2013)¹¹ e Ruane et al. (2013)¹², neste projeto utilizou-se da abordagem estatística para a geração de cenários climáticos futuros que, por sua vez, baseiam-se na premissa de que o clima regional é condicionado por dois fatores principais: as condições climáticas de larga escala e as feições fisiografias (topografia, uso do solo, continentalidade/oceanidade), de modo que a informação climática é determinada, a princípio, por um modelo estatístico que relaciona as variáveis climáticas de larga escala (no caso deste projeto, os dados oriundos dos GCMs com as variáveis locais (base de dados que compõem a linha de base de cada uma das 142 estações meteorológicas utilizadas). Assim, os dados dos GCMs alimentam um modelo estatístico que estima o clima regional, considerando a estrutura temporal da série da linha de base. O ponto forte desta abordagem é permitir a obtenção de dados para localidades específicas (pontos ou bacias hidrográficas de pequena escala), como é o caso da abordagem utilizada neste projeto. Ademais, este método já leva em conta a correção de possíveis erros sistemáticos dos modelos, fazendo a geração dos cenários futuros a partir da perturbação estatística originada a partir dos GCM's sobre a série climática da linha de base, e não usando os dados futuros gerados pelos modelos propriamente ditos, o que permite fazer a comparação entre cenários agrícolas futuros e atuais com a devida coerência técnica.

Após a constituição da base climática, foi feita organização da base de solos do Paraná em três tipos de solos conforme classe textural (arenoso, médio, argiloso) também foram estabelecidos foram três tipos de materiais genéticos (ciclo precoce, médio e tardio) representando os diferentes tipos de comprimentos de ciclo da soja no Estado do Paraná. Seguindo a abordagem atualmente utilizada no ZARC, os solos foram agrupados em três categorias quanto à capacidade de retenção de água: arenoso (Tipo 1); textura média (Tipo 2);

¹¹ Hudson, N; Ruane, A. 2013. Guide for Running AgMIP Climate Scenario Generation Tools with R in Windows. Disponível em <http://www.agmip.org/wp-content/uploads/2013/10/Guide-for-Running-AgMIP-Climate-Scenario-Generation-with-R-v2.3.pdf>

¹² Ruane, A.C., J.M.Winter, S.P.McDermid, N.I.Hudson. 2013. AgMIP Climate Data and Scenarios for Integrated Assessment. Chapter 3. Handbook of Climate Change and Agroecosystems.

e argiloso (Tipo 3), conforme disposto na Instrução Normativa nº 2, de 09 de outubro de 2008 (<http://www.agricultura.gov.br/assuntos/riscos-seguro/risco-agropecuario/documentos/INn2de09.10.2008.pdf>).

Os modelos parametrizados foram DSSAT/*CropGro-Soybean* e o modelo APSIM-*Soybean* e todas as simulações foram feitas com ambos os modelos em paralelo. Um terceiro cenário agrícola futuro foi gerado a partir da média dos dois modelos (DSSAT e APSIM). Por fim, para cada cenário climático, cada tipo de solo e cada genótipo foram simuladas 16 datas de semeadura, em passos decendiais (ou seja, a cada 10 dias), entre agosto e dezembro de cada ano da base de dados (30 anos, entre 1980 e 2010).

Tais simulações geraram um banco de dados que foi analisado para a geração das zonas homogêneas de risco conforme descrito na próxima seção deste documento. Além disso, as simulações foram organizadas e os dados apresentados da forma de mapas de produtividade, risco climático e zonas homogêneas e compartilhadas, ainda no mês de Maio de 2019, com o comitê gestor do projeto.

A metodologia de mapeamento atende integralmente aos requisitos para a recomendação das datas de semeadura (assumindo-se ZARC como referência técnica para esse procedimento de recomendação de datas), uma vez que cada município é classificado integralmente em uma única classe de risco levando-se em conta a área de cada faixa de risco dentro do município. Esta técnica de classificação é importante para os casos em que um município é dividido em duas classes de risco diferentes. Nestes casos, tomou-se como classe aquela de menor risco e que representasse pelo menos 20% da área do município.

Como já mencionado, os mapas foram gerados para cada data de semeadura, tipo de solo e material genético utilizando os dois modelos baseados em processos (DSSAT-CropGro e APSIM-Soybean), além da média entre esses dois modelos, utilizando as bases climáticas

oriundas da linha de base e dos oito GCM/RCP's selecionados. Entende-se, com base no que foi proposto no termo de referência do projeto e na descrição metodológica e produtos gerados, que este produto está finalizado.

Em suma, os passos que foram seguidos para a consecução deste produto foram os seguintes:

- Coletar e organizar e depurar as séries de dados meteorológicos em escala diárias, envolvendo as seguintes variáveis: temperatura máxima, temperatura mínima, chuva e irradiância solar global. Dados de umidade do ar e velocidade do vento, se disponíveis, devem ser utilizados para melhoria na qualidade das simulações, mas não são essenciais.
- Definição dos perfis de solos típicos por região para as simulações, sendo os parâmetros hidráulicos (umidade na capacidade de campo, umidade no ponto de murcha permanente, umidade no ponto de saturação, densidade, condutividade hidráulica saturada, número da curva pelo método do servido de hidrologia dos Estados Unidos, e albedo do solo) são os essenciais para processamento, mas dados referentes a textura, as características químicas, ao teor de matéria orgânica pode ser incluídos nos arquivos referentes aos solos;
- Coleta e geração dos cenários climáticos futuros, segundo os protocolos do Projeto AgMIP, como descrito em Hudson & Ruane (2013)¹³. Este banco de dados permite obter cenários oriundos de 20 GCM's sendo possível processar todos eles. Assim como neste projeto, pode-se estabelecer um critério para seleção de GCM representativos, reduzindo-se assim a necessidade de processamento e a massa de dados gerados;

¹³ Hudson, N; Ruane, A. 2013. Guide for Running AgMIP Climate Scenario Generation Tools with R in Windows. Disponível em <http://www.agmip.org/wp-content/uploads/2013/10/Guide-for-Running-AgMIP-Climate-Scenario-Generation-with-R-v2.3.pdf>

- Executar a simulação com os modelos baseados em processos e obter-se a média entre eles para efeito de análise. Caso seja usado um número de modelos suficiente, sugere-se calcular a mediana entre os modelos;
- Representar espacialmente na forma de mapas, para cada decêndio, tipo de solo e genótipo os valores de produtividade média entre os modelos, interpolando-se os valores para os municípios com dados faltantes;
- Representar na forma de tabelas, para cada decêndio, tipo de solo, genótipo e município, os valores da produtividade média entre os modelos.

8.4 NOTA METODOLÓGICA DO PRODUTO 3

Neste produto, os dados de entrada são os valores de produtividade gerados pelos modelos agrícolas descritos no produto 2, e ela se baseia numa simples premissa de cálculo frequência para estratificação do risco agrícola em escala de município. A estratificação municipal do risco foi feita considerando-se cinco níveis de risco (NR, 60%, 65%, 70%, 75% e 80%) aplicado sobre a produtividade média de cada município e, a partir dela fazendo uma análise frequencial, como mostrado na equação 1. Com ela, pode-se estabelecer qual o nível de risco para cada combinação (clima, solo, genética) e quantificar o nível de risco para cada data de semeadura. No presente caso, optou-se por não delimitar uma janela de semeadura fixa por entendermos que isso implicaria em perda de informações e empobrecimento dos resultados- na forma atual, basta ao tomador de decisão definir qual o nível de risco que se pretende admitir e então define-se a janela de semeadura para este nível de risco.

$$F_{i,j,k,l,r,q} = F(Y_{i,j,k,l,r} \geq Y_{c,j,k,l,r} * NR_q) = \frac{(1 - m_{i,j,k,l,r,q})}{n} * 100$$

em que F é a frequência (em porcentagem) da produtividade para um determinada data de semeadura, solo, clima e genética ser maior do que o produto da produtividade média (30 anos e todas as datas de semeadura) (Y_c) para um determinado solo, clima e genética aquela data, pelo nível de risco selecionado (NR), sendo dado em cinco (q) níveis (60%, 65%, 70%, 75% e 80%). i é data da semeadura, j é o município, k é o clima (linha de base ou climas futuros), l é o tipo de solo, r é o material genético. Neste estudo, o cálculo foi feito com a produtividade média aritmética de todos os decêndios, solos e genótipos, mas após uma análise mais detalhada, que não pôde ser incorporada nos resultados deste estudo, sugere-se que o uso futuro desta abordagem seja feita considerando-se a produtividade média ponderada pela curva de semeadura regional, tipos de solos predominantes e genótipos preponderantes.

Este produto foi concluído e os resultados estão disponíveis no mesmo repositório do portal do Sistema Tempocampo, sendo lá apresentados como Tabelas de Risco. Tais tabelas estão organizadas para cada combinação de solo, grupo genético e clima, apresentando em cada uma delas o nível de risco referente a cada data de semeadura, permitindo assim recomendar as datas mais adequadas. Aqui, diferentemente da abordagem tradicional do ZARC de indicar ou não cada uma das datas, apresentadas o nível de risco na forma de classes, permitindo assim ao usuário selecionar quais níveis de risco pretende considerar na sua análise. Caso deseje, há também tabelas (no mesmo formato) apresentando o valor absoluto do risco climático (como opção no seletor *critério* do website). Assim, considerando-se os mapas de produtividade gerados no produto 2 e somando-se aqui os mapas gerados para cada nível de risco (e considerando-se a ramificação em termos de modelos, GCM's, RCP's, solos, genéticas e datas de semeadura) tem-se na pasta compartilhada com o comitê gestor do projeto cerca de 23.000 mapas referentes a todas as

combinações simuladas. Entende-se, com base no que foi proposto no termo de referência do projeto e na descrição metodológica e produtos gerados, que este produto está finalizado.

Em suma, os passos que foram seguidos para a consecução deste produto foram os seguintes:

- Com base nas produtividades simuladas, estimar as frequências de se alcançar a produtividade crítica a ser definida conforme o escopo do estudo, utilizando uma abordagem de cálculo frequencial;
- Representar espacialmente na forma de mapas, para cada decêndio, tipo de solo e genótipo os valores das frequências calculadas, interpolando-se os valores para os municípios com dados faltantes;
- Representar na forma de tabelas, para cada decêndio, tipo de solo, genótipo e município, os valores das frequências calculadas.

8.5 NOTA METODOLÓGICA DO PRODUTO 4

Este produto trata da probabilidade de perda gerada tendo por base os dados de produtividade agrícola gerada nos produtos 2 e 3. A probabilidade de perda foi calculada a partir da seguinte expressão:

$$\text{Probabilidade de Perda (PP)} = \text{Prob}(Y \leq y_g)$$

em que Y é a produtividade agrícola e y_g é a produtividade garantida, ou seja, o nível de cobertura multiplicado pela produtividade estimada. O nível de cobertura variou de 60 a 80%, em múltiplos de 5%.

O ajuste de distribuições é fundamental para mensurar com o maior nível de precisão a probabilidade de perda. Nesse sentido, utilizou-se as seguintes distribuições paramétricas: Normal Assimétrica, Skew-T e a distribuição Odd Log-logística, por apresentarem robustez e flexibilidade na detecção de assimetrias e bimodalidades.

É importante destacar que poder-se-ia utilizar a Distribuição Normal pela facilidade matemática, mas poderia gerar resultados imprecisos, visto que os dados de produtividade podem apresentar certo grau de assimetria. Por isso, optamos pela Normal Assimétrica que capta a assimetria tanto negativa quanto positiva na distribuição dos dados.

Por outro lado, a Skew-T além de captar a assimetria na distribuição dos dados possui caudas “mais pesadas”, acomodando melhor os cálculos da probabilidade no início e no final da distribuição. Por fim, pode ser que os dados apresentem alguma forma de bimodalidade. Nesse caso, a distribuição Odd Log-logística é recomendada para modelar os dados.

O critério utilizado para a escolha do melhor modelo foi o critério de informação de Akaike (AIC). O AIC é um estimador da qualidade relativa de modelos estatísticos para um dado conjunto de dados (Akaike, 1974).

Dada uma coleção de modelos para os dados, o AIC estima a qualidade de cada modelo, em relação a cada um dos outros modelos. Assim, o AIC fornece um meio para a seleção de modelos. Suponha que tenhamos um modelo estatístico de alguns dados. Seja k o número de parâmetros estimados no modelo. Seja \hat{L} o valor máximo da função de verossimilhança para o modelo. Então, o valor AIC do modelo é dado por:

$$AIC = 2k - 2\ln(\hat{L})$$

Dado um conjunto de modelos candidatos, o modelo preferido é aquele com o valor mínimo de AIC. Assim, o AIC recompensa a qualidade do ajuste (avaliado pela função de

verossimilhança), mas também inclui uma penalidade que é uma função crescente do número de parâmetros.

Outro critério que pode ser utilizado nessa situação é chamado de BIC (critério de informação Bayesiano), que penaliza mais fortemente modelos com um maior número de parâmetros do que o AIC tendendo, dessa forma, selecionar modelos com menor número de parâmetros.

Os passos a serem seguidos até a probabilidade de perda são:

- Coletar/Calcular e depurar as séries temporais de produtividade agrícola, por cultura, por unidade espacial (talhão, propriedade agrícola, distrito, município etc.)
- Para cada localidade, separar por:
 - Modelo (Dssat, Apsim etc.);
 - GCM;
 - RCP;
 - Genética;
 - Solo; e,
 - Data de plantio.
- Modelar cada série temporal de produtividade, ajustando as 3 distribuições de probabilidade;
- Escolher o melhor modelo utilizando o critério AIC; e,
- Calcular a probabilidade de perda, de acordo com os 5 níveis de cobertura (60 a 80%).

8.6 NOTA METODOLÓGICA DO PRODUTO 5

O índice de perda foi gerado tendo-se como base as informações de produtividade agrícola geradas no produto 2. O mercado segurador, em geral, utiliza um índice de perda que é base para o cálculo das indenizações. Esse índice depende da produtividade garantida, que é a multiplicação do nível de cobertura pela produtividade esperada, e a produtividade observada no ano t. Os índices de perda foram gerados tendo-se por base a seguinte equação:

$$\text{Índice de Perda (IP)} = (y_g - y_o)/y_g$$

em que y_o é a produtividade observada no ano t. Para o cálculo dos índices de perda utilizou-se as mesmas nove bases climáticas, ou seja, o clima atual (linha de base) e os oito cenários futuros/gcm combinado com os 2 RCPs, 4.5 e 8.5, respectivamente, além dos dois modelos agrícolas Dssat e Apsim.

Os passos a serem seguidos até o índice de perda são:

- Coletar/Calcular e depurar as séries temporais de produtividade agrícola, por cultura, por unidade espacial (talhão, propriedade agrícola, distrito, município etc.)
- Para cada localidade, separar por:
 - Modelo (Dssat, Apsim etc.);
 - GCM;
 - RCP;
 - Genética;
 - Solo; e,
 - Data de plantio.
- Modelar cada série temporal de produtividade utilizando a equação do índice de perda para cada um dos 5 níveis de cobertura (60 a 80%).

8.7 NOTA METODOLÓGICA DO PRODUTO 6

A taxa de prêmio é um parâmetro essencial em qualquer tipo de seguro, seja ele público ou privado. Uma taxa mal ajustada pode levar a desequilíbrios financeiro-atuariais chegando, em alguns casos, a inviabilizar a continuidade da carteira de seguros. Ademais, leva a um problema perverso no mercado segurador, denominado seleção adversa ou anti-seleção.

Uma vez escolhida a abordagem probabilística por meio do critério de informação de Akaike (AIC) ou critério de informação bayesiano (BIC), gera-se a estimativa da probabilidade de perda que é um dos parâmetros da taxa de prêmio e que pode ser calculada através da área sob a curva menor do que a produtividade garantida.

Desta forma, se denominarmos o nível de cobertura por λ , tal que $0 < \lambda < 1$, e a produtividade esperada por y_e , então a probabilidade da produtividade ser menor do que λy_e será calculada através da área sob a densidade. A taxa de prêmio do seguro agrícola de produtividade/custeio é dada por (Goodwin e Ker, 1998):

$$\text{Taxa de prêmio} = \frac{F_Y(\lambda y_e) E_Y[\lambda y_e - (Y | y < \lambda y_e)]}{\lambda y_e}$$

onde E é o operador de esperança e F a distribuição cumulativa da produtividade.

Em alguns casos, a importância segurada é composta por mais de uma variável aleatória. Um exemplo é o seguro de faturamento. Nesse caso, há necessidade de se avaliar o comportamento da produtividade agrícola e o comportamento do preço futuro. Caso o preço seja fixado em dólar, existe ainda uma terceira variável que é a taxa de câmbio. Nota-se que, no caso anterior, a taxa de prêmio é calculada de forma univariada considerando apenas a produtividade agrícola. No caso bivariado, a taxa de prêmio é fixada levando-se em conta a relação entre as duas variáveis, de acordo com a seguinte equação:

$$\text{Taxa de Prêmio} = \frac{F_y(X < x; Y < y | x, y < x_g, y_g) \cdot [x_g \cdot y_g - E(XY | x, y < x_g, y_g)]}{x_g \cdot y_g}$$

onde E é o operador de esperança, F a distribuição cumulativa da produtividade, x_g representa o preço garantido, que em geral baseia-se em uma proporção dos preços futuros negociados em bolsas de mercadorias e futuros; y_g diz respeito à produtividade garantida, obtida através de uma proporção da produtividade média histórica do produtor; x representa o preço ao final da safra do produtor enquanto que y representa a produtividade obtida ao final da safra.

Os passos a serem seguidos até a taxa de prêmio são:

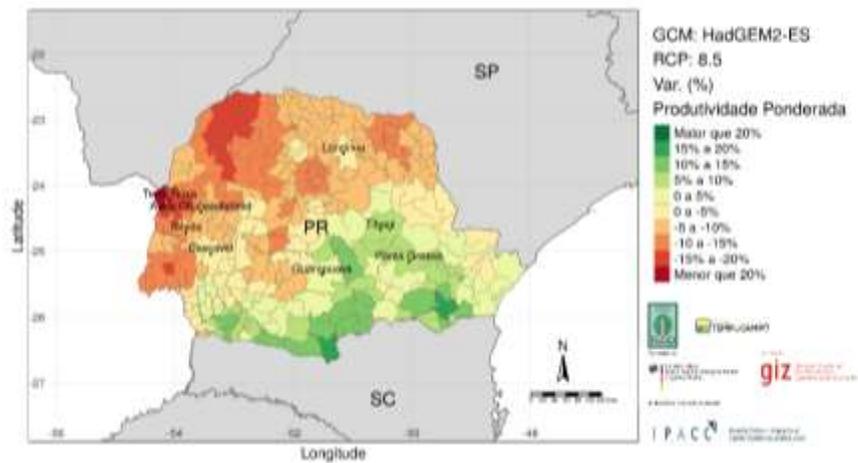
- Coletar/Calcular e depurar as séries temporais de produtividade agrícola, por cultura, por unidade espacial (talhão, propriedade agrícola, distrito, município etc.)
- Para cada localidade, separar por:
 - Modelo (Dssat, Apsim, etc);
 - GCM;
 - RCP;
 - Genética;
 - Solo; e,
- Para cada série de produtividade estimar a produtividade esperada;
- Multiplicar por cada um dos níveis de cobertura;
- Calcular a probabilidade de perda;
- Multiplicar pela perda esperada; e,
- A taxa será encontrada após dividir o resultado pela produtividade garantida.

8.8 NOTA METODOLÓGICA DO PRODUTO 7

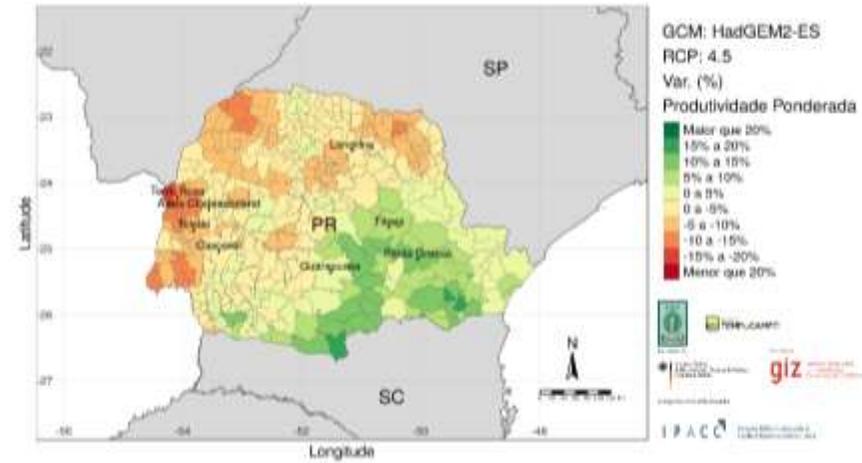
Este produto revela qual foi o “custo” em termos de aumento ou redução da produtividade estimada, probabilidade de perda (PP) e índice de perda (IP), em média, tomando-se por base o cenário “baseline” e os cenários futuros. As entradas utilizadas foram os resultados dos produtos 2, 4 e 5. Em outras palavras, visando a medição do impacto de cada um dos cenários RCP (Representative Concentration Pathways), desenvolvidos pelo Intergovernmental Panel on Climate Change (IPCC), na PP e IP foi calculado o percentual de variação entre cada cenário e o nível atual (linha de base).

Para cada combinação de nível de cobertura e cenários (baseline e RCP) os valores dos índices de perda e probabilidade de perda foram obtidos pela média ponderada de todas as variáveis utilizadas para compor as variações tanto da produtividade quanto dos índices e probabilidades de perda, quais sejam: o tipo de solo (1, 2 e 3), o mês de semeadura (decêndios de agosto a dezembro), material genético (precoce, médio e tardio) e dois modelos Dssat e Apsim.

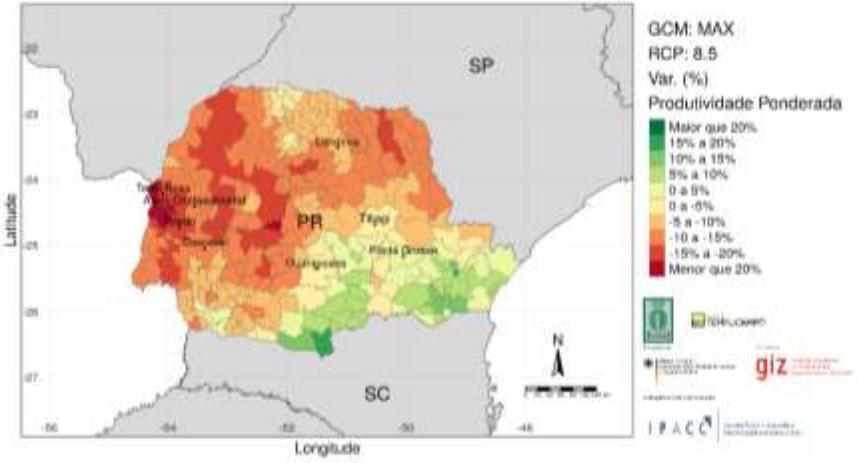
Para cada combinação do nível de cobertura e cenários (baseline e RCP) os valores dos índices de perda e probabilidade de perda foram obtidos por média ponderada dos resultados previamente obtidos. O produto final da ponderação consiste em 88 mapas, em nível de microrregião, detalhando a variação tanto no PP quanto no IP para cada cenário RCP. Oito desses mapas (Fig. 1) foram gerados a partir da diferença entre as produtividades simuladas com base no clima atual (linha de base) e os cenários futuros, permitindo avaliar a variação possível que seria causada pelo clima futuro.



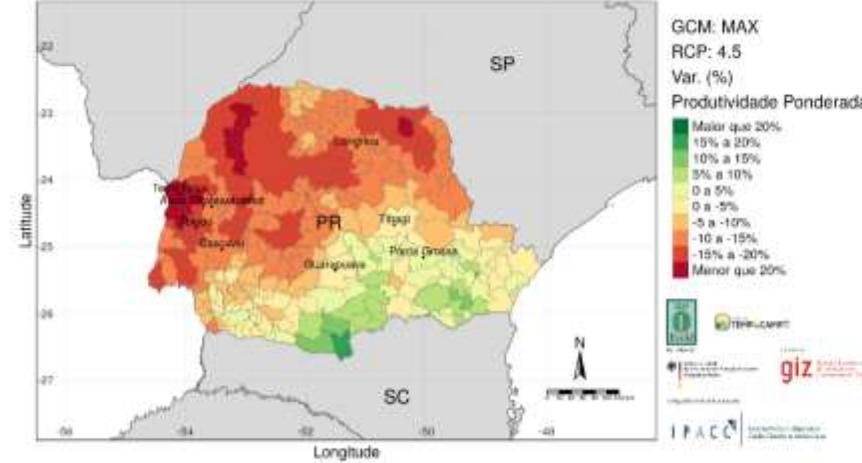
A



B



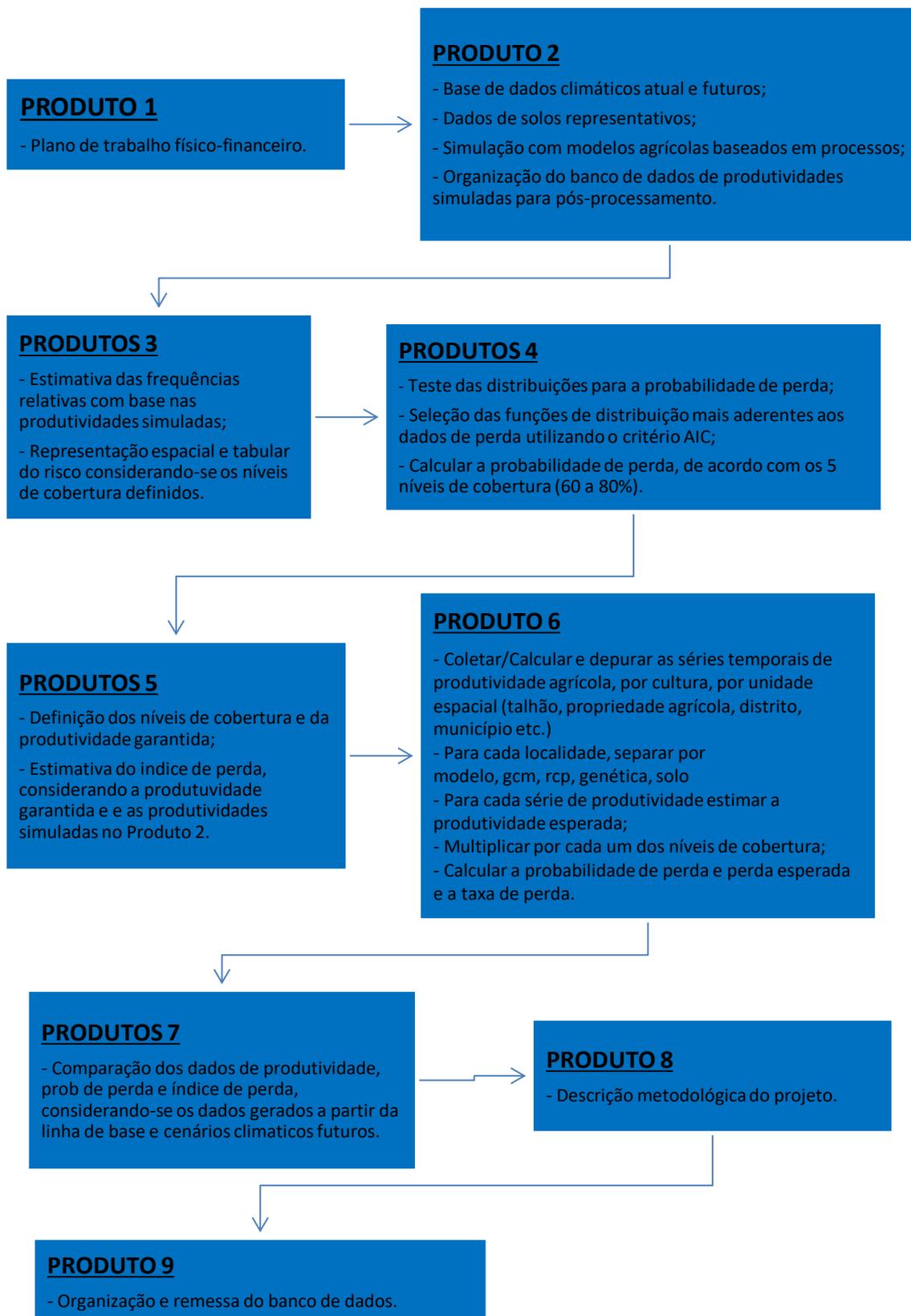
C



D

Figura 1. Variação relativa da produtividade ponderada simulada para oito cenários climáticos futuros, considerando como referência a série climática atual (linha de base).

9. FLUXOGRAMA DO PROJETO



10. ANEXO 1 - CÓDIGOS EM R REFERENTES AO SEÇÃO 5 (PRODUTO 6)

```
rm(list=ls(all=TRUE))
setwd("~/Área de Trabalho/PR/produtividade")
require(ggplot2)
dados=read.csv("nortecentralPR_IPARDES.csv",sep=";",header = TRUE)
(ano=dados[,1])
(y=dados[,2]) ## produtividade em cascavel
par(mfrow=c(2,2))
plot(ts(y,start=c(1980,1),end=c(2014,1)),type="o",lwd=1,main="Cascavel",ylim=c(500,3900))
hist(y,main="Cascavel",nclass=15)

#####

##### yt = alfa+ beta*t + epsilont

#####

y1=ts(y,start=c(1980,1),end=c(2014,1))
ano=dados[,1]
cor(y1,ano) # coeficiente de correlacao linear
m1= lm(y1~ano);m1 # MMQ
summary(m1)

par(mfrow=c(1,1))
plot(ts(y1,start=c(1980,1),end=c(2014,1)),type="o",lwd=1,main="Cascavel",ylim=c(500,3900))
abline(m1,col=20)
names(m1)
```

```

res= m1$residuals
ypred= m1$fitted.values
par(mfrow=c(2,2))
plot(m1)

shapiro.test(res)
shapiro.test(rstudent(m1))

m2= lm(y1~ano+l(ano^2));m2
summary(m2)
predict(m2)
lines(ano,predict(m2),col=2)
names(m2)
res= m2$residuals
ypred= m2$fitted.values

ynew= ypred[35]*(1+res/ypred)
ynew2=res+ ypred[35]
ynew=ts(ynew,start=c(1980,1),end=c(2014,1))
ynew2=ts(ynew2,start=c(1980,1),end=c(2014,1))

par(mfrow=c(1,1))
plot(y1,type="o",ylim=c(700,4200),main="Cascavel ",ylab="Productivity (Kg/ha)",xlab="Years")
lines(ynew,col="red",type="o",lwd=1,lty=2)
lines(ynew2,col="green ",type="o",lwd=1,lty=2)
legend('bottomleft', c("Original","Corrigida Metodo Multiplicativo","Corrigida Metodo Aditivo"),
      col=c('1','red',"green"),lty=c(1,2,2),lwd=c(1,1,1))

y=ynew

```

```

y=y1
plot(y)
hist(y,nclass=15)

####Estatistica DEscritiva

require(moments)

cbind(n=length(y),media=mean(y),mediana=median(y),desviopadrao=sd(y),assimetria=skewness(y),c
urtose= kurtosis(y))

cbind(max=max(y),min=min(y))# OBS A mediana > media entao tem dist. assimetrica a esquerda

# Verificacao da autocorrelaçao

library(lmtest)

lmy=lm(y~ano)

#pvalor>0.05 entao nao rejeito H0

Box.test(y,type=c("Ljung-Box")) #pvalor>0.05 nao rejeito H0
Box.test(y*y,type=c("Ljung-Box")) #pvalor>0.05 nao rejeito H0

par(mfrow=c(2,2))

acf(y)
pacf(y)
acf(y*y)
pacf(y*y)

require(tseries)

m2= arima(y,order=c(1,0,0));m2
tsdiag(m2)

Box.test(m2$residuals,type=c("Ljung-Box"))

r=m2$residuals

```

```

plot(series)

#####

Ye=Produtividade esperada
Ye2=mean(y[30:35]);Ye2

xt<- c(0.6, 0.65, 0.7, 0.75, 0.8) # nivel de cobertura 70%, 75%,80,85 e 90%
Yg =seq(1,5)
for(i in 1:5) { Yg[i] = Ye2*xt[i]}
Yg # Produtividade garantida

#####

pdfnorm=function(par,x){
  mu = par[1]
  sigma = par[2]
  (1/(sigma*sqrt(2*pi)))*exp(-0.5*((x - mu)/sigma)^2)
}

cdfnorm=function(par,x){
  mu = par[1]
  sigma = par[2]
  pnorm(x,mean=mu,sd=sigma)
}

require(AdequacyModel)

aj=goodness.fit(pdf=pdfnorm,cdf=cdfnorm, starts=c(mean(y),sd(y)),data=y, method="Nelder-Mead",
domain=c(0,Inf));aj

aj$mle

result=goodness.fit(pdf=pdfnorm,cdf=cdfnorm, starts=c(aj$mle[1],aj$mle[2]),data=y,
method="Nelder-Mead", domain=c(0,Inf));result

```

```
#####

x = seq(min(y)-100,max(y)+150,0.1)

fnorm =pdfnorm(result$mle,x)

dev.off()

hist(y,freq=F,main="Cascavel",ylim=c(0,0.0015))

hist(y,freq=F,main="",ylim=c(0,0.0015),xlabel="")

#lines(density(y,bw="nrd0",kernel="gaussian"),col=6,lwd = 2,lty=2) #kernel

lines(x,fnorm , col = '1',lwd =2,lty=1) #Normal

####

plot(ecdf(y),main="Cascavel ")#ylabel="Cumulative Dist. Function",

lines(x,cdfnorm(result$mle,x),col=2,lwd = 1,lty=1) #Dist. Acumulada Normal

##cdfNormal =fcao probal. cumulativa

cbind("Prob. Perda(60%)" = cdfnorm(result$mle,Yg[1]),"Prob.
Perda(65%)"=cdfnorm(result$mle,Yg[2]),"Prob. Perda(70%)"=cdfnorm(result$mle,Yg[3]),

      "Prob. Perda(75%)"=cdfnorm(result$mle,Yg[4]),"Prob. Perda(80%)"=cdfnorm(result$mle,Yg[5])) #
Probabilidade de Perda

#####

pdf_sn=function(par,x){

  mu = par[1]

  sigma = par[2]

  gama=par[3]

  z = (x - mu)/sigma

  f=dnorm(z,0,1)

  F=pnorm(gama*z)

  (2/sigma)*f*F

}

require(sn)
```

```

cdf_sn=function(par,x){
  mu = par[1]
  sigma = par[2]
  gama=par[3]
  z = (x - mu)/sigma
  pnorm(z)-2*TOwen(z,gama)
}

aj=goodness.fit(pdf=pdf_sn,cdf=cdf_sn, starts=c(mean(y),sd(y),0.2),data=y, method="L",
domain=c(0,Inf));aj

aj$mle

result=aj

result=goodness.fit(pdf=pdf_sn,cdf=cdf_sn, starts=c(aj$mle[1],aj$mle[2],aj$mle[3]),data=y,
method="Nelder-Mead", domain=c(0,Inf));result

cbind(result$mle[1],result$Erro[1],result$mle[2],result$Erro[2],result$mle[3],result$Erro[3])

#####

x = seq(min(y)-100,max(y)+150,0.1)

fsn =pdf_sn(result$mle,x)

lines(x,fsn, col = '3', lwd = 2,lty=3) #Skew-Normal

####Distribuicao acumulada

lines(x,cdf_sn(result$mle,x),col=3,lwd = 2,lty=2) #Dist. Acumulada Normal

cbind("Prob. Perda(70%)" = cdf_sn(result$mle,Yg[1]),"Prob.
Perda(75%)"=cdf_sn(result$mle,Yg[2]),"Prob. Perda(80%)"=cdf_sn(result$mle,Yg[3]),

  "Prob. Perda(85%)"=cdf_sn(result$mle,Yg[4]),"Prob. Perda(90%)"=cdf_sn(result$mle,Yg[5]))

#####

pdf_olln=function(par,x){
  mu = par[1]
  sigma = par[2]
  alpha = par[3]
  z = (x - mu)/sigma

```

```

f = (1/(sigma*sqrt(2*pi)))*exp(-0.5*(z)**2)
F = pnorm(z)
num= alpha*f*( F*(1-F))^(alpha-1)
den = (F^(alpha) + (1-F)^(alpha))^2
num/den
}

cdf_olln=function(par,x){
  mu = par[1]
  sigma = par[2]
  alpha = par[3]
  z = (x - mu)/sigma
  G = pnorm(z)
  (G^(alpha))/((G^(alpha))+(1-G)^(alpha))
}

aj=goodness.fit(pdf=pdf_olln,cdf=cdf_olln, start=c(mean(y),sd(y),2),data=y, method="Nelder-Mead",
domain=c(0,Inf))

aj$mle

result=goodness.fit(pdf=pdf_olln,cdf=cdf_olln, starts=c(aj$mle[1],aj$mle[2],aj$mle[3]),data=y,
method="L-BFGS-B", domain=c(0,Inf));result

#cbind(result$mle[1],result$Erro[1],result$mle[2],result$Erro[2],result$mle[3],result$Erro[3])

#####

x = seq(min(y)-100,max(y)+150,0.1)
folln = pdf_olln(result$mle,x)
lines(x,folln , col = '4', lwd = 2,lty=1)
lines(x,cdf_olln(result$mle,x),col=4,lwd = 2,lty=1) #Dist. Acumulada Normal

#####

```

```

cbind("Prob. Perda(70%" = cdf_olln(result$mle,Yg[1]),"Prob.
Perda(75%)=cdf_olln(result$mle,Yg[2]),"Prob. Perda(80%)=cdf_olln(result$mle,Yg[3]),
"Prob. Perda(85%)=cdf_olln(result$mle,Yg[4]),"Prob. Perda(90%)=cdf_olln(result$mle,Yg[5])) #
Probabilidade de Perda

```

```
#####
```

```

require(ExtDist)
pdfbeta=function(par,x){
  par1=par[1]
  par2=par[2]
  par3=par[3]
  par4=par[4]
  dBeta_ab(x, par1,par2,par3,par4)
}

```

```

cdfbeta=function(par,x){
  par1=par[1]
  par2=par[2]
  par3=par[3]
  par4=par[4]
  pBeta_ab(x,par1,par2,par3,par4)
}

```

```

require(AdequacyModel)
pp=eBeta_ab(y);pp
aj=goodness.fit(pdf=pdfbeta,cdf=cdfbeta, starts = c(pp$shape1,pp$shape2,pp$a,pp$b),data=y,
method="N", domain=c(0,Inf));aj
aj$mle
result=aj
result=goodness.fit(pdf=pdfbeta,cdf=cdfbeta,
starts=c(aj$mle[1],aj$mle[2],aj$mle[3],aj$mle[4]),data=y, method="N", domain=c(0,Inf));result

```

```

aj=goodness.fit(pdf=pdfbeta,cdf=cdfbeta, starts=c(1.65,0.78,2362.50,3661.33),data=y, method="S",
domain=c(0,Inf));aj

result=aj

require(EnvStats)

#fb= dBeta_ab(y, pp$shape1,pp$shape2,pp$a,pp$b)#params =
list(pp$shape1,pp$shape2,pp$a,pp$b)

#facumb= pBeta_ab(y, pp$shape1,pp$shape2,pp$a,pp$b);facumb

t = seq(min(y)-100,max(y)+150,0.1)

length(t)

fb =pdfbeta(result$mle,t)

length(fb)

####

plot(ecdf(y),main="Castro")#ylabel="Cumulative Dist. Function",
lines(x,cdfbeta(result$mle,x),col="Sienna",lwd =2,lty=2) #Dist. Acumulada Normal
legend("topleft",lty=c(1,2,1,2),lwd=c(1,2,2,,2), c("N","SN","OLL-N","Beta"),col=c(2,3,4,"Sienna"))

cbind("Prob. Perda(70%)" = cdfbeta(result$mle,Yg[1]),"Prob.
Perda(75%)"=cdfbeta(result$mle,Yg[2]),"Prob. Perda(80%)"=cdfbeta(result$mle,Yg[3]),

      "Prob. Perda(85%)"=cdfbeta(result$mle,Yg[4]),"Prob. Perda(90%)"=cdfbeta(result$mle,Yg[5])) #
Probabilidade de Perda

#####3

require(skewt)

pdf_skewt=function(par,x){

  mu = par[1]

  sigma = par[2]

  gama=par[3]

```

```

df= par[4]
z = (x - mu)/sigma
dskt(z,df,gama)
}
require(sn)
cdf_skewt=function(par,x){
mu = par[1]
sigma = par[2]
gama=par[3]
df= par[4]
z = (x - mu)/sigma
pskt(z,df,gama)
}
aj=goodness.fit(pdf=pdf_skewt,cdf=cdf_skewt, starts=c(mean(y),sd(y),gama=2,df=2),data=y,
method="Nelder-Mead", domain=c(0,Inf));aj
aj$mle
result=aj
result=goodness.fit(pdf=pdf_sn,cdf=cdf_sn, starts=(aj$mle[1],aj$mle[2],aj$mle[3],aj$mle[4]),data=y,
method="Nelder-Mead", domain=c(0,Inf));result
cbind(result$mle[1],result$Erro[1],result$mle[2],result$Erro[2],result$mle[3],result$Erro[3])

#####
x = seq(min(y)-100,max(y)+150,0.1)
fst =pdf_skewt(result$mle,x)
hist(y,prob=T,ylim=c(0,0.0025))
lines(x,fst,col="2",lty=4,lwd=2)
legend("topleft",lty=c(1,3,1,4),lwd=c(2,2,2,2),cex=0.8, c("N","SN","OLL-N","RS"),col=c(1,3,4,2))
lines(x,cdf_skewt(result$mle,x),col=3,lwd = 2,lty=2) #Dist. Acumulada Normal

```


11. ANEXO RESUMO DAS Etapas da metodologia de precificação (SEÇÃO 5 – PRODUTO 6)

I) Tratamento da série de produtividade quanto a tendência e autocorrelação.

II) Seleção da distribuição de probabilidade (Normal, Skew-Normal, OLLN e Skew-t) que melhor se ajusta aos dados, através do procedimento descrito em 3.

III) Cálculo da distribuição acumulada de probabilidade, ou também chamada de probabilidade de perda, escolhida para o valor de Y_g (produtividade garantida pelo seguro).

IV) Cálculo da Expectativa de Perda, através da fórmula:

$$E(perda) = E(\max(y_g - Y); 0) = (\lambda y_e - E(Y|Y < \lambda y_e)).$$

Observe que é necessário calcular a probabilidade condicional acumulativa da produtividade dado que ela é menor do que a produtividade garantida.

V) Cálculo da Taxa de Prêmio, através da fórmula:

$$TP = \frac{F_y(\lambda y_e) E_y[\lambda y_e - (Y|y < \lambda y_e)]}{\lambda y_e}$$

Em que E é o operador de esperança e F a distribuição cumulativa da produtividade (probabilidade de perda), λ representa o nível de cobertura, tal que $0 < \lambda < 1$, e y^e é a produtividade esperada. O numerador é formado por dois termos, a probabilidade de perda e a perda média dada pelo operador de esperança, ou seja, a média das perdas considerando toda a série temporal. A multiplicação da probabilidade de perda pela perda esperada, dividida pela produtividade garantida resulta na taxa de prêmio para cada nível de cobertura.

11.1 TRATAMENTO PRELIMINAR DOS DADOS

Para a precificação do risco de produtividade municipal é necessário fazer um tratamento preliminar nos dados com relação à tendência crescente da produtividade ao longo dos anos, de acordo com a metodologia proposta por Gallagher (1987).

Uma regressão linear é ajustada aos dados de produtividade tendo o tempo como a variável explicativa, então o último rendimento observado é corrigido utilizando-se os resíduos do modelo para cada ano. Os rendimentos corrigidos pela tendência são definidos pela seguinte equação:

$$\tilde{y}_t = \widehat{y}_{2016} \left(1 + \frac{\hat{e}_t}{\hat{y}_t} \right)$$

onde \tilde{y}_t , \hat{y}_t e \hat{e}_t são, respectivamente, o rendimento corrigido, o rendimento ajustado e o residual do ano t , y_{2016} é o rendimento ajustado para 2016.

No próximo passo foi utilizado o teste de Ljung and Box(1978) para tratar a autocorrelação serial da produtividade. Ljung e Box propuseram uma pequena alteração no teste de Box - Pierce, na qual torna o teste mais generalizado. Foi observado na literatura que o teste de Ljung - Box apresenta melhores resultados do que o teste de Box - Pierce.

As hipóteses do teste são as mesmas:

$$\begin{cases} H_0: Os\ resíduos\ são\ i.\ i.\ d.; \\ H_1: Os\ resíduos\ não\ são\ i.\ i.\ d.. \end{cases}$$

Calculamos as estimativas de autocorrelações por

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{a}_t \widehat{a_{t-k}}}{\sum_{t=1}^n \hat{a}_t^2}$$

É possível mostrar que $\hat{r}_k \sim N\left(0, \frac{1}{n}\right)$. Se o modelo for apropriado, a estatística do teste

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

terá aproximadamente uma distribuição χ^2 com $(K - p - q)$ graus de liberdade, onde K é o número de defasagens tomada na função de autocorrelação, p e q são as ordens do modelo ajustado. Portanto, rejeitamos a hipótese nula se $Q > \chi_{1-\alpha, k-p-q}^2$ com um nível de significância α .

Para a modelagem da produtividade testa-se modelos paramétricos simétricos e assimétricos, que incorporam a unimodalidade e bimodalidade dos dados. Todas estas análises foram realizadas no software R utilizando os pacotes “Optim” e “AdequacyModel”.

Num terceiro momento foi utilizado o teste de Breusch and Pagan (1979) para tratar a heterogeneidade de variâncias. Baseado no teste multiplicador de Lagrange, o teste de Breusch-Pagan é bastante utilizado para testar a hipótese nula de que as variâncias dos erros são iguais (homoscedasticidade) versus a hipótese alternativa de que as variâncias dos erros são uma função multiplicativa de uma ou mais variáveis, sendo que esta(s) variável(eis) pode(m) pertencer ou não ao modelo em questão.

Inicialmente, ajusta-se um modelo de regressão linear (simples ou múltiplo) e obtem-se os resíduos $e = (e_1, \dots, e_n)$ e os valores ajustados $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$. Em seguida, consideram-se os resíduos ao quadrado e estes são padronizados de modo que a média do vetor de resíduos padronizados u seja igual a 1. Esta padronização é feita dividindo cada resíduo ao quadrado pela SQE/n em que SQE é a Soma de Quadrados dos Resíduos do modelo ajustado e n é o número de observações. Desta forma, temos que cada resíduo padronizado é dado por

$$u_i = \frac{e_i^2}{SQE/n}, i = 1, \dots, n, \text{ em que } SQE = \sum_{i=1}^n e_i^2.$$

Por fim, faz-se a regressão entre $u = (u_1, \dots, u_n)$ (variável resposta) e o vetor \hat{y} (variável explicativa) e obtem-se a estatística do teste χ_{BP}^2 calculando-se a Soma de Quadrados da Regressão de u sobre \hat{y} e dividindo-se o valor encontrado por 2. Sob a hipótese nula, esta estatística tem distribuição qui-quadrada com 1 grau de liberdade. Para a modelagem da produtividade foram testados modelos paramétricos simétricos e assimétricos, que incorporam a unimodalidade e bimodalidade dos dados. Todas estas análises foram realizadas no software R utilizando os pacotes “Optim” e “AdequacyModel”.

Testa-se também se os dados seguem uma distribuição Normal, por meio do teste de Kolmogorov-Smirnov (KS). O teste de Kolmogorov - Smirnov pode ser utilizado para avaliar as hipóteses: H_0 : os dados seguem uma distribuição Normal, contra H_1 : os dados não seguem uma distribuição Normal.

Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada dos dados, no caso a Normal, e a função de distribuição empírica dos dados. Como critério, comparamos esta diferença com um valor crítico, para um dado nível de significância. O valor-p resultante do no programa R pode ser interpretado como a medida do grau de concordância entre os dados e a hipótese nula. Quanto menor for o valor-p, menor é a consistência entre os dados e a hipótese nula. Então, a regra de decisão é rejeita-se H_0 , se o valor-p for menor ou igual a um certo valor arbitrado pelo analista.

Questões relativas à melhoria do ponto 6 ("Aplicação da metodologia para o caso da soja no Estado do Paraná: comparativo do Proagro") do Produto 6

11.2 AJUSTE E TESTES DAS DISTRIBUIÇÕES DE PROBABILIDADE

As distribuições de probabilidade consideradas para a análise foram:

Skew-Normal

$$f(x) = \frac{2}{\omega} \phi\left(\frac{x - \xi}{\omega}\right) \Phi\left(\alpha \left(\frac{x - \xi}{\omega}\right)\right)$$

Onde ϕ representa a densidade de probabilidade da distribuição normal e Φ a sua densidade acumulada.

Odd log-logística normal (OLLN)

$$F(x; \mu, \sigma, \alpha) = \int_0^{\frac{\Phi(x; \mu, \sigma)}{\bar{\Phi}(x; \mu, \sigma)}} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2} dt = \frac{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right)}{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha}$$

em que $\bar{\Phi}(x; \mu, \sigma) = 1 - \Phi(x; \mu, \sigma)$. A fdp é dada por:

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right) \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}}{\sigma \left\{ \Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha \right\}^2}$$

respectivamente. Note que $\alpha > 0$ é um parâmetro de forma.

Skew-t

$$f_{ST}(x; \mu, \sigma, \lambda, p, q) = \frac{p}{2v\sigma q^{1/p} B\left(\frac{1}{p}, q\right) \left(\frac{|x-\mu+m|^p}{q(v\sigma)^p (\lambda \text{sign}(x-\mu+m) + 1)^p} + 1\right)^{\frac{1}{p}+q}}$$

onde B é a função beta, μ é o parâmetro de localização, $\sigma > 0$ é o parâmetro de escala, $-1 < \lambda < 1$ é o parâmetro de assimetria, e $p > 0$ e $q > 0$ são os parâmetros que controlam a curtose. Note que m e v não são parâmetros, mas funções dos outros parâmetros que são usados aqui para mudar a escala ou deslocar a distribuição. Na parametrização original, tem-se que:

$$m = \frac{2\nu\sigma\lambda q^{\frac{1}{p}} B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)}$$

e

$$v = \frac{q^{-\frac{1}{p}}}{\sqrt{(3\lambda^2 + 1) \frac{B\left(\frac{3}{p}, q - \frac{2}{p}\right)}{B\left(\frac{1}{p}, q\right)} - 4\lambda^2 \frac{B\left(\frac{2}{p}, q - \frac{1}{p}\right)^2}{B\left(\frac{1}{p}, q\right)^2}}}$$

Para escolha da melhor distribuição paramétrica foram utilizados os critérios AIC de Akaike (1974) e BIC, Schwarz et al (1978). O Critério de Informação de Akaike (AIC) é definido como

$$AIC_p = -2\log(L_p) + 2[(p + 1) + 1]$$

em que L_p é a função de máxima verossimilhança do modelo e p é o número de variáveis explicativas consideradas no modelo.

O Critério de Informação Bayesiano (BIC) é definido como

$$BIC_p = -2\log(L_p) + [(p + 1) + 1]\log(n).$$

Tanto o AIC quanto o BIC aumentam conforme a *SQE* aumenta. Além disso, ambos critérios penalizam modelos com muitas variáveis sendo que valores menores de AIC e BIC são preferíveis.

Como modelos com mais variáveis tendem a produzir menor SQE mas usam mais parâmetros, a melhor escolha é balancear o ajuste com a quantidade de variáveis.

Além disso, pode-se utilizar estatísticas modificadas de Anderson-Darling (W^*) e Cramer-von Mises (A^*), proposto por Lin et al. (2008) e Pakyari and Balakrishnan (2012), respectivamente.

Seja x_1, x_2, \dots, x_n uma amostra aleatória e suponha que um provável candidato para a f.d.a. dos dados seja $F(x)$, o teste de hipóteses para verificar a adequação da distribuição $F(x)$ aos dados é

$$\begin{cases} H_0: a \text{ amostra tem distribuição } F(x) \\ H_1: a \text{ amostra não tem distribuição } F(x) \end{cases} \quad (1)$$

Anderson e Darling (1952, 1954) propuseram a seguinte estatística para testar (1):

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]}{F(x)(1 - F(x))} dF(x)$$

sendo $F_n(x)$ a função de distribuição acumulada empírica definida por

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq x\}} = \begin{cases} 0, & \text{se } x < x_{(1)}, \\ \frac{k}{n}, & \text{se } x_{(k)} \leq x < x_{(k+1)}, \\ 1, & \text{se } x > x_{(n)}, \end{cases}$$

sendo $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, as estatísticas de ordem da amostra aleatória e $\mathbb{1}_{\{x_i \leq x\}}$ a função indicadora que vale 1 se $x_i \leq x$ e 0 se $x_i > x$, $i = 1, \dots, n$.

A estatística A^2 pode ser representada numa forma equivalente como

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n \left[(2i - 1) \ln(F(x_{(i)})) + (2(n - i) + 1) \ln(1 - F(x_{(i)})) \right]$$

A transformação $F(x_{(i)})$ leva $x_{(i)}$ em $U_{(i)}$, sendo $U_{(1)}, \dots, U_{(n)}$ uma amostra de tamanho n com distribuição uniforme em $(0,1)$. Logo,

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n \left[(2i - 1) \ln(U_{(i)}) + (2(n - i) + 1) \ln(1 - U_{(i)}) \right]$$

Para calcular o valor da estatística A^2 , devemos seguir os passos abaixo:

1. Ordene os valores da amostra: $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$;
2. Quando necessário, estime os parâmetros da distribuição de interesse;

3. Calcule $U_i = F(x_{(i)})$ e calcule o valor da estatística de Anderson Darling:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1)(\ln(U_i) + \ln(1 - U_{n+1-i}))]$$

4. Para cada uma das distribuições calcule, se for o caso, o valor da estatística modificada de acordo com as tabelas dadas para cada uma delas.

Similarmente se obtém a estatística de Cramér-von-Mises:

$$W^2 = \sum_{i=1}^n \left(U_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}$$

11.3 CÁLCULO DA PROBABILIDADE DE PERDA

Para o cálculo da probabilidade de perda é feito o cálculo da distribuição acumulada no valor de Y_g (produtividade garantida).

Depois de selecionada a função de distribuição que melhor se ajustou os dados de acordo com o foi calculada densidade acumulada de cada distribuição utilizando-se o software R. Esta fase da modelagem tem como input os níveis de cobertura desejados e como output os valores de probabilidade de perda.

11.4 CÁLCULO DA TAXA DE PRÊMIO

O cálculo da taxa de prêmio é realizado pela equação

$$TP = \frac{F_y(\lambda y^e) E_y[\lambda y^e - (Y|y < \lambda y^e)]}{\lambda y^e}$$

em que E é o operador de esperança e F a distribuição cumulativa da produtividade (probabilidade de perda), λ representa o nível de cobertura, tal que $0 < \lambda < 1$, e y^e é a produtividade esperada. O numerador é formado por dois termos, a probabilidade de perda e a perda média dada pelo operador de esperança, ou seja, a média das perdas considerando toda a série temporal. A multiplicação da probabilidade de perda pela perda esperada, dividida pela produtividade garantida resulta na taxa de prêmio para cada nível de cobertura.