



Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

JOÃO EUDES MIQUÉIAS MACIEL TORRES

AMOSTRAGEM DE ÁREA E APLICAÇÕES EM AGROPECUÁRIA

Recife

2018

João Eudes Miquéias Maciel Torres

AMOSTRAGEM DE ÁREA E APLICAÇÕES EM AGROPECUÁRIA

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de mestre em Estatística.

Área de Concentração: Ciência Exatas e da Terra

Orientador: Cristiano Ferraz

Coorientador: André Leite Wanderley

Recife

2018

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

T693a Torres, João Eudes Miquéias Maciel
 Amostragem de área e aplicações em agropecuária / João Eudes Miquéias
 Maciel Torres. – 2018.
 99 f.: il., fig., tab.

 Orientador: Cristiano Ferraz.
 Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN,
 Estatística, Recife, 2018.
 Inclui referências e apêndice.

 1. Estatística aplicada. 2. Amostragem. I. Ferraz, Cristiano (orientador). II.
 Título.

 310 CDD (23. ed.) UFPE- MEI 2018-115

JOÃO EUDES MIQUEIAS MACIEL TORRES

AMOSTRAGEM DE ÁREA E APLICAÇÕES EM AGROPECUÁRIA

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 31 de julho de 2018.

BANCA EXAMINADORA

Prof.^o Cristiano Ferraz
UFPE

Prof.^o Carlos Feitosa Luna
FIOCRUZ

Prof.^o Raydonal Ospina Martínez
UFPE

À minha família, por me apoiar em todo momento e por ser um porto seguro. Mãe, seu amor e cuidado me deram forças pra continuar. Pai, meu maior exemplo.

AGRADECIMENTOS

Em primeiro lugar, a Deus. Reconheço o seu amor infinito em todas as coisas que Ele tem me proporcionado, em especial, a conclusão do Mestrado. Por muitas vezes me deu alívio e em seus braços pude descansar. A Ele toda honra, glória, majestade e poder, entre os séculos dos séculos. Amém.

À minha família, meu porto seguro, meu lugar de retorno. Tive em vocês todo apoio e suporte que precisei, infalivelmente. Agradeço a minha mãe, Judite, por todo amor e carinho, orando sempre por mim. Agradeço a meu pai, José Silva, por ser um exemplo de pai e um amigo. Agradeço as minhas irmãs, Donina e Waléria, por todo encorajamento e pelos bons conselhos. Mesmo longe, me senti abraçado por vocês todos os dias.

A todos os amigos com quem tive o prazer de dividir grande parte dos meus dias de pós-graduação, e que de agora em diante farão parte da minha vida. Em especial, a Abel Borges e Raíza Oliveira, Bruna Palm, César Diogo, José Jairo, Adenice Ferreira e Anny Kerollayny, por de muitas formas e em muitos momentos me ajudarem a concluir esse trabalho.

Aos meus amigos Álvaro Mateus e Assíria Dantas, por todas as suas orações e amizade. À Flávia e Rute Pereira, por seus valiosos conselhos, apoio e orações; são como pais para mim. Ao meu amigo e irmão José Kleyton, por toda amizade, companheirismo e auxílio.

Aos meus orientadores, Cristiano Ferraz e André Leite, por toda ajuda, apoio e compreensão, por me ensinar e pressionar sempre que necessário, por todos ensinamentos não só no âmbito acadêmico.

À banca examinadora, pelas valiosas sugestões que, sem dúvida, enriqueceram a qualidade desse trabalho.

Aos professores do Programa de Pós-Graduação em Estatística (PPGE) da UFPE, por proporcionarem uma formação com excelência. De muitos tenho aprendido desde a graduação. À secretaria do PPGE pelo bom atendimento, solução de dúvidas e conselhos fornecidos que contribuíram para a conclusão do Mestrado.

À todos amigos na Pró-Reitoria de Planejamento, Orçamento e Finanças (PROPLAN) da UFPE, por toda ajuda, apoio e suporte que me deram na etapas finais desse trabalho.

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

Ó profundidade das riquezas, tanto da sabedoria,
como da ciência de Deus! Quão insondáveis são
os seus juízos, e inescrutáveis, os seus caminhos!
Porque quem compreendeu o intento do senhor?
Ou quem foi seu conselheiro? Ou quem lhe deu
primeiro a ele, para que lhe seja recompensado?
Porque dele, e por ele, e para ele são todas as
coisas; glória, pois, a ele eternamente. Amém.

(BÍBLIA, 1995)

RESUMO

Técnicas de amostragem de área são aplicáveis quando a população-alvo pode ser delimitada geograficamente e um cadastro construído a partir de informações do solo, situação comum em pesquisas agropecuárias de âmbito nacional e regional. Os estimadores são baseados em como as unidades amostrais (segmentos) se relacionam com as unidades observacionais (campos e propriedades agrícolas). Estimação por segmento fechado com observação direta, por exemplo, não é suscetível a erros decorrentes de respostas imprecisas dos agricultores, mas é difícil de ser implementada na prática e está sujeito a erros de mensuração devido ao uso de instrumentos que auxiliam no processo de geração de medidas. Estimação por segmento ponderado, com subamostragem de pontos, baseia-se na opinião dos agricultores para obter os valores necessários para as variáveis de interesse, mas pode ser mais eficiente em pesquisas de múltiplos propósitos. O presente trabalho apresenta técnicas de estimação com cadastros de área com segmentos quadrados e exemplifica suas aplicações por meio de um estudo experimental realizado em Goiana-PE pelo laboratório CAST, para a FAO. O desempenho estatístico dos estimadores apresentados foi investigado através de um estudo de simulação de Monte Carlo. Os resultados do estudo mostraram que o estimador por segmento ponderado com subamostragem de pontos é mais preciso, em termos de erro quadrático médio, para estimar totais de área cultivada.

Palavras-chave: Estimação de área cultivada. Cadastro de área de segmentos quadrados. Pesquisas agropecuárias.

ABSTRACT

Area sampling techniques are used when the target population is delimited geographically and a frame constructed from soil information, a common situation in national and regional agricultural surveys. In such scenario, estimators take into account how sampling units (segments) and observational units (parcels and agricultural holdings) are related to each other. Estimation using closed segments with direct observation, for example, are not susceptible to errors due to imprecise answers of holders but are difficult to implement in practice and are subject to measurement errors related to the use of instruments that help in the process of generating measures. Estimation using weighted segments with subsampling of farms by points rely on the opinion of farmers to get the needed values for the variables of interest but can be more cost efficient in multi-purpose surveys. In this thesis, area sampling estimation techniques for square segments are introduced and their application exemplified using an experimental study carried out in Goiana by the CAST laboratory, for FAO. The statistical performance of the introduced estimators is investigated by a Monte Carlo simulation study. The results have shown that subsampling farmers using points within squared segments, and estimating the total area cultivated with a given crop using the weighted segment approach is the more efficient method of estimation, leading to the smallest values of mean square error among the studied area sampling and estimation techniques.

Keywords: Crop area estimation. Area frame of square segments. Agricultural surveys.

LISTA DE FIGURAS

Figura 1 – Divisão geopolítica da UF de Pernambuco em municípios.	23
Figura 2 – Malha de segmentos quadrados sobreposta na região de Goiana.	24
Figura 3 – Abordagem por pontos não agrupados.	25
Figura 4 – Amostragem de pontos dentro de segmentos (pontos agrupados).	25
Figura 5 – Campos de cultivo dentro de segmentos delineadas por meio de GPS.	28
Figura 6 – Amostragem por segmento fechado com observação direta dos tracts.	31
Figura 7 – Erro de especificação devido a identificação incorreta dos limites do segmento.	33
Figura 8 – Campos de cultivo dentro de segmento localizado em área urbana no município de Goiana.	36
Figura 9 – Identificação de campos de produção sobrepostos em Goiana.	42
Figura 10 – Pontos no segmento 364 em Goiana-PE.	44
Figura 11 – Segmentos estratificados no município de Goiana-PE.	46
Figura 12 – População de campos simulada para o cenário 1.	59
Figura 13 – Distribuição de frequência das áreas das propriedades no cenário 1.	60
Figura 14 – Distribuição de frequência das áreas das propriedades com as culturas I e II no cenário 1.	60
Figura 15 – População de campos simulada para o cenário 2.	61
Figura 16 – Distribuição de frequência das áreas das propriedades no cenário 2.	61
Figura 17 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura I, cenário 1.	64
Figura 18 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura II, cenário 1.	66
Figura 19 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura III, cenário 1.	68
Figura 20 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura IV, cenário 1.	70
Figura 21 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura I, cenário 2.	73
Figura 22 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura II, cenário 2.	75

Figura 23 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura III, cenário 2.	77
Figura 24 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura IV, cenário 2.	78

LISTA DE TABELAS

Tabela 1 – Definições dos estratos.	45
Tabela 2 – Classificação dos segmentos do cadastro de área de Goiana no momentos que a amostra foi selecionada.	45
Tabela 3 – Distribuição dos segmentos amostrados e efetivos, por estrato.	49
Tabela 4 – Estatísticas de frequência das unidades amostrais associadas às propriedades com cana-de-açúcar, em Goiana.	49
Tabela 5 – Estimativas para o total de área cultivada com cana-de-açúcar, por estrato e estimador, em Goiana.	50
Tabela 6 – Contribuição para a estimativa da variância do total de área estimada com cana-de-açúcar em Goiana, por estimador e estágio.	51
Tabela 7 – Estimativas para o total de fertilizantes (em toneladas) utilizado em Goiana, por estrato e estimador.	52
Tabela 8 – Contribuição para a estimativa da variância do total da quantidade de fertilizantes estimado (em toneladas) em Goiana, por estimador e estágio.	52
Tabela 9 – Eficiência da estratificação obtida no experimento de Goiana.	53
Tabela 10 – Probabilidades de associação entre tipos de cultivo e um polígono.	56
Tabela 11 – Número de segmentos na amostra, por estrato, para cada valor de aresta considerado.	62
Tabela 12 – Número de segmentos por estrato, cenário 1.	62
Tabela 13 – Estimativas para o total de área com a cultura I, cenário 1.	64
Tabela 14 – Estimativas para a variância do total estimado de área com a cultura I, cenário 1.	65
Tabela 15 – Estimativas para o total de área com a cultura II, cenário 1.	65
Tabela 16 – Estimativas para a variância do total estimado de área com a cultura II, cenário 1.	67
Tabela 17 – Estimativas para o total de área com a cultura III, cenário 1.	67
Tabela 18 – Estimativas para a variância do total estimado de área com a cultura III, cenário 1.	68
Tabela 19 – Estimativas para o total de área com a cultura IV, cenário 1.	69
Tabela 20 – Estimativas para a variância do total estimado de área com a cultura IV, cenário 1.	70

Tabela 21 – Número de segmentos por estrato, cenário 2.	71
Tabela 22 – Estimativas para o total de área com cultura I, cenário 2.	72
Tabela 23 – Estimativas para a variância do total estimado de área com a cultura I, cenário 2.	73
Tabela 24 – Estimativas para o total de área com cultura II, cenário 2.	74
Tabela 25 – Estimativas para a variância do total estimado de área com a cultura II, cenário 2.	75
Tabela 26 – Estimativas para o total de área com cultura III, cenário 2.	76
Tabela 27 – Estimativas para a variância do total estimado de área com a cultura III, cenário 2.	76
Tabela 28 – Estimativas para o total de área com cultura IV, cenário 2.	78
Tabela 29 – Estimativas para a variância do total estimado de área com a cultura IV, cenário 2.	79

LISTA DE ABREVIATURAS E SIGLAS

CV	Coeficiente de Variação
DP	Desvio Padrão
EQM	Erro Quadrático Médio
FAO	Food and Agriculture Organization of the United Nations
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GS	Global Strategy
HT	Horvitz-Thompson
LUCAS	Eurostat's Land Use/Cover Area frame Survey
MARS	Monitoring Agriculture with Remote Sensing
NASS	National Agricultural Statistics Service
NRCS	Natural Resource Conservation Service
USDA	United States Department of Agriculture

SUMÁRIO

1	INTRODUÇÃO	17
2	AMOSTRAGEM COM CADASTROS DE ÁREA	21
2.1	UNIDADES AMOSTRAIS EM CADASTROS DE ÁREA	21
2.1.1	Segmentos	22
2.1.2	Pontos	24
2.2	UNIDADES OBSERVACIONAIS EM CADASTROS DE ÁREA	26
2.2.1	Observação direta	27
2.2.2	Observação de unidades de produção	29
2.3	AMOSTRAGEM DE PROPRIEDADES COM CADASTROS DE SEGMENTOS	30
2.3.1	Segmento fechado	30
2.3.2	Segmento aberto	32
2.3.3	Segmento ponderado	34
2.3.4	Subamostragem de propriedades por pontos dentro de segmentos	35
2.3.5	Amostragem de propriedades por pontos	38
2.4	FERRAMENTAS PARA MELHORAR A EFICIÊNCIA AMOSTRAL	39
2.4.1	Estratificação	39
2.4.1.1	Estratificação em duas fases	41
2.4.1.2	Estratificação em dois estágios	41
2.4.2	Sistema de posicionamento geográfico	41
3	APLICAÇÃO EM UM ESTUDO AGROPECUÁRIO	43
3.1	O EXPERIMENTO DA FAO EM GOIANA	43
3.1.1	Cadastro de área	43
3.1.2	Estratificação dos segmentos	44
3.1.3	Estimadores	45
3.1.3.1	Estimador 1	47
3.1.3.2	Estimador 2	47
3.1.3.3	Estimador 3	48
3.1.4	Resultados do experimento	49
3.1.4.1	Estimativas de área cultivada com cana-de-açúcar	49
3.1.4.2	Estimativas da quantidade de fertilizantes	51

3.1.5	Eficiência da estratificação	52
4	AVALIAÇÃO NUMÉRICA DOS ESTIMADORES	54
4.1	PROTOCOLO DE SIMULAÇÃO	54
4.1.1	Geração de uma população de campos	55
4.1.2	Geração dos cadastros de área	56
4.1.3	Extração de variáveis resposta	57
4.2	EXPERIMENTO DE MONTE CARLO	57
4.2.1	Cenários gerados	59
4.3	RESULTADOS NUMÉRICOS	61
4.3.1	Cenário 1	62
4.3.2	Cenário 2	69
4.4	DISCUSSÃO DOS RESULTADOS	79
5	CONSIDERAÇÕES FINAIS	81
	REFERÊNCIAS	83
	APÊNDICE A – CÓDIGOS COMPUTACIONAIS	85

1 INTRODUÇÃO

Técnicas de amostragem são largamente utilizadas em pesquisas agropecuárias, pois possibilitam uma coleta de dados com maior rapidez, com qualidade de informação e a um menor custo que estudos censitários, viabilizando a realização de inferências para uma população de interesse. As reduções de custo e tempo devem-se ao fato de que informações são coletadas apenas para uma fração da população. Pelas razões supracitadas, técnicas de amostragem são empregadas em diferentes contextos, além de estudos agropecuários, como investigações epidemiológicas, estudos de impacto ambiental e estudos de impacto de políticas públicas, dentre outros.

Métodos probabilísticos de amostragem garantem a realização de inferência, a partir de uma amostra. É necessário que seja atribuída a todos os elementos da população-alvo uma probabilidade de seleção estritamente positiva para a amostra ser dita probabilística. A probabilidade de seleção é determinada por meio do processo de aleatorização, que, através de um protocolo de instruções, conhecido como plano amostral, define como os elementos da população são selecionados para compor a amostra. Na elaboração de qualquer plano amostral, é imprescindível que exista (ou que possa ser construído) um cadastro (COCHRAN, 1977; LOHR, 2009).

Cadastro é a designação dada a qualquer material ou instrumento usado para identificar e obter acesso a um elemento ou um conjunto de elementos (*cluster*) que compõem a população-alvo. Os elementos de um cadastro são conhecidos como unidades amostrais, pois o processo de seleção aleatória é aplicado a elas para determinar a amostra. Unidades amostrais podem ser número de telefone ou *e-mail* de indivíduos, endereços residenciais, limites territoriais e porções geográficas do solo, dentre outras. A implementação de um plano amostral depende da natureza das unidades amostrais no cadastro disponível bem como da necessidade de utilização de instrumentos ou informações auxiliares para acessar os elementos quando as unidades amostrais são clusters (FERRAZ, 2015).

Em pesquisas nas quais a população-alvo pode ser delimitada geograficamente, uma abordagem viável é utilizar o próprio território como base para construir um cadastro capaz de

fornecer cobertura completa. A região que cobre toda a população-alvo é dividida em parcelas de território com limites definidos e não sobrepostas, denominadas segmentos. A partição de uma área pode considerar limites físicos identificáveis na configuração do próprio solo, tais como estradas, rios, diferenças de vegetação e tipo de cultivo, ou limites que seguem um padrão geométrico virtualmente sobreposto à região, como por exemplo, uma malha de segmentos quadrados (GALLEGO *et al.*, 1995).

No contexto de pesquisas agropecuárias, cadastros cujas unidades amostrais são nomes de produtores e explorações agrícolas ou endereços de residências rurais são frequentemente identificados como cadastros de lista e são construídos com base em dados de censos agrícolas e populacionais e de registros administrativos. Cadastros cujas unidades amostrais são segmentos são chamados de cadastros de área e os procedimentos amostrais adotados fazem parte do escopo de amostragem de área. As unidades de um cadastro de área podem ser também pontos com uma pequena área associada e representação em um sistema de coordenadas (GSARS, 2015).

Embora a utilização de cadastros de lista tenha baixo custo, é comum que cadastros desse tipo apresentem problemas de cobertura e que se tornem defasados devido à grande quantidade de mudanças nas populações-alvo de estudos agropecuários. Um cadastro de lista construído com base em um censo agropecuário, por exemplo, não incorpora informações sobre o êxodo rural ao longo dos anos, tornando-o desatualizado para cobrir a população de estabelecimentos agropecuários de cunho familiar. Cadastros de área não apresentam problemas de cobertura, a menos que a população se estenda para áreas não cobertas pelo cadastro, mas esse problema é facilmente solucionado se forem incorporadas tais regiões ao cadastro. Todas as mudanças no uso da terra e na localização das explorações agrícolas, dentre outras, são captadas por um cadastro de área (HOUSEMAN, 1975; FAO, 1996).

Desde 1954, o *National Agricultural Statistics Service* (NASS) do *United States Department of Agriculture* (USDA) tem desenvolvido, usado e analisado cadastros de área como principal ferramenta para conduzir pesquisas agropecuária que visam múltiplas informações: área de cultivo, custo de produção e despesas da fazenda, produção e produtividade de grãos, inventários de pecuária e outros itens associados à agricultura (COTTER; NEALON, 1987). Para cada estado daquele país, com exceção do Alasca, um cadastro de área é construído usando limites físicos com informações auxiliares de imagens de satélites, orto-imagens e mapas topo-

gráficos e de cultivo. Após um período entre 15 e 20 anos, os cadastros precisam passar por um processo de atualização (DAVIES, 2009). Em paralelo, o *Natural Resource Conservation Service* (NRCS), também vinculado ao USDA, tem realizado levantamentos de dados longitudinais sobre as condições e tendências para o solo, água e outros recursos naturais (GOEBEL *et al.*, 1998). O NRCS utiliza um esquema de amostragem em dois estágios em que são selecionados segmentos quadrados no primeiro estágio e pontos dentro dos segmentos no segundo estágio.

O projeto MARS (*Monitoring Agriculture with Remote Sensing*) da Comunidade Europeia foi desenvolvido com um plano amostral em dois estágios semelhante ao aplicado pelo NRCS. Neste projeto, segmentos quadrados são selecionados no primeiro estágio e pontos com uma configuração pré-definida são usados para selecionar propriedades agrícolas no segundo estágio. Imagens de satélite foram utilizadas para estratificar os segmentos e fornecer informações auxiliares para melhorar a precisão de estimativas de áreas (TAYLOR *et al.*, 1997; GALLEGU; DELINCÉ; CARFAGNA, 1994).

O objetivo dessa dissertação é descrever e difundir técnicas de estimação com amostragem de área, pouco exploradas nos cursos básicos de amostragem. Em particular, são apresentadas técnicas baseadas em cadastros de segmentos quadrados cuja aplicabilidade vem ganhando espaço com os recentes avanços no campo da navegação por satélites (GSARS, 2017). Também são analisadas as principais vantagens e desvantagens de alguns estimadores, levando em consideração as atuais tecnologias de georreferenciamento. Um estudo experimental realizado no município de Goiana - Pernambuco é descrito a fim de ilustrar a aplicação de planos amostrais com cadastros de área (FERRAZ, 2018). Análises comparativas da eficiência de planos amostrais de área também são realizadas através de simulações de Monte Carlo. São apresentadas algumas ferramentas computacionais implementadas no ambiente de *software* de livre distribuição, R, que são úteis para planejar, selecionar e analisar amostras de área.

Essa dissertação está dividida em 5 capítulos, sendo o primeiro esta breve introdução. No capítulo 2, são descritos os principais conceitos associados à amostragem de área, como as principais unidades amostrais (segmentos e pontos) e observacionais (propriedades e campos de cultivo) de um cadastro de área, ferramentas para estratificar as unidades amostrais e alguns estimadores. No capítulo 3, está descrito o experimento com amostragem de área realizado em Goiana com o objetivo de ilustrar a amostragem de área na agropecuária. Os métodos

utilizados para construir e estratificar o cadastro de segmentos, os estimadores de área propostos e alguns dos principais resultados obtidos no experimento são pontuados. No capítulo 4, um estudo de simulação de Monte Carlo é apresentado, onde foram comparados os desempenhos dos estimadores de área aplicados no experimento de Goiana em diferentes cenários. No capítulo 5, são expostas as principais conclusões desse trabalho, bem como sugestões para análises e pesquisas futuras.

2 AMOSTRAGEM COM CADASTROS DE ÁREA

Problemas com a cobertura do cadastro se constituem como uma das principais fontes de erro não amostral. Pesquisas no setor agropecuário que almejam o levantamento de dados como área cultivada, produção, rebanhos, infraestrutura e socioeconômicos, dentre outras, são afetadas por esse problema. Devido a dinâmica das populações agrícolas, tanto de produtores quanto de núcleos familiares, a obtenção de um cadastro que contenha todos os elementos dessas populações torna-se um trabalho árduo e oneroso (GSARS, 2015). Um cadastro de baixo custo e boa cobertura pode ser obtido a partir de censos populacionais e agrícolas precedentes ao estudo amostral. Contudo, cobrir uma população inteira requer um maior empenho em termos de tempo e recursos financeiros; isso faz com que censos sejam realizados em intervalos de cinco a dez anos, o que torna as informações desatualizadas, a depender da diferença entre o ano do levantamento amostral e o ano de referência do censo.

Quando a população-alvo pode ser delimitada geograficamente, uma solução com grande potencial é utilizar o próprio território como base para construir um cadastro que apresenta, intrinsecamente, cobertura completa, i. e., um cadastro que contenha todos os elementos da população-alvo. O primeiro objetivo em uma pesquisa com amostragem de área é definir a região de interesse, com limites geográficos bem estabelecidos. Em seguida, as informações espaciais da região delimitada devem ser transformadas, de alguma maneira, em unidades amostrais de um cadastro.

2.1 UNIDADES AMOSTRAIS EM CADASTROS DE ÁREA

Utilizar o próprio território para atingir a cobertura completa trás consigo outros benefícios inerentes ao método; não importa a dinâmica das populações agrícolas, um cadastro de área consegue absorver mudanças, como o tipo de pecuária e o uso do solo, dentre outras. Contudo, é necessário um meio de relacionar o território com as unidades amostrais e, ademais, com as unidades observacionais (fazendas, campos de cultivo). As unidades amostrais de um cadastro de área podem ser segmentos, pontos ou transectos (faixas de solo com certo comprimento). Nessa dissertação, são exploradas técnicas de amostragem suportadas por

cadastros de área que utilizam como unidades amostrais segmentos, em particular segmentos quadrados, e pontos.

2.1.1 Segmentos

No âmbito da amostragem de área, segmentos são definidos como parcelas territoriais de área não nula, cujas formas são delineadas sobre a própria região de interesse com moldes regulares ou irregulares. Os segmentos são delineados garantindo que não haja nenhuma sobreposição entre dois ou mais segmentos. Isso garante que um cadastro que tenha como unidades amostrais segmentos apresente identificabilidade, uma vez que qualquer área na região poderá ser associada de maneira única aos segmentos. Por outro lado, a união dos segmentos deve cobrir toda a região de interesse, para garantir a completude da cobertura.

Uma questão primordial quando segmentos são usados em um cadastro de área é escolher a forma que eles apresentarão: se irregular, considerando limites físicos presentes na configuração do próprio solo, seguindo estradas, rios, diferenças de vegetação e fronteiras, dentre outros; se regular, seguindo um padrão de formas geométricas, como quadrados e retângulos.

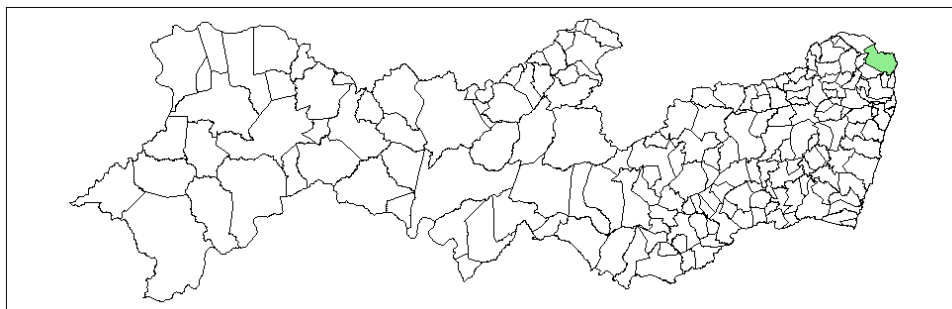
No caso em que limites físicos são utilizados, existe um custo adicional com materiais auxiliares por meios dos quais os segmentos são delineados, tais como imagens de satélite, orto-imagens e mapas topográficos e de cultivo. Também é necessário atualizar regularmente, tanto o cadastro quanto os materiais auxiliares, devido a alterações que a paisagem sofre eventualmente, pois os pesquisadores em campo devem identificar com facilidade as fronteiras dos segmentos. Os cadastros produzidos pelo NASS através dessa abordagem são utilizados por um período de 15 à 20 anos até que uma atualização seja realizada.

Segmentos geométricos são facilmente projetados em qualquer região, pois ignoram a configuração presentes na paisagem. Isto faz a produção de um cadastro ser mais rápida e suprime a necessidade de atualizações. Contudo, ignorar a configuração do solo pode acarretar maiores problemas no tocante a locomoção no trabalho de campo. A principal dificuldade dessa abordagem é a identificação correta dos segmentos em solo, antes feita por meio de pontos de referência e algum tipo de material de apoio. Porém esse procedimento ficou obsoleto devido ao

desenvolvimento de sistemas de navegação por satélites (GNSS – do inglês, Global Navigation Satellite System) que garantem certa confiabilidade na identificação das fronteiras dos segmentos (CARFAGNA, 2013). Em particular, a utilização de dispositivos GPS (do inglês, Global Positioning System) pode simplificar a localização dos segmentos em solo.

Em estudos amostrais que abrangem vastas regiões, pesquisas de âmbito nacional em países de grande extensão territorial, por exemplo, um procedimento usual é primeiro dividir a região em grandes parcelas de solo, compreendidas com unidades amostrais primárias (uap's), que podem conter múltiplos segmentos, unidades amostrais secundárias (uas's). O projeto do NASS usa limites físicos para determinar as uap's que contenham de 6 a 8 segmentos também com limites físicos. O NRCS e o projeto MARS apresentam uap's em blocos quadrados, subdivididos em segmentos quadrados. A divisão geopolítica Brasileira permite misturar as abordagens por limites físicos e geométricos. Toda a extensão territorial do Brasil está dividida em 27 Unidades Federativas (UF) que são compostas por municípios. A figura 1 mostra a divisão da UF de Pernambuco (PE) em municípios, em destaque o município de Goiana. A figura 2 mostra um cadastro de segmentos geométricos sobreposto à região de Goiana.

Figura 1 – Divisão geopolítica da UF de Pernambuco em municípios.



Um fator importante é determinar o tamanho do segmento. GSARS (2015) sugere que o tamanho dos segmentos seja escolhido de modo que os pesquisadores em solo não ultrapassem o limite de um dia de trabalho (de 2 a 4 horas se possível) para observar um segmento por completo. Logo, deve-se considerar o tipo de relevo e da cobertura vegetal presente no solo bem como a disponibilidade de veículos que auxiliem a locomoção. Outro fator que deve ser considerado é o número de campos (porção de terra separada por linhas de demarcação facilmente reconhecíveis) dentro do segmento, caso o esquema amostral exija que todos os campos sejam visitados.

Figura 3 – Abordagem por pontos não agrupados.

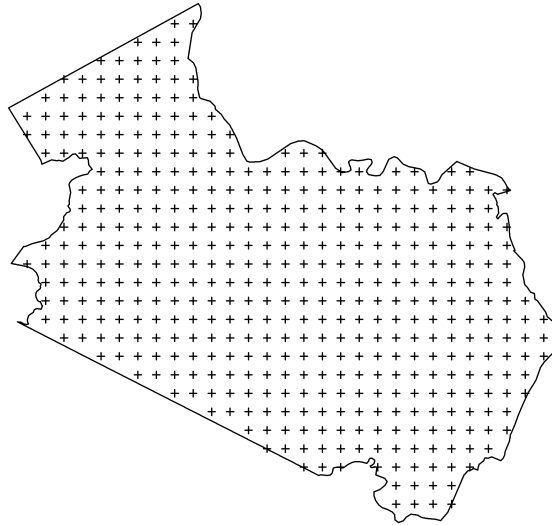


Figura 4 – Amostragem de pontos dentro de segmentos (pontos agrupados).



Alguns fatores devem ser levados em consideração na etapa de planejamento do estudo amostral. A disponibilidade de veículos para todos os pesquisadores e a existência de uma rede de estradas de qualidade favorecem o uso de pontos não agrupados, pois o deslocamento de alguns quilômetros entre pontos pode ser facilmente realizado. Caso isso não aconteça, um só veículo pode deixar vários pesquisadores em um segmento que percorrem as distâncias entre pontos caminhando (o número de pontos depende do tamanho do segmento e da distância que se pretende percorrer em solo.). Além disso, a cobertura do solo deve ser observada. Se a região apresentar como principal fonte de produção grandes latifúndios ou cobertura vegetal homogênea é provável que pontos não agrupados sejam mais eficientes. Se for predominante

na região a agricultura de subsistência, muito comum em países em desenvolvimento, onde a produção concentra-se em pequenas propriedades, a amostragem de pontos agrupados dentro de segmentos deve ser mais eficiente.

Quando são usados para fins de estratificação, estimação (por imagens ou observação direta) ou classificação de solo, os pontos passam a ter área ao considerar a região adjacente à suas coordenadas, como um círculo (ou uma matriz de *pixels*) (CARFAGNA; GALLEGU, 2005; EUROPEAN, 2002). O tamanho da vizinhança depende da resolução das imagens e materiais ancilares. Por exemplo, com imagens de satélite de baixa resolução uma vizinhança maior precisa ser analisada.

O projeto MARS e o estudo do NRCS utilizam pontos subamostrados em segmentos quadrados. Um esquema semelhante a este foi adotado nos estudos experimentais realizados no Brasil e no Nepal em 2016. O projeto LUCAS (*Eurostat's Land Use/Cover Area frame Survey*) da Comunidade Europeia utiliza pontos para amostragem de fazendas em duas fases: na primeira fase, uma amostra é selecionada uma grande amostra de pontos que são foto-interpretados por meio de fotos aéreas e imagens de satélite para propostas de estratificação; na segunda fase, uma amostra final de pontos é retirada em cada estrato (EUROPEAN, 2002).

2.2 UNIDADES OBSERVACIONAIS EM CADASTROS DE ÁREA

Pretender que um estudo amostral seja capaz de fornecer informações úteis para múltiplos propósitos é um importante fator na elaboração do plano amostral; quanto mais informações forem obtidas mais relevante e eficiente será o estudo. Por exemplo, é desejável que em uma pesquisa agropecuária, além de produzir informações de produção, área de cultivo e uso de fertilizantes, sejam produzidos dados socioeconômicos, geográficos, ambientais, hidrológicos, geológicos, entre outras. Isso dependerá da existência de meios para coletar dados nas unidades observacionais (unidades nas quais são aplicados questionários e procedimentos de medição). Neste sentido, a amostragem de área apresenta-se muito versátil, pois o levantamento de informações pode ser realizado em campo pelos pesquisadores, portando instrumentos de mensuração adequados, ou por entrevistas aos proprietários de unidades de produção.

2.2.1 Observação direta

Qualquer informação obtida pelo pesquisador com auxílio instrumental, sem que seja necessário o uso de questionários ou entrevistas, pode ser entendido como uma observação direta. No contexto de pesquisas agropecuárias, a abordagem da observação direta tem como principal alvo os campos de cultivo; neles podem ser coletadas informações acerca da produção (e.g. quantidade de culturas presentes) e área de cultivo. Cada segmento é composto por múltiplos campos produtivos, ou seja, um *cluster* de unidades observacionais. Todos os campos dentro de um mesmo segmento podem ser observados (isto lembra amostragem por conglomerado) e, neste caso, realiza-se a observação completa dos segmentos. Contudo, se o segmento possuir uma grande quantidade de campos a observação completa torna-se árdua. Neste caso, é aconselhável reduzir o tamanho ou subamostrar o segmento. Quando pontos são utilizados para subamostrar segmentos faz-se necessário um maior cuidado com problemas de localização incorreta das unidades amostrais.

Utilizar um dispositivo de GPS de preço e precisão razoáveis melhora o desempenho de pesquisas com observação direta, especialmente se informações geoespaciais são desejadas. Sua utilização justifica-se não apenas por facilitar a localização das unidades amostrais (segmentos e pontos), mas também por delimitar os campos de plantio fornecendo dados acerca de suas formas através das coordenadas geográficas. O GPS também é capaz de fornecer uma aproximação da área de um campo em tempo real (CARFAGNA, 2013). A figura 5 mostra algumas campos de cultivo, delimitado com um dispositivo de GPS, no município de Goiana.

Por outro lado, as observações não necessariamente precisam ser realizadas em solo, mas podem ser realizada por meio de imagens aéreas e de satélite. Em Ruanda, a aplicação de imagens de satélite foi testada para classificar o tipo de cultivo presente no solo. Esse estudo mostrou resultados promissores apenas quando os campos tem área superior a 0,5 ha (hectares) e com a presença de uma única cultura. Entretanto, em campos de área menor de 0,3 ha e com a presença de múltiplas culturas, cenário comum naquele país, a classificação mostrou resultados pobres, sendo aconselhável sua utilização como complemento em um estudo por amostragem de

2.3 AMOSTRAGEM DE PROPRIEDADES COM CADASTROS DE SEGMENTOS

Selecionar uma amostra de propriedades agrícolas a partir de um cadastro do tipo lista é, sem dúvida, mais barato do que usar um cadastro de área. No entanto, devido a problemas de atualização e abrangência, pode haver uma diferença substancial entre a população presente no cadastro de lista e a real população. A aplicação de cadastros de área para amostrar propriedades requer alguma regra de associação entre as propriedades e as unidades amostrais do cadastro. Quando segmentos são utilizados como meio para selecionar propriedades existem três abordagens clássicas conhecidas como segmento aberto, fechado e ponderado (HENDRICKS; SEARLS; HORVITZ, 1965; NEALON, 1984). GALLEGO (1994) apresenta uma abordagem de seleção de propriedades com subamostragem dos segmentos através de pontos.

2.3.1 Segmento fechado

Na abordagem por segmento fechado, todas as propriedades que possuem áreas de operação dentro do segmento são selecionadas para compor a amostra. Todas as áreas agrícolas dentro do segmento sob a supervisão de um mesmo proprietário são chamadas de *tracts* e as informações fornecidas pelos proprietários devem ser relativas ao tract e não à propriedade completa. O estimador por segmento fechado é comumente usado para estimar áreas de cultivo e números acerca de rebanhos. No tocante a estimação de áreas de cultivo, os tracts podem ser delineados em mapas para verificações e correções posteriores (possivelmente usando imagens aéreas). Deve-se ainda ter cuidado com informações de rebanho dos tracts localizados nas bordas do segmento, pois parte do rebanho relatado pode estar fora do segmento.

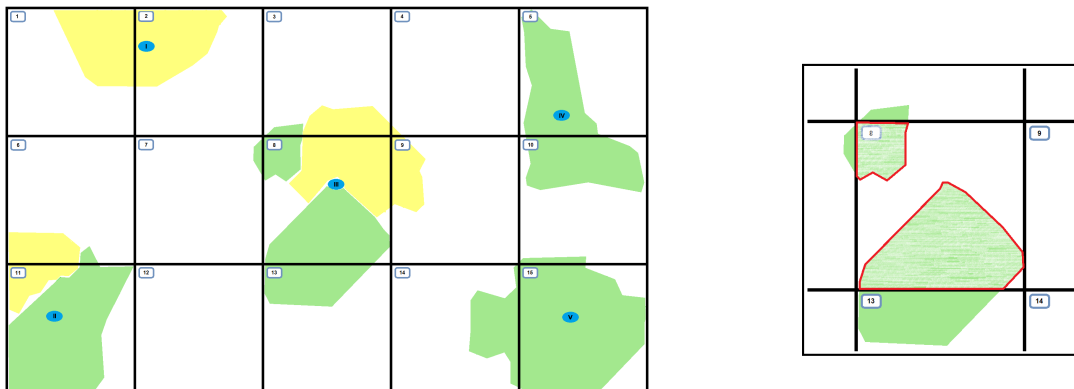
A principal limitação dessa abordagem é que muitas vezes os proprietários são capazes de fornecer dados globais da propriedade, mas suas respostas são imprecisas para o tract. Se os limites dos segmentos são geométricos obter respostas precisas dos proprietários torna-se ainda mais difícil. O levantamento de informações, tais como produção e quantidade de fertilizantes, são desfavorecidas por essa abordagem. Além disso, o número de informações que podem ser obtidas é limitado. Por exemplo, o proprietário consegue informar quantos tratores são usados na propriedade, mas para o tract essa informação não faz sentido. A obtenção de

dados socioeconômicos também exemplifica essa limitação.

Embora essa abordagem apresente algumas limitações quando as informações são coletadas dos proprietários, ela pode ser útil se técnicas de observação direta forem aplicadas. A área de plantio com determinada cultura interna ao segmento pode ser medida com auxílio de um dispositivo de GPS (ou imagens aéreas). Essa foi uma das abordagens aplicadas no experimento realizado em Goiana. Se um GPS for usado para delinear e medir a área dos campos de cultivo presentes nos tracts, fica opcional a coleta dessas informações diretamente dos proprietários, que podem ser obtidas com intuito de verificar a qualidade das respostas fornecidas.

A figura 6 ilustra como essa abordagem funciona: o segmento 8 selecionado para compor a amostra apresenta dois tracts. Os proprietários fornecem as informações relativas aos tracts, que são delineados pelo pesquisador obedecendo aos limites geográficos do segmento.

Figura 6 – Amostragem por segmento fechado com observação direta dos tracts.



Considere que y_i representa o valor da variável de interesse no i -ésimo segmento, mensurada com a abordagem de segmento fechado através da observação direta ou entrevistando os proprietários. O peso amostral de y_i é o inverso da probabilidade de inclusão do segmento na amostra. Assim, se a região geográfica é totalmente coberta por N segmentos, dos quais uma amostra aleatória simples de tamanho n é selecionada, a probabilidade de seleção de quaisquer segmento é $\frac{n}{N}$. Dessa forma, o estimador de Horvitz-Thompson (HT) para o total da variável y será dado por

$$\hat{\tau}_y = \frac{N}{n} \sum_{i \in s} y_i = N\bar{y}_s, \quad (2.1)$$

e a variância de $\hat{\tau}_y$

$$\text{Var}(\hat{\tau}_y) = N^2 \left(1 - \frac{n}{N}\right) \frac{\sigma_y^2}{n},$$

em que, s é o conjunto de segmentos na amostra e σ_y^2 é a variância da variável resposta na população de segmentos. Um estimador não viesado para $\text{Var}(\hat{\tau}_y)$ é dado por

$$\widehat{\text{Var}}(\hat{\tau}_y) = N^2 \left(1 - \frac{n}{N}\right) \frac{S_y^2}{n}, \quad (2.2)$$

em que, S_y^2 é a variância da variável resposta nos segmentos da amostra, i. e.,

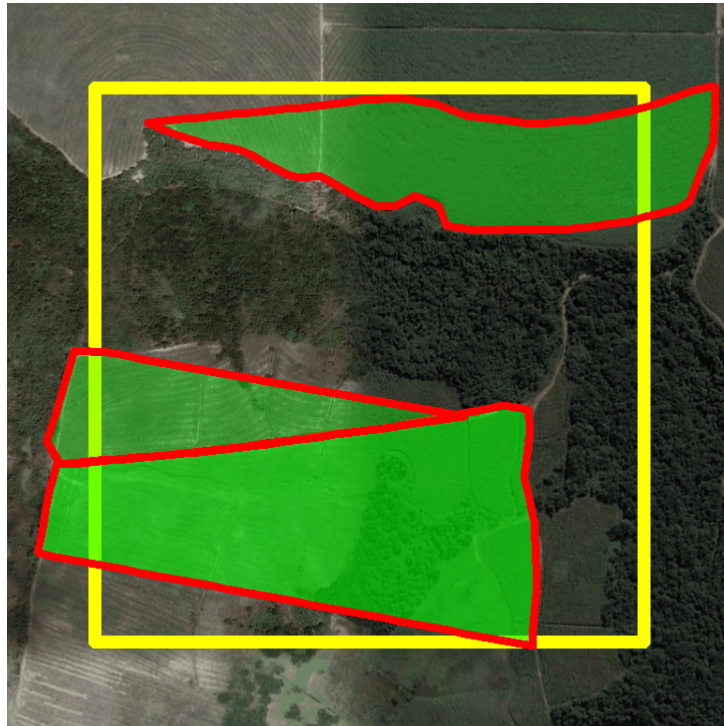
$$S_y^2 = \frac{1}{n-1} \left(\sum_{i \in S} y_i^2 - N\bar{y}_s^2 \right).$$

Existem alguns problemas em utilizar a abordagem de segmento fechado com observação direta no tocante ao trabalho de campo. Dificuldades podem surgir devido à locomoção em solo, pois determinar, por exemplo, a área exata de plantio exige uma incursão direta do pesquisador. Outra dificuldade que pode aparecer é a existência de múltiplas culturas numa mesma área. A depender do tamanho escolhido do quadrado, a cobertura de todo o segmento torna-se muito demorada. A identificação incorreta dos limites do segmento é um possível erro não amostral de especificação que resulta diretamente em erros de mensuração das variáveis de interesse. A figura 7 mostra que o pesquisador incluiu áreas dos campos de cultivo fora dos limites do segmento, no experimento de Goiana.

2.3.2 Segmento aberto

Na abordagem por segmento aberto, são observadas todas as propriedades que possuam “sede” dentro dos limites do segmento e as informações obtidas dos proprietários são relativas à propriedade inteira. Caso um segmento seja selecionado para compor a amostra, não é necessário que todos os campos de cultivo sejam visitados, mas que sejam localizadas áreas com

Figura 7 – Erro de especificação devido a identificação incorreta dos limites do segmento.



algum tipo de construção que possam ser classificadas com uma sede. Para tanto, a definição de sede deve ficar patente para os pesquisadores, a fim de evitar subjetividade e imprecisão nesta classificação. Por exemplo, a sede pode ser a casa onde o proprietário reside, o prédio de maior valor na propriedade ou até mesmo a entrada principal da propriedade.

A principal vantagem em relação à abordagem de segmento fechado é que o número de itens que podem ser estudados não é restrito, justamente porque os dados requisitados são relativos à propriedade. Além disso, existe uma redução de custos devido a um menor número de entrevistas (observações); mesmo que sua propriedade possua tracts em mais de um segmento na amostra o proprietário é entrevistado uma única vez, se o segmento que contém a sede for selecionado. O mesmo não acontece com as abordagens por segmento fechado e ponderado onde todos os proprietários de tracts no segmento são entrevistados (mais de uma vez para o segmento ponderado, inclusive).

A estimação por segmento aberto determina que uma propriedade seja associada a um único segmento. Assim, o fator de expansão (peso amostral) da variável resposta é o inverso da probabilidade de inclusão de um único segmento na amostra. Portanto, sob amostragem aleatória simples de segmentos, o estimador para o total de um variável de interesse y decorre de (2.1)

$$\hat{t}_y = \frac{N}{n} \sum_{i \in s} \sum_{j \in s_i} y_{ij}$$

em que, s_i é conjunto de propriedades no segmento i e y_{ij} é o valor assumido pela variável de interesse na propriedade j com sede no segmento i .

Em contrapartida, a variância amostral é bem maior para o segmento aberto se comparado com outras abordagens. O número de propriedades pode variar muito entre os segmentos, dependendo se a paisagem apresenta grandes latifúndios ou agricultura familiar. A fórmula para o estimador da variância é a mesma de (2.2), mas neste caso as estimativas são maiores, uma vez que essa abordagem pode ser vista como uma amostragem por conglomerados com números distintos de observações (LOHR, 2009). Por exemplo, se a região de interesse apresentar uma paisagem com grandes latifúndios é esperado que muitos segmentos apresentem variáveis resposta iguais a zero.

Além do mais, pode haver uma alta taxa de não resposta, devido ao receio dos proprietários em fornecer informações sobre as propriedades inteiras. Pequenas explorações tendem a não possuir nenhuma sede, porque o proprietário reside em áreas urbanas e, portanto, essas áreas devem ser incluídas na amostra. Deve-se ainda tentar incorporar à propriedade informações de áreas arrendadas para exploração de terceiros.

2.3.3 Segmento ponderado

Nesta abordagem, todos os proprietários que possuam tracts dentro do segmento selecionado devem fornecer dados pertinentes a toda propriedade. Naturalmente, as variáveis de interesse obtidas de algumas propriedades podem ser utilizadas mais de uma vez, caso a propriedade possua tracts em mais de um segmento. Uma ponderação através da razão entre a área do tract e a área de operação agrícola na propriedade é imposta às variáveis respostas, gerando uma nova variável

$$x_{ij} = \frac{T_{ij}}{A_j} y_j \quad (2.3)$$

em que T_{ij} é a área do tract da propriedade j no segmento i e A_j a área de operação na propriedade

j . Essa abordagem cria uma variável aleatória fictícia x uniformemente distribuída em A_j . Calcular o total de x é o mesmo que calcular o total de y ,

$$\tau_x = \sum_{j \in F} x_j = \sum_{j \in F} \sum_{i=1}^N x_{ij} = \sum_{j \in F} \sum_{i=1}^N \frac{T_{ij}}{A_j} y_j = \sum_{j \in F} \frac{y_j}{A_j} \sum_{i=1}^N T_{ij} = \sum_{j \in F} \frac{y_j}{A_j} A_j = \tau_y$$

em que, F é o conjunto de todas as propriedades na região. Sob amostragem aleatória simples, o estimador de HT para o total na abordagem por segmento ponderado é expresso por

$$\hat{\tau}_y = \frac{N}{n} \sum_{i=1}^N \sum_{j \in s_i} \frac{T_{ij}}{A_j} y_j.$$

A estimação por segmento ponderado fornece estimativas mais precisas se comparado com a estimação por segmento aberto e fechado. A ponderação reduz o efeito das respostas de grandes propriedades, que aumentam a variância do estimador por segmento aberto, e reduz a influência de valores de tracts muito grandes, que afetam a variância do estimador por segmento fechado. Para as culturas com maiores áreas de plantio, o estimador por segmento ponderado apresenta apenas uma pequena redução na variância em relação ao segmento fechado (NEALON, 1984). Adicionalmente, nessa abordagem não se faz necessário determinar onde está localizada a sede da propriedade, apenas localizar o proprietário, além de não restringir o escopo de itens que podem ser pesquisados, limitação apresentada para a estimação por segmento fechado.

O estimador pode ter problemas de subestimação caso os proprietários informam erroneamente o total de área de operação por incluir, por exemplo, áreas florestais e ociosas. As áreas relatadas dos tracts são subjetivas e podem estar distantes das reais áreas, principalmente se forem utilizados segmentos geométricos. Uma alternativa seria determinar as áreas dos tracts através de um dispositivo GPS, com custo adicional. Além disso, regiões urbanas devem ser incluídas por poder conter áreas de cultivo, como visto na figura 8.

2.3.4 Subamostragem de propriedades por pontos dentro de segmentos

As abordagens por segmento fechado e ponderado exigem que dados sejam levantados para cada tract no segmento. Caso o segmento apresente um grande número de campos

Figura 8 – Campos de cultivo dentro de segmento localizado em área urbana no município de Goiana.



de cultivo, a tarefa de identificar e localizar os proprietários dos tracts corretamente torna-se onerosa, especialmente se for empregada a observação direta. Nesse caso, uma alternativa para reduzir o trabalho de campo é usar subamostragem por pontos. Para cada ponto que cair em áreas de uso agrícola, o pesquisador deve localizar e entrevistar o proprietário (ou fazer mensurações no campo por observação direta).

A principal vantagem da subamostragem por pontos em relação à estimação por segmento ponderado é que não se faz necessário saber a área dos tracts, devido à construção do estimador de HT. Supondo que as áreas dos tracts são uniformemente distribuídas dentro do segmento, a probabilidade inclusão na amostra da propriedade j dentro do i -ésimo segmento, onde são subamostrados P_i pontos, será

$$\pi_{j|i} = P_i \frac{T_{ij}}{T_i}$$

onde T_{ij} é a área do tract da propriedade j no segmento i e T_i é a área do segmento i (inclusas áreas não agrícolas). A estimativa para o total de uma variável dentro de um segmento, usando a transformação dada em (2.3), é calculada por

$$\hat{X}_i = \sum_{j=1}^{P_i} \frac{x_{ij}}{\pi_{j|i}} = \sum_{j=1}^{P_i} \frac{T_i}{T_{ij}P_i} \frac{y_j T_{ij}}{A_j} = \frac{T_i}{P_i} \sum_{j=1}^{P_i} \frac{y_j}{A_j}.$$

Assim, sob uma amostragem aleatória simples de segmentos, com subamostragem de pontos dentro dos segmentos, o estimador de HT para o total de uma variável y é explicitado por

$$\hat{\tau}_y = \sum_{i \in s} \frac{\hat{X}_i}{\pi_i} = \sum_{i \in s} \frac{N}{n} \sum_{j=1}^{P_i} \frac{T_{ij} y_j}{P_i A_j} = \frac{N}{n} \sum_{i \in s} \frac{T_i}{P_i} \sum_{j=1}^{P_i} \frac{y_j}{A_j}. \quad (2.4)$$

Esse estimador não requer o computo da variável artificial x , tampouco que a área dos tracts sejam conhecidas, apenas que as informações globais da propriedade sejam conhecidas. Se os segmentos apresentarem forma geométrica, como quadrados, então $T_i = T$ e, usualmente, $P_i = P$ para todos segmentos no cadastro. As expressões para a variância e para um estimador não enviesado da variância do estimador dado em (2.4) podem ser derivadas a partir da fórmula para a variância de planos amostrais em dois estágios (COCHRAN, 1977)

$$\text{Var}(\hat{\tau}_y) = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) \sigma_X^2 + \frac{N}{n} \sum_{i \in s} \frac{1}{P_i(P_i - 1)} \sum_{j=1}^{P_i} \left(\frac{y_j}{A_j} T_i - \hat{X}_i\right)^2$$

e

$$\widehat{\text{Var}}(\hat{\tau}_y) = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) S_X^2 + \frac{N}{n} \sum_{i \in s} \frac{1}{P_i(P_i - 1)} \sum_{j=1}^{P_i} \left(\frac{y_j}{A_j} T_i - \hat{X}_i\right)^2, \quad (2.5)$$

onde

$$\sigma_X^2 = \frac{1}{N-1} \sum_{i \in U} \left(\hat{X}_i - \bar{\hat{X}}\right)^2, \quad S_X^2 = \frac{1}{n-1} \sum_{i \in s} \left(\hat{X}_i - \bar{\hat{X}}\right)^2 \quad \text{e} \quad \bar{\hat{X}} = \frac{1}{n} \sum_{i \in s} \hat{X}_i,$$

em que U é o conjunto, com tamanho N , de segmentos na população.

É importante salientar que nessa abordagem os tracts são selecionados com reposição, com probabilidade de seleção proporcional a T_{ij}/T_i , e incluir na amostra as propriedades j e j' tem probabilidade diferente de incluir a propriedade j duas vezes, desde que exista uma distância relativamente grande entre as propriedades. Na construção do estimador (2.4) esse fato foi ignorado (GALLEGO; DELINCÉ; CARFAGNA, 1994).

Alternativamente, podem-se usar apenas pontos que caíam em áreas agrícolas para obter outro estimador. Neste caso, o número F_i de pontos que recaem em áreas agrícolas no segmento i são usados para selecionar tracts uniformemente distribuídos em $G_i = \sum_{j \in s_i} T_{ij}$, o total de área agrícola no segmento i . No segmento i são observados quantos pontos forem necessários para observar F_i pontos agrícolas. As expressões (2.4) e (2.5) podem ser adaptadas substituindo P_i por F_i e T_i por G_i . Assim,

$$\hat{\tau}_y = \sum_{i \in s} \frac{\hat{X}_i}{\pi_i} = \frac{N}{n} \sum_{i \in s} \frac{G_i}{F_i} \sum_{j=1}^{F_i} \frac{y_j}{A_j} \quad (2.6)$$

e

$$\widehat{\text{Var}}(\hat{\tau}_y) = \frac{N^2}{n} \left(1 - \frac{n}{N}\right) S_X^2 + \frac{N}{n} \sum_{i \in s} \frac{1}{F_i(F_i - 1)} \sum_{j=1}^{F_i} \left(\frac{y_j}{A_j} G_i - \hat{X}_i\right)^2 \quad (2.7)$$

são estimadores não viesados para o total τ_y e para $\text{Var}(\hat{\tau}_y)$, respectivamente.

Um ponto crucial é que a definição usada para decidir se um ponto recai sobre terras agrícolas deve está em consonância com a definição de terra agrícola operada, usada na entrevista com os proprietários e na medição das áreas dos tracts. No caso em que o segmento não possua áreas de uso agrícola uma variável fictícia de valor zero pode ser criada e a segunda parte da expressão (2.7) tem um termo nulo em sua soma. O estimador da variância de $\hat{\tau}_y$ não pode ser calculado caso $F_i = 1$. Nesse caso, o valor zero pode ser atribuído, embora isso acarrete na subestimação da variância dentro dos segmentos.

2.3.5 Amostragem de propriedades por pontos

Propriedades podem ser amostradas através de cadastro compostos exclusivamente por pontos, como exemplificado na figura 3. Nessa abordagem se pelo menos um ponto recair em terras agrícolas dentro de uma propriedade ela é inclusa na amostra, e o proprietário deve fornecer dados relativos a toda propriedade. Assim, a probabilidade de selecionar uma propriedade é proporcional a razão entre a área agrícola na propriedade e a área total da região. É importante perceber que a amostra é selecionada com reposição e, sob certas circunstâncias, é razoável supor que as razões entre as áreas operadas das propriedades e a área total da região são relativamente

pequenas. Sob essas condições, a probabilidade de inclusão pode ser aproximada por

$$\pi_j \approx \frac{mA_j}{R}$$

onde m é o número de pontos selecionados, A_j é a área agrícola da propriedade j e R é área da região. O estimador para o total de y , sob um plano de amostragem aleatória simples com reposição, conhecido como estimador de Hansen–Hurwitz (HH), é expresso por

$$\hat{\tau}_y = \sum_{j=1}^m \frac{y_j}{\pi_j} = \frac{R}{m} \sum_{j=1}^m \frac{y_j}{A_j} \quad (2.8)$$

e o estimador não viesado para variância de $\hat{\tau}_y$, encontrado em (LOHR, 2009), é dado por

$$\hat{\text{Var}}(\hat{\tau}_y) = \frac{1}{m(m-1)} \sum_{j=1}^m \left(\frac{Ry_j}{A_j} - \hat{\tau}_y \right)^2.$$

Os pontos que não incidem em terras agrícolas têm imputados valores $y_j = 0$. Isso inflaciona a variância do estimador (2.8). Uma estratégia que reduz a variância desse estimador é, antes da seleção dos pontos, fazer uma simples estratificação em dois estratos, pontos não agrícolas e pontos agrícolas, por meio de fotointerpretação de imagens aéreas e de satélites. A maior limitação da amostragem por pontos é a cobertura de propriedades que possuem rebanhos, mas nenhuma terra agrícola. Uma possível solução para este problema é considerar uma densa amostra de pontos nas áreas construídas das regiões rurais, em busca de currais e granjas, por exemplo.

2.4 FERRAMENTAS PARA MELHORAR A EFICIÊNCIA AMOSTRAL

2.4.1 Estratificação

A estratificação das unidades amostrais é uma estratégia usada para aumentar a precisão do estimador em termos de variância, se em cada estrato as variáveis de interesse apresentem comportamento homogêneo. Em cadastros de área, os estratos são baseados geralmente em características presentes na paisagem do terreno, que podem ser identificadas por meio de fotointerpretação de imagens aéreas e de satélites. Por exemplo, os segmentos podem ser estratificados

considerando a cobertura de solo dominante, como áreas florestais, de cultivo permanente ou de pastagem. Uma definição comum para os estratos é considerar o percentual de área agrícola nos segmentos, e. g., “altamente agrícola” se área agrícola $> 70\%$ e “pouco agrícola” se área agrícola $< 30\%$. Em verdade, existe uma ampla variedade de possíveis rótulos para os estratos. Por isso, deve-se tentar garantir primordialmente a consonância com os objetivos do estudo amostral para que a estratificação seja eficiente.

Se as imagens pelas quais a estratificação é realizada estiverem desatualizadas, pode haver uma inconsistência entre o estrato da unidade amostral e paisagem observada na pesquisa de campo. Isso deve reduzir a eficiência da estratificação em termos de variância, mas não introduz viés nas estimativas baseadas na amostra sob estratificação. De fato, o viés dos estimadores de área pode derivar da presença de erros sistemáticos nas observações feitas pelos pesquisadores, mas não da estratificação inconsistente (DELINCÉ, 2015). A única situação que introduz um viés negativo nas estimativas é quando unidades amostrais classificadas como não agrícolas não forem inclusas na pesquisa amostral, embora elas apresentem um número significativo de atividades agropecuárias (GSARS, 2017). Esse viés é reduzido ao selecionar uma pequena amostra de segmentos não agrícolas.

Uma estratificação eficiente e de baixo custo pode ser realizada através de imagens de *sensoriamento remoto*, imagens geradas a partir da captura da radiação eletromagnética refletida pela superfície terrestre por sensores colocados em satélites e aeronaves. Essas imagens podem ser obtidas gratuitamente em fontes de livre distribuição, como o Bing e o Google Earth. Com essas imagens, especialistas podem construir mapas de cobertura terrestre que são usados para auxiliar na etapa de estratificação. Ainda algoritmos de classificação podem ser empregados para classificar essas imagens (CARFAGNA; GALLEG0, 2005).

Outra forma de estratificação é através da observação de pontos que podem ser visitados ou foto-interpretados remotamente. Ao contrário de segmentos, pontos são estratificados diretamente pela observação da cobertura do solo, permitindo a criação de estratos específicos como “arroz” e “cana-de-açúcar”. Por simplicidade, os pontos podem ser classificados como agrícolas ou não agrícolas, observando a localização exata do ponto e também sua área adjacente. Uma desvantagem surge devido a grande quantidade de pontos que precisam ser classificados em um cadastro de pontos de uma região muito ampla. Nessa situação, a estratificação pode ser

feita através de *crowdsourcing* (em português, colaboração coletiva), onde voluntários recebem imagens que são por eles classificadas. A acurácia da classificação dos voluntários pode ser medida por meio de imagens de teste (GSARS, 2017).

2.4.1.1 Estratificação em duas fases

Quando o número de unidades amostrais é muito grande, e a estratificação completa dessas unidades torna-se onerosa, uma solução razoável é selecionar uma grande amostra para fins de estratificação numa primeira fase (estratificação incompleta). Em seguida, uma amostra final é selecionada a partir das unidades que foram estratificadas na primeira fase. Essa abordagem tem sido aplicada no projeto LUCAS (EUROPEAN, 2002). Ao contrário de segmentos, a exclusão dos pontos não agrícolas não acarreta em problemas de subestimação.

2.4.1.2 Estratificação em dois estágios

A observação incompleta de segmentos por meio de pontos também pode ser aplicada para fins de estratificação. Determinado número de pontos dentro dos segmentos são classificados e os estratos são construídos com base nos rótulos atribuídos aos pontos, por exemplo, “segmento agrícola com áreas de pastagem”. Uma forma mais simples é observar a proporção de pontos agrícolas dentro dos segmentos, e. g., “altamente agrícola” se proporção de pontos agrícolas $> 70\%$, “agrícola” se $70\% > \text{proporção de pontos agrícolas} > 30\%$ e “não agrícola” se proporção de pontos agrícolas $< 30\%$. Essa abordagem foi aplicada no estudo de Goiana.

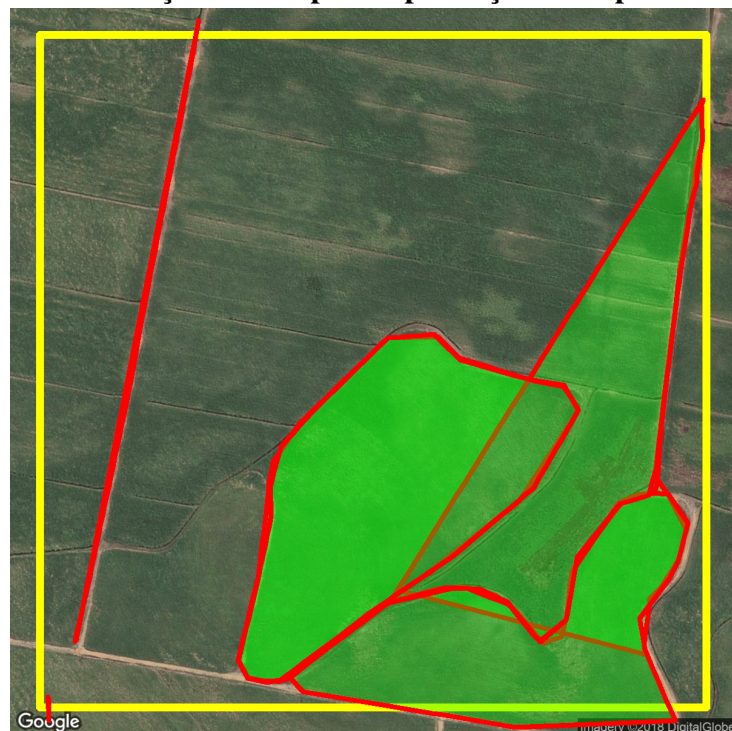
2.4.2 Sistema de posicionamento geográfico

Um GNSS é um sistema baseado em uma rede de satélites de navegação, controlada por estações terrestres, que continuamente transmitem sinais de rádio que, quando capturados por receptores, determinam a sua localização geográfica (longitude, latitude e elevação) na superfície da Terra (MARTINEZ, 2013). O mais antigo e popular sistema GNSS é o GPS. Além de auxiliar

na localização das unidades amostrais e na mensuração de variáveis respostas por observação direta, o GPS é uma excelente ferramenta de controle de qualidade. Ele pode registrar o endereço da sede das propriedades (caso uma lista de propriedades seja construída a partir da amostra de área) ou a localização a partir da qual o pesquisador em campo realizou a observação, o que pode ser confrontado com a real localização do ponto no cadastro (DELINCÉ, 2017). Quando utilizado para delinear os campos de cultivo, o GPS permite checar se os limites do segmento foram respeitados (quando utilizado) ou se houve sobreposição de campos.

A figura 9 mostra campos de produção identificados erroneamente pelo recenseador na etapa da pesquisa de campo, em Goiana. As linhas em amarelo são os limites do segmento e as linhas em vermelho os limites dos campos. As áreas em verde escuro foram medidas (delineadas) mais de uma vez. De fato, dos quatro campos reportados, dois já seriam suficientes para cobrir toda a área de cultivo no segmento. Essa verificação só foi possível graças à utilização de dispositivos de GPS.

Figura 9 – Identificação de campos de produção sobrepostos em Goiana.



3 APLICAÇÃO EM UM ESTUDO AGROPECUÁRIO

3.1 O EXPERIMENTO DA FAO EM GOIANA

Objetivando melhorar as estatísticas agrícolas e rurais, principalmente em países em desenvolvimento, a *Food and Agriculture Organization of the United Nations* (FAO) em conjunto com a *Global Strategy* (GS) tem fomentado a construção de um sistema de cadastros, conhecido como cadastro mestre, que possa ser utilizado nos mais diversos tipos de pesquisas agropecuárias e rurais. Uma das características mais importantes de um cadastro mestre para pesquisas agropecuárias é o potencial para lidar com fazendas, domicílios e parcelas de terra como unidades amostrais. Neste sentido, utilização de cadastros de área tem sido amplamente defendida como parte integrante desse sistema de cadastros (GSARS, 2015).

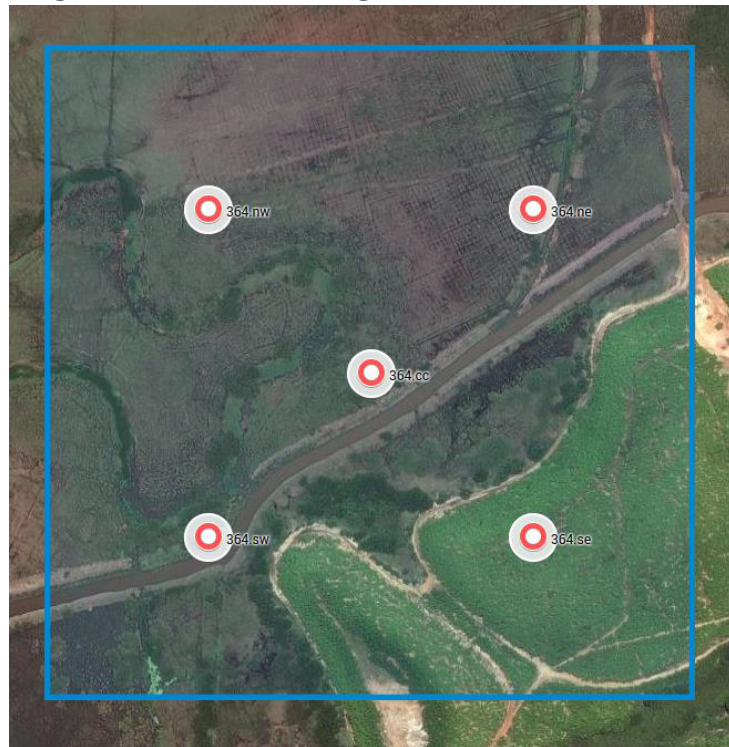
A fim de avaliar aspectos relacionados ao custo, complexidade, viabilidade e eficiência de diferentes tipos de cadastros tomados como um cadastro de amostragem mestre, o GS lançou uma série de experimentos que foram projetados e implementados durante os anos de 2016 e 2017. Um desses experimentos foi conduzido no município de Goiana, região da mata pernambucana, e contou com a utilização de um cadastro de área cobrindo quase toda a área desse município. Nos experimentos realizados no Brasil, uma pesquisa de área foi projetada com foco em métodos de implementação de baixo custo (FERRAZ, 2018).

3.1.1 Cadastro de área

O cadastro de área construído para o experimento consistiu de um grid de segmentos quadrados sobrepostos a área de 445.814 km², correspondente a região de Goiana. Segmentos de 49ha (700m × 700m) foram usados, resultando em um cadastro de 922 segmentos que cobriam quase toda região. Os segmentos que cruzaram as fronteiras do município foram mantidos no cadastro se mais de 50% de suas áreas estivessem internas às fronteiras.

Em cada unidade do cadastro de segmentos, cinco pontos igualmente espaçados foram selecionados, seguindo um padrão pré-fixado de cinco pontos. Isso permitiu estudar desempenhos de amostragem em dois estágios, com subamostragem por pontos. A Figura 10 mostra um exemplo de segmento (delineado por linhas em azuis) em Goiana, com seus respectivos pontos (esferas em branco).

Figura 10 – Pontos no segmento 364 em Goiana-PE.



3.1.2 Estratificação dos segmentos

A estratificação foi realizada usando fotointerpretação para classificar cada ponto dentro de um segmento como “Agrícola” ou “Não agrícola”, dependendo em que tipo de cobertura de solo eles caíssem. A análise do tipo de cobertura do solo foi realizada através de imagens obtidas gratuitamente do Google Earth e Bing. O processo de classificação envolveu seis voluntários trabalhando de maneira não exclusiva, usando crowdsourcing, no qual cada voluntário recebia automaticamente pontos para classificar. Nestas condições, o tempo de trabalho para alcançar a estratificação completa foi de aproximadamente uma semana. Depois da classificação, os segmentos foram categorizados entre os estratos conforme as regras descritas na tabela 1.

Tabela 1 – Definições dos estratos.

Estrato	Descrição
1. Altamente Agrícola	Segmentos com 4 ou 5 pontos classificados como “Agrícola”
2. Agrícola	Segmentos com 2 ou 3 pontos classificados como “Agrícola”
3. Não Agrícola	Segmentos com no máximo 1 ponto classificado como “Agrícola”

Apesar de ser chamado de estrato não agrícola, os segmentos nesse grupo podem, por definição, apresentar algum tipo de atividade agropecuária, sempre que um dos pontos for classificado como agrícola ou quando apresentar áreas de pasto. Não incluir esse estrato no processo de amostragem pode introduzir viés negativo nas estimativas. Uma amostra aleatória de 60 segmentos foi selecionada respeitando a estratificação. O tamanho da amostra foi definido levando em consideração a restrição orçamentária do experimento e uma alocação *ad hoc* foi utilizada para definir o tamanho amostral em cada estrato. A tabela 2 mostra o número de segmentos necessários para cobrir a região de Goiana e alocação da amostra, por estrato. A figura 11, mostra o cadastro de área estratificado para Goiana, onde pode ser identificadas áreas agrícolas (altamente agrícolas e agrícolas) e não agrícolas.

Tabela 2 – Classificação dos segmentos do cadastro de área de Goiana no momentos que a amostra foi selecionada.

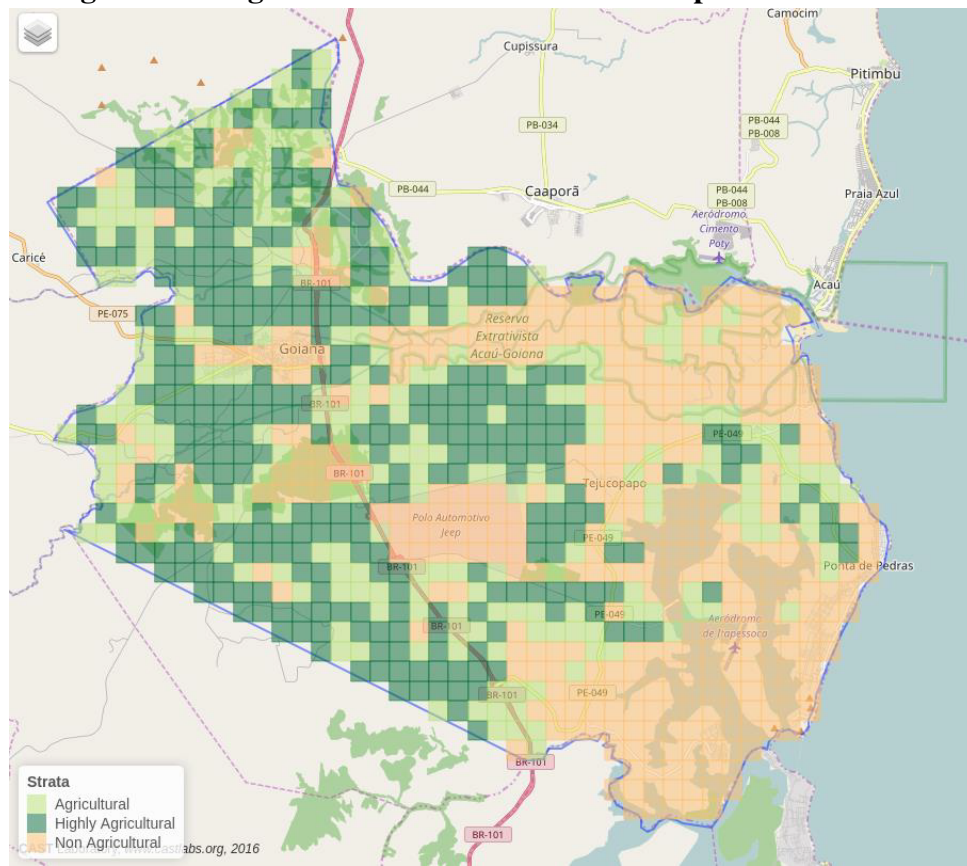
Número de Segmentos	Altamente Agrícola	Agrícola	Não Agrícola	Total
Cadastro	314	232	376	922
Amostra	42	15	3	60

3.1.3 Estimadores

Os estimadores empregados basearam-se nas técnicas de estimação por segmento fechado e ponderado, descritas no capítulo 2. Para que a aplicação e comparação dos estimadores fossem possíveis, duas abordagens distintas foram utilizadas para coletar as variáveis de interesse:

- Abordagem 1: Uma pesquisa de área de segmentos quadrados com observação direta da utilização do solo em cada segmento amostrado. Nesta abordagem, os pesquisadores deviam percorrer o perímetro dos campos com cada tipo de cultivo identificado dentro dos limites do segmento e suas respectivas áreas medidas com o auxílio de um dispositivo GPS. Aqui, a unidade amostral foi o segmento e a unidade de observacional foi o campo de cultivo correspondente a um segmento fechado.

Figura 11 – Segmentos estratificados no município de Goiana-PE.



- **Abordagem 2:** Uma pesquisa de área de segmentos quadrados com subamostragem de pontos em cada segmento para seleção de propriedades. Nesta abordagem, os proprietários foram subamostrados dentro de cada segmento amostrado se os pontos caíssem nos tracts relativos as suas propriedades. Depois de verificar que houve pelo menos um ponto no tract da propriedade, o pesquisador de campo deveria localizar e entrevistar o proprietário. Nesse caso, as unidades amostrais foram os proprietários e a unidades observacionais foram as propriedades.

A primeira abordagem usa a estimação por segmento fechado com observação direta de campos de cultivo. A subamostragem de proprietários por pontos fornece uma maneira de evitar a medição de áreas por dispositivo GPS, usando a definição de segmento ponderado para unidades observacionais. A abordagem 2 fornece uma alternativa para a eficiência na abordagem 1 nos casos em que o número de tracts por segmento é grande. Deve-se salientar, porém, que na segunda abordagem as variáveis mensuraram a área relacionada às declarações dos agricultores, enquanto na primeira abordagem, a variável foi mensurada diretamente.

3.1.3.1 Estimador 1

O Estimador 1 foi construído com base na primeira abordagem. Os estimadores para segmentos fechados, dados em (2.1) e (2.2), foram generalizados para o contexto de amostragem estratificada, com amostragem aleatória simples dentro dos estratos. Considere que y_i seja a soma de todas as áreas no segmento i cultivadas com uma cultura de interesse, então,

$$\hat{\tau}_1 = \sum_{h=1}^H \sum_{i \in S_h} \frac{y_i}{\pi_i} = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} y_i$$

é um estimador não viesado para total de área cultivada na região com a cultura de interesse e

$$\widehat{\text{Var}}(\hat{\tau}_1) = \sum_{h=1}^H \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_h^2 \quad (3.1)$$

é um estimador não viesado para a variância de $\hat{\tau}_1$, onde N_h e n_h são o número de segmentos do estrato h no cadastro e na amostra, respectivamente, s_h é o conjunto, de tamanho n_h , de segmentos do estrato h na amostra e S_h^2 é a variância amostral nos segmentos do estrato h calculada por

$$S_h^2 = \frac{1}{n_h - 1} \left(\sum_{i \in s_h} y_i^2 - N_h \bar{y}_{s_h}^2 \right).$$

Estimar a área total para uma cultura de interesse é um problema de estimativa de domínio. Para calcular a variância correta usando o estimador de variância (3.1), é necessário definir $y_i = 0$ sempre que no segmento i não houver áreas com a cultura de interesse.

3.1.3.2 Estimador 2

A elaboração do Estimador 2 foi baseada na segunda abordagem como alternativa em situações onde o número de campos dentro de cada segmento é muito grande, cenário onde a aplicação da abordagem 1 é ineficiente devido a grande quantidade de trabalho de campo. Dentro dos segmentos selecionados em um primeiro estágio, os proprietários foram subamostrados P_i vezes, onde P_i é número de pontos agrupados dentro do segmento i . Assim, para cada segmento na amostra, foi computada a variável

$$\hat{X}_i = \frac{T_i}{P_i} \sum_{j=1}^{P_i} \frac{y_j}{A_j}$$

em que y_j é uma variável de interesse. Para uma amostra estratificada, com amostra aleatória simples dentro dos estratos, o estimador para o total de y pode ser obtido de (2.4):

$$\hat{\tau}_2 = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{T_i}{P_i} \sum_{j=1}^{P_i} \frac{y_j}{A_j}.$$

Aos pontos que caíram em áreas não agrícolas ou em propriedades que não apresentavam associação com a variável de interesse foram imputados valores zero. No primeiro estágio do experimento, uma amostragem aleatória simples foi aplicada aos segmentos dentro de cada estrato. No segundo estágio, as propriedades foram selecionadas através de pontos em uma amostragem com reposição. Dessa forma, o estimador para a variância dado em (2.5) pode ser estendido para o Estimador 2 como

$$\widehat{\text{Var}}(\hat{\tau}_2) = \sum_{h=1}^H \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_{Xh}^2 + \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{1}{P_i(P_i - 1)} \sum_{j=1}^{P_i} \left(\frac{y_j}{A_j} T_i - \hat{X}_i\right)^2, \quad (3.2)$$

em que

$$S_{Xh}^2 = \frac{1}{n_h - 1} \sum_{i \in s_h} \left(\hat{X}_i - \bar{\hat{X}}_h\right)^2 \quad \text{e} \quad \bar{\hat{X}}_h = \frac{1}{n_h} \sum_{i \in s_h} \hat{X}_i.$$

3.1.3.3 Estimador 3

O Estimador 2 foi adaptado para considerar apenas pontos que recaíram em áreas agrícolas. Entretanto, como os pontos foram selecionados através de um padrão pré-fixado, o número de pontos agrícolas foi menor ou igual ao número de pontos subamostrados, $F_i < P_i$. Dessa forma, o Estimador 3, derivado de (2.6), é dado por

$$\hat{\tau}_3 = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{G_i}{F_i} \sum_{j=1}^{F_i} \frac{y_j}{A_j},$$

em que, $G_i = \sum_{j \in s_i} T_{ij}$ é o total de área agrícola no segmento i e

$$\widehat{\text{Var}}(\hat{\tau}_3) = \sum_{h=1}^H \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_{Xh}^2 + \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{1}{F_i(F_i - 1)} \sum_{j=1}^{F_i} \left(\frac{y_j}{A_j} G_i - \hat{X}_i\right)^2 \quad (3.3)$$

é um estimador não viesados para $\text{Var}(\hat{\tau}_3)$.

Nos segmentos que registraram apenas um ponto agrícola a variância foi calculada fazendo $F_i(F_i - 1) = 1$. Em segmentos onde não houve pontos agrícolas uma variável fictícia

com valor zero foi computada, $\hat{X}_i = 0$. Além do mais, no experimento não foram mensuradas as áreas agrícolas operadas (tracts) e os valores dos G_i 's foram aproximados pela área total dos campos de cultivo em cada segmento. Isto pode ter causado uma subestimação para o estimador.

3.1.4 Resultados do experimento

As estimativas foram obtidas a partir de uma amostra estratificada de 60 segmentos selecionados do cadastro de área de 922 segmentos. Dos segmentos selecionados, 49 possuíam algum tipo de cultivo, alguns segmentos mais de um tipo. Estes segmentos são chamados de segmentos efetivos. A tabela 3 mostra, por estrato, o número de segmentos efetivos na amostra.

Tabela 3 – Distribuição dos segmentos amostrados e efetivos, por estrato.

Número de Segmentos	Altamente Agrícola	Agrícola	Não Agrícola	Total
Amostra	42	15	3	60
Efetivos	35	12	2	49

3.1.4.1 Estimativas de área cultivada com cana-de-açúcar

Em Goiana, a principal cultura é a cana-de-açúcar. Essa cultura foi encontrada em 75% dos segmentos amostrados e em 92% da amostra efetiva. A alta frequência de observação permitiu o uso dos três estimadores propostos para estimar o total de área cultivada. Oito propriedades registraram o cultivo de cana-de-açúcar. A tabela 4 mostra estatísticas de frequência para essas propriedades, onde percebe-se que uma propriedade é responsável por grande parte desse tipo cultivo (máximo).

Tabela 4 – Estatísticas de frequência das unidades amostrais associadas às propriedades com cana-de-açúcar, em Goiana.

	Número de pontos	Número de Segmentos
Mínimo	1	1
Máximo	153	37
Mediana	4,5	1
Média	23,25	5,75

A tabela 5 discrimina por estrato e estimador: a estimativa para a área total com cana-de-açúcar em hectares (Total), o desvio padrão dessa estimativa (DP) e o coeficiente de variação (CV). O número de segmentos efetivos com cana-de-açúcar foi 45, com 34 no estrato altamente agrícola, 10 no estrato agrícola e 1 no estrato não agrícola. O desvio padrão para o Estimador 1 foi calculado usando apenas segmentos efetivos com cana-de-açúcar e, portanto, utilizou-se a estimação por domínio.

Tabela 5 – Estimativas para o total de área cultivada com cana-de-açúcar, por estrato e estimador, em Goiana.

Estrato	Estimador	Total	DP	CV
Altamente Agrícola	Estimador 1	7.115	658,7	0,09
	Estimador 2	7.169	665,9	0,09
	Estimador 3	4.778	616,5	0,13
Agrícola	Estimador 1	3.535	726,2	0,21
	Estimador 2	3.441	871,5	0,25
	Estimador 3	2.082	554,5	0,27
Não Agrícola	Estimador 1	177	0,0	0,0
	Estimador 2	774	774,3	1,0
	Estimador 3	111	111,4	1,0
Total	Estimador 1	10.828	980,4	0,12
	Estimador 2	11.383	1.342,6	0,12
	Estimador 3	6.972	836,6	0,12

O Estimador 2 obteve a maior estimativa para o total de área cultivada com cana-de-açúcar em Goiana, 5,13% e 63,28% maior em relação as estimativas dos estimadores 1 e 3, respectivamente. Contudo, a estimativa da variância também foi maior para o Estimador 2. Nos segmentos não agrícolas da amostra, apenas um ponto foi classificado como agrícola e, portanto, uma única observação teve valor diferente de zero. Isto acarretou no aumento da estimativa da variância para esse estrato.

O Estimador 3 apresentou a menor estimativa de total de área cultivada, provavelmente devido a aproximação das áreas agrícolas pelo total de área dos campos de cultivo em cada segmento. Isto deve ter causado uma subestimação do total de área com cana-de-açúcar. Por outro lado, o desvio padrão foi o menor entre os estimadores, justamente por esse estimador utilizar apenas pontos classificados como agrícola.

Os estimadores para a variância dos estimadores 2 e 3, dados em (3.3) e (3.2), são compostos por duas parcelas: a primeira é a contribuição do 1º estágio do plano amostral na estimativa da variância; a segunda é a contribuição do 2º estágio do plano amostral na estimativa da variância. A tabela 6 mostra, por estrato, a contribuição de cada parcela na estimativa da variância dos estimadores 2 e 3.

Tabela 6 – Contribuição para a estimativa da variância do total de área estimada com cana-de-açúcar em Goiana, por estimador e estágio.

	Estágio	Total	Altamente Agrícola	Agrícola	Não Agrícola
Estimador 1	1º estágio	1.790.764,61	439.888,4	756.054,7	594.821,49
	2º estágio	11.756,18	3.575,6	3.396,5	4.784,09
Estimador 2	1º estágio	577.261,7	286.625,43	278.331,74	12.304,52
	2º estágio	122.602,6	93.394,36	29.109,29	98,96

3.1.4.2 Estimativas da quantidade de fertilizantes

Visando exemplificar o uso dos estimadores para variáveis que não sejam áreas de cultivo, foram obtidas estimativas para o total de fertilizante utilizado pelos produtores agrícolas em Goiana. A informação sobre a utilização de fertilizantes pôde ser obtida em 47 dos segmentos na amostra. As tabelas a seguir sumarizam as principais estimativas, com os valores expressos em toneladas.

A informação da quantidade de fertilizante não pode ser obtida nos tracts através da observação direta. Para que a aplicação do Estimador 1 fosse possível, o valor informado pelos proprietários para a quantidade de fertilizante utilizada em toda a propriedade foi multiplicada pela razão entre as áreas de cultivo no segmento e a área com operações agrícolas na propriedade. A tabela 7 mostra as estimativas para o total de fertilizantes, em toneladas, utilizados em Goiana, por estrato e estimador.

O Estimador 2 apresentou a maior estimativa para o total de fertilizantes entre todos os estimadores, com estimativa 72,26% e 63,06% maior em relação as estimativas dos estimadores 1 e 3, respectivamente. Esse estimador apresentou maior estimativa de variância, pela mesma razão supracitada na estimação de cana-de-açúcar.

Tabela 7 – Estimativas para o total de fertilizantes (em toneladas) utilizado em Goiana, por estrato e estimador.

Estrato	Estimador	Total	DP	CV
Altamente Agrícola	Estimador 1	1.832	442,12	0,24
	Estimador 2	3.483	372,2	0,11
	Estimador 3	2.329	331,2	0,14
Agrícola	Estimador 1	1.257	532,3	0,42
	Estimador 2	1.880	469,7	0,25
	Estimador 3	1.160	307,1	0,26
Não Agrícola	Estimador 1	271	156,5	0,58
	Estimador 2	426	425,9	1,0
	Estimador 3	61	61,2	1,0
Total	Estimador 1	3.361	709,5	0,21
	Estimador 2	5.789	735,2	0,13
	Estimador 3	3.550	455,8	0,13

O Estimador 3 obteve estimativas para a quantidade total de fertilizantes próximas as estimativas do Estimador 1, 5,64% maior. Isto pode indicar que a estratégia adotada para tornar o Estimador 1 aplicável na estimação da quantidade de fertilizantes resultou em uma subestimação desse total.

A contribuição de cada estágio do plano amostral para a estimativa da variância dos estimadores 2 e 3 é mostrada na tabela 8. A estimativa da variância para o Estimador 3 foi a menor entre os três estimadores.

Tabela 8 – Contribuição para a estimativa da variância do total da quantidade de fertilizantes estimado (em toneladas) em Goiana, por estimador e estágio.

	Estágio	Total	Altamente Agrícola	Agrícola	Não Agrícola
Estimador 1	1º estágio	537.607,39	137.973,18	219.700,70	179.933,50
	2º estágio	2.974,28	561,11	965,98	1.447,18
Estimador 2	1º estágio	174.918,15	85.467,56	85.728,46	3.722,11
	2º estágio	32.806,60	24.200,73	8.575,92	29,93

3.1.5 Eficiência da estratificação

O pressuposto da amostragem estratificada é que a variância dos elementos dentro

dos estratos é pequena e, entre os estratos, a variância é alta. O critério para avaliar a qualidade de uma estratificação é a diminuição atingida nas variâncias das estimativas. A eficiência da estratificação é calculada como uma razão

$$\widehat{Eff} = \frac{\widehat{Var}_{Noest}(\hat{\tau})}{\widehat{Var}_{est}(\hat{\tau})},$$

em que, $\widehat{Var}_{est}(\hat{\tau})$ é a variância estimada sob estratificação, dada por (2.2), e

$$\widehat{Var}_{Noest}(\hat{\tau}) = \left(\frac{N}{n} - 1 \right) \sum_{h=1}^H N_h [S_h^2 + (\bar{y}_{s_h} - \bar{y})^2]$$

é a variância estimada sem estratificação. A justificativa para o uso da estimativa da variância do Estimador 1 para o computo da variância sob a estratificação é que os estratos foram construídos com base em segmentos.

A eficiência da estratificação fornece uma medida da economia feita em termos de tamanho da amostra: sem estratificação, precisaríamos de um tamanho de amostra de $n \times Eff$ em vez de n para obter a mesma precisão. A tabela 9 mostra eficiências de 1,2 e 3,2, permitindo concluir que a estratificação realizada foi eficiente. Na estimativa da área de cana-de-açúcar, por exemplo, o resultado significa que, se não fosse feita uma estratificação, o tamanho da amostra necessária para alcançar a mesma precisão observada no campo seria três vezes maior do que a utilizada.

Tabela 9 – Eficiência da estratificação obtida no experimento de Goiana.

	Variância Estratificada	Variância Não Estratificada	Eficiência
Cana-de-açúcar	$9,6127 \times 10^5$	$3,1053 \times 10^6$	3,2305
Fertilizantes	$5,0333 \times 10^5$	$5,9857 \times 10^5$	1,1892

4 AVALIAÇÃO NUMÉRICA DOS ESTIMADORES

Os desempenhos dos estimadores aplicados no experimento de Goiana foram avaliados através de simulações de Monte Carlo. As comparações entre estimadores de área são baseadas em sua aplicabilidade, eficiência na amostragem e suscetibilidade a erros não amostrais (FORD *et al.*, 1986). Nesse estudo numérico são avaliados os desempenhos de três planos amostrais (estimadores), considerando diferentes tamanhos de segmentos. Outro aspecto levado em consideração nesta avaliação numérica é a densidade de produtores na região de interesse. As simulações aqui apresentadas fazem a comparação da eficiência em termos de precisão dos estimadores através do computo do viés e erro quadrático médio (EQM). Aspectos como a subjetividade em respostas dos proprietários, erros de mensuração e estratificação incorreta não são tratados nesta dissertação. O estudo foi realizado utilizando o ambiente de software livre para computação gráfica e estatística R, versão 3.4.4 (R CORE TEAM, 2017).

4.1 PROTOCOLO DE SIMULAÇÃO

Preliminar ao experimento de Monte Carlo foi necessário gerar uma população de campos produtivos com características espaciais intrínsecas à amostragem de área. A criação e manipulação de objetos espaciais, como polígonos, linhas e pontos, dentre outros, foi realizada utilizando o pacote SP (PEBESMA; BIVAND, 2005), operações topológicas de interseção e união de objetos espaciais foram realizadas utilizando o pacote RGEOS (BIVAND; RUNDEL, 2017) e operações de projeção e transformação de objetos espaciais no sistema de coordenadas geográficas feitas através do pacote RGDAL (BIVAND; KEITT; ROWLINGSON, 2018), todos estes pacotes disponíveis no R. A partir da criação de uma população-alvo fictícia, foram construídos cadastros de área de segmentos e de pontos dentro de segmentos, nos quais foram mensurados em todas as unidades amostrais os valores das variáveis de interesse, aplicando as duas abordagens utilizadas em Goiana.

4.1.1 Geração de uma população de campos

O ponto de partida desta simulação é gerar uma população de explorações agrícolas na qual os totais das variáveis de interesse podem ser determinados. São enumeradas abaixo as principais etapas seguidas na tentativa de emular uma população agrícola próxima da realidade:

1. Normalmente os campos de produção apresentam formas geométricas complexas. Para reproduzir essa característica, e por ser uma solução simples, os campos foram representados por polígonos gerados a partir das coordenadas geográficas de municípios brasileiros, com uma redução de escala adequada. Dados espaciais dessa natureza são distribuídos gratuitamente e podem ser facilmente obtidos¹.
2. Os polígonos gerados foram posicionados obedecendo apenas à restrição de que os seus respectivos centroides estivessem dentro dos limites espaciais de Goiana. Isso permite emular campos que ultrapassam os limites de Goiana. A fim de introduzir correlação espacial, cada polígono foi posicionado de acordo com a sua dimensão; se a área do polígono k pertencia ao intervalo I_a , então a posição do centroide do polígono k foi dada através da distribuição normal bivariada $(X, Y) \sim N_2(\mu_a, \Sigma_a)$, onde μ_a é um ponto aleatoriamente escolhido interno à região de Goiana. Os valores de (x, y) são gerados até que não haja interseção entre os polígonos.
3. Por fim são associadas informações aos polígonos. Para atribuir proprietários aos polígonos foram sorteados com reposição números entre $1, \dots, P$ com probabilidades de seleção $p_k = u_k / \sum_{k=1}^P u_k$, em que, $u_k \sim U(0, 0.25)$, $k = 1, \dots, P$, e P é o número máximo de proprietários. Os números sorteados representaram os proprietários dos polígonos e, então, uma propriedade pôde ser formada por vários polígonos. Um campo de cultivo pode apresentar mais de um tipo de cultivo. A distribuição multinomial foi utilizada para determinar quantos tipos de cultivo os polígonos teriam e a proporção de área plantada com cada cultura definida aleatoriamente através da distribuição uniforme (com a soma das proporções igual a 1). A tabela 10 mostra as probabilidades de associação entre os tipos de cultivo e os polígonos.

¹ <<https://gadm.org/>>

Tabela 10 – Probabilidades de associação entre tipos de cultivo e um polígono.

Probabilidade (%)	Tipo de cultivo
70	Apenas cultura I
10	Apenas cultura II
10	Mistura de cultura II e cultura III
5	Apenas cultura IV
5	Mistura de cultura II, cultura III e cultura IV

4.1.2 Geração dos cadastros de área

Com apenas os limites territoriais da região de interesse é possível realizar uma divisão espacial em segmentos quadrados. Entretanto, construir um cadastro de área de segmentos com uma estratificação semelhante à aplicada em Goiana requer a sobreposição de uma camada de informações espaciais proveniente de uma população. Os procedimentos feitos na construção dos cadastros bem como a estratificação das unidades cadastrais estão descritas a seguir:

1. Com base apenas nos limites territoriais de Goiana foram geradas duas camadas de objetos espaciais: a primeira composta por segmentos quadrados, podem ser vistos como polígonos espaciais quadrados, que particionam toda a região, e a segunda camada composta por pontos agrupados dentro dos segmentos. Os segmentos nas bordas da região que tinham área de interseção com a região menores que 50% foram retirados do cadastro e, nos centroides dos segmentos restantes, uma configuração de cinco pontos foi sobreposta, um ponto posicionado sobre o centroide e os demais pontos posicionados nas mediatrizes das retas definidas do centroide aos vértices.
2. Um fator avaliado através da simulação foi a influência do tamanho do segmento na precisão dos estimadores. Foram considerados três cadastros de segmentos quadrados com arestas $d \in \{1000\text{m}, 700\text{m}, 400\text{m}\}$; a área dos segmentos, para os valores de arestas considerados, são 100ha, 49ha e 16ha, respectivamente. O número de unidades amostrais é função da aresta do segmento, $N_d = f(d)$. Para esses valores de aresta o cadastro gerado foi composto por 443, 914 e 2796 segmentos, respectivamente.
3. A estratificação dos segmentos utilizou as localizações espaciais dos polígonos da população gerada nas etapas anteriores. Em cada segmento os pontos que estavam localizados dentro de algum polígono foram rotulados como agrícolas e, então, o segmento foi estratificado de acordo com o número de pontos agrícolas: se 4 ou 5 pontos, o segmento foi classificado como altamente agrícola, se 2 ou 3 pontos, o segmento foi classificado como agrícola, se 1 ou menos pontos o segmento foi classificados como não agrícola.

4.1.3 Extração de variáveis resposta

As duas abordagens de observação das variáveis respostas aplicadas em Goiana também foram emuladas e seus valores computados em todas as unidades amostrais, tanto segmentos quanto pontos, a fim de tornar a simulação de Monte Carlo mais rápida.

1. Abordagem 1: para cada segmento do cadastro foi verificado se havia interseção com os polígonos da população fictícia. Caso houvesse, a área no segmento com a cultura de interesse seria proporcional à fração de área plantada nos polígonos com essa cultura. Também se somou a área de todas as interseções de polígonos com cada segmento para computar a área agrícola utilizada, necessária para o uso do estimador 3; neste caso há equivalência entre estas duas medidas de área.
2. Abordagem 2: uma cadastro de lista foi construído com as propriedades da população fictícia, somando as áreas de plantio com cada cultura em todos os polígonos que tinham o mesmo proprietário. Para cada ponto no segmento foi verificado se ele estava localizado dentro de algum polígono da população. Caso estivesse, o proprietário do polígono era identificado e as informações da propriedade associadas ao ponto.

4.2 EXPERIMENTO DE MONTE CARLO

Estudos de Monte Carlo com 10.000 réplicas foram conduzidos a fim de avaliar a precisão dos estimadores em vários cenários, construídos com base em contextos realísticos. A saber, os cenários são constituídos de populações com (i) múltiplos campos com tamanhos semelhantes e alta densidade espacial e (ii) múltiplos campos com tamanhos distintos, alguns de grande porte e outros de pequeno porte. A justificativa para a escolha dessas configurações de populações está na heterogeneidade agrícola brasileira, composta em grande parte por grandes latifúndios nas regiões centrais do país e por pequenas propriedades em áreas mais litorâneas. Em cada réplica de Monte Carlo, uma amostra estratificada de segmentos foi retirada utilizando o pacote `SAMPLING` do R (TILLÉ; MATEI, 2016), com tamanho amostral em cada estrato proporcional a fração amostral aplicada no estudo em Goiana, e foram computadas as estimativas

para o total e para a variância do estimador do total. Os estimadores foram avaliados no tocante a estimação de totais de área cultivada.

Com as réplicas de Monte Carlo foram computadas, para os estimadores apresentados na seção 3.1.3, as médias das estimativas do total bem como a média das variâncias dos estimadores, pois são estimativas numéricas para o valor esperado de $\hat{\tau}$ e $\widehat{\text{Var}}(\hat{\tau})$, respectivamente. Também foram computados numericamente o viés e viés relativo, expresso em percentual, para mensurar a proximidade com o valor esperado dos estimadores (total populacional e variância do estimador). O desvio padrão foi computado para avaliar como mudanças na amostra afetam os estimadores. Apresenta-se também o EQM que faz o *trade-off* entre o viés e a variância dos estimadores. O estimador de Monte Carlo para o valor esperado dos estimadores é dado por

$$\hat{\tau}_{mc} = \hat{E}(\hat{\tau}) = \sum_{i=1}^R \frac{\hat{\tau}_i}{R} \quad (4.1)$$

e a variância desse estimador é

$$\text{Var}(\hat{\tau}_{mc}) = \frac{\sum_{i=1}^R (\hat{\tau}_i - \hat{\tau}_{mc})^2}{R}, \quad (4.2)$$

em que, $\hat{\tau}_i$ é a estimativa do total para a i -ésima réplica, R é o número de réplicas e $E(.)$ é o operador de esperança matemática. De maneira análoga, o estimador para a variância do estimador do total é

$$\hat{\nu}_{mc} = \hat{E}[\widehat{\text{Var}}(\hat{\tau})] = \sum_{i=1}^R \frac{\widehat{\text{Var}}(\hat{\tau}_i)}{R} \quad (4.3)$$

e sua variância é

$$\text{Var}(\hat{\nu}_{mc}) = \frac{\sum_{i=1}^R [\widehat{\text{Var}}(\hat{\tau}_i) - \text{Var}(\hat{\tau}_{mc})]^2}{R}, \quad (4.4)$$

em que $\widehat{\text{Var}}(\hat{\tau}_i)$ é a estimativa da variância do estimador do total na i -ésima réplica. Os desvios padrão de (4.2) e (4.4) também foram computados. Os vieses para (4.1) e (4.3) foram computados, respectivamente, por

$$\text{Viés}(\hat{\tau}_{mc}) = \hat{\tau}_{mc} - \tau$$

e

$$\text{Viés}(\hat{\nu}_{mc}) = \hat{\nu}_{mc} - \text{Var}(\hat{\tau}_{mc}).$$

Com essas duas quantidades foram computados os EQM's e os vieses relativos.

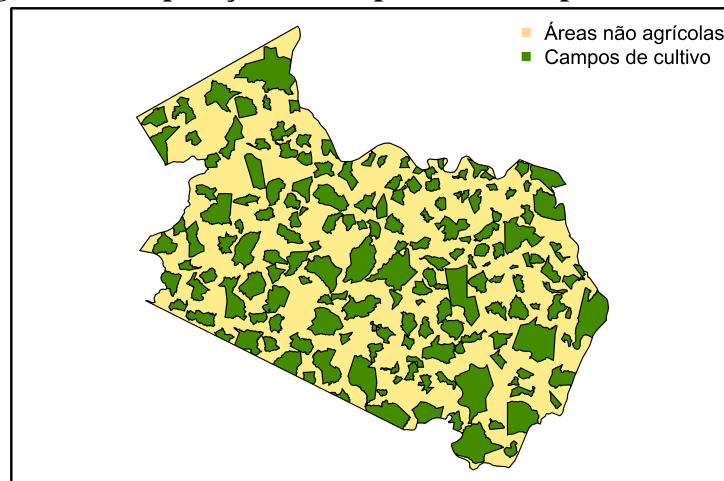
4.2.1 Cenários gerados

Tomando por base o município de Goiana, duas populações de produtores foram geradas através dos procedimentos descritos acima. As principais diferenças entre as populações geradas são a quantidade de campos de produção e o número de proprietários.

Cenário 1

No primeiro cenário, uma população de 173 polígonos, referentes a municípios pernambucanos com uma redução de escala de 40 vezes, foram associados a 72 proprietários, dos quais 29 estavam associados a um único polígono e 7 foi o máximo de polígonos com um mesmo proprietário. A figura 12 mostra a população de campos gerada com área total de cultivo igual a 16.785,54ha, 37,64% da área total da região, distribuída em 12.344,07ha (73,54%) com a cultura I, 2.866,84ha (17,08%) com cultura II, 1.403,09ha (8,36%) com a cultura III e 171,35ha (1,02%) com a cultura IV.

Figura 12 – População de campos simulada para o cenário 1.



A área das 72 propriedades geradas apresentaram assimetria a esquerda, como visto na figura 13, isso também refletiu em assimetria para as áreas cultivadas. A figura 14 mostra assimetria a esquerda para as áreas nas propriedades com a cultura I e II.

Figura 13 – Distribuição de frequência das áreas das propriedades no cenário 1.

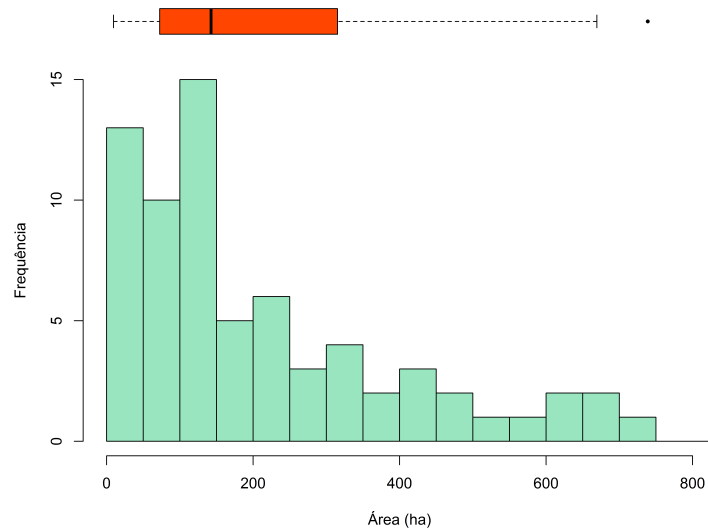
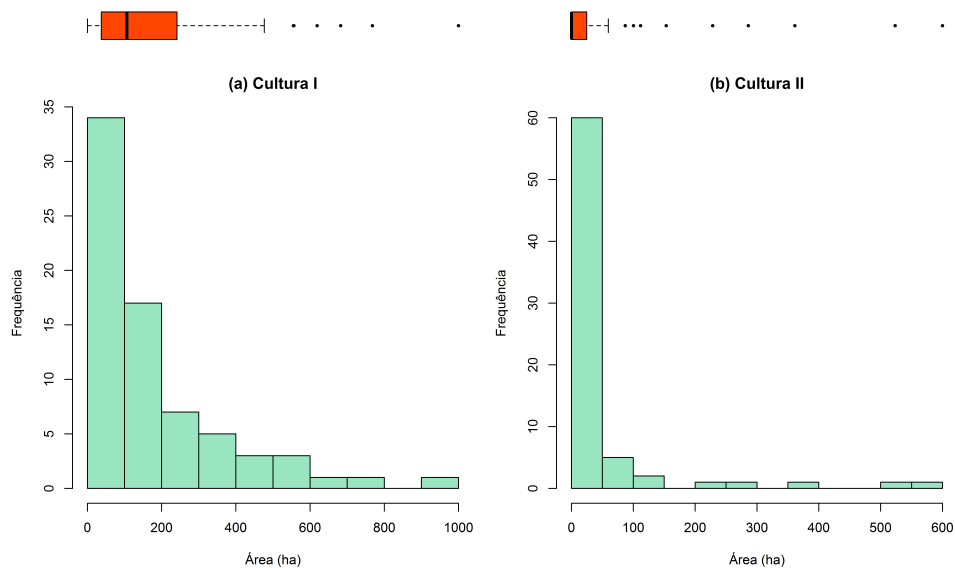


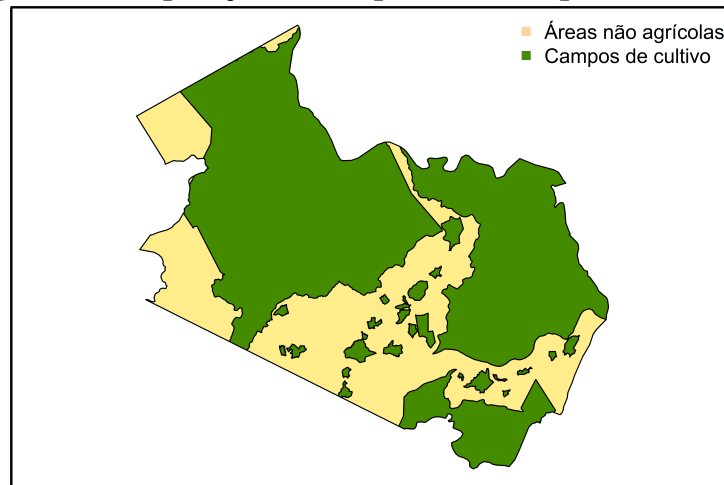
Figura 14 – Distribuição de frequência das áreas das propriedades com as culturas I e II no cenário 1.



Cenário 2

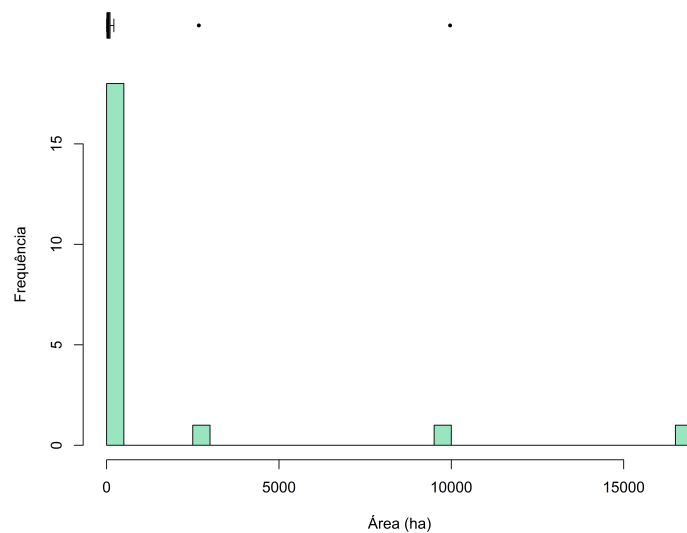
No segundo cenário foi gerada uma população de 28 polígonos, 3 com uma redução de escala de 4 vezes e 25 com redução de 35 vezes, que foram associados a 21 proprietários. A área total de cultivo na população foi de 29.749,47ha, 66,73% da área total da região, distribuída em 29.164,10ha (98,03%) com cultura I, 453,57ha (1,52%) com cultura II, 127,41ha (0,42%) com cultura III e 4,40ha (0,01%) com cultura IV. Os polígonos menores foram introduzidos nesse cenário (figura 15) para representar pequenos produtores que, geralmente, são negligenciados em cadastros de lista contendo apenas grandes propriedades.

Figura 15 – População de campos simulada para o cenário 2.



Neste cenário, um comportamento assimétrico das áreas das propriedades já era esperado, devido a redução de escala entre polígonos serem diferentes. A figura 16 mostra a distinção entre as áreas das propriedades que foram associadas aos três polígonos com menor redução de escala.

Figura 16 – Distribuição de frequência das áreas das propriedades no cenário 2.



4.3 RESULTADOS NUMÉRICOS

Para cada cenário, foram construídos três cadastros de segmentos quadrados, com diferentes tamanhos de aresta. Uma amostra estratificada de segmentos foi retirada a partir de

cada cadastro, utilizando a mesma fração amostral aplicada no estudo em Goiana. A tabela 11 mostra os tamanhos amostrais, por estrato, que foram retiradas nesta simulação para as diferentes dimensões de segmento.

Tabela 11 – Número de segmentos na amostra, por estrato, para cada valor de aresta considerado.

Aresta do segmento	Estrato			Total
	Altamente agrícola	Agrícola	Não agrícola	
$d = 1000$	21	8	2	31
$d = 700$	42	15	3	60
$d = 400$	128	46	10	184

4.3.1 Cenário 1

Devido à retirada de segmentos que possuísem menos de 50% de área dentro da região, o total de área cultivada coberta pelo cadastro foi menor que o total de área cultivada na região. Para $d = 1000$, os segmentos continham 97,03% da área cultivada na região, 76,81% em segmentos classificados como agrícola e altamente agrícola. Para $d = 700$, os segmentos continham 98,27% da área cultivada na região, 82,50% em segmentos classificados como agrícola e altamente agrícola. Para $d = 400$, os segmentos continham 98,98% da área cultivada na região, 88,41% em segmentos classificados como agrícola e altamente agrícola. A redução da dimensão do segmento resultou em diminuição de áreas agrícolas identificadas em segmentos não agrícolas. A tabela 12 mostra os resultados da estratificação para diferentes valores de aresta.

Tabela 12 – Número de segmentos por estrato, cenário 1.

Aresta do segmento	Estrato			Total
	Altamente agrícola	Agrícola	Não agrícola	
$d = 1000$	75	157	211	443
$d = 700$	219	233	465	917
$d = 400$	789	458	1549	2796

As tabelas a seguir resumem as estimativas obtidas para o total e para a variância dos estimadores do total. Para as culturas mais frequentes (cultura I e cultura II) o Estimador 1, com $d = 1000$ e $d = 700$, apresentou um viés relativo (negativo) próximo a diferença entre os totais nos segmentos e os totais na região. Para as culturas menos frequentes o viés do Estimador

1 foi um pouco menor, com uma pequena sobrestimação. Contudo, o Estimador 1 mostrou maior sensibilidade à mudanças na amostra, apresentando o maior desvio padrão entre os três estimadores.

Embora o Estimador 2 tenha apresentado em alguns casos um maior viés, se comparado ao Estimador 1, o desvio padrão mostra que esse estimador apresenta menor sensibilidade quanto à mudanças na amostra; seu EQM foi o menor em todos os casos. Isso mostrou que se um (possível) maior viés for aceitável, o Estimador 2 é a melhor escolha de estimador para os totais de área. O Estimador 3 não apresentou resultados satisfatórios, tanto em termos de viés quanto em termos de dispersão.

No geral, com o tamanho de aresta de 400 metros houve um aumento nos vieses dos estimadores. Isto pode ser decorrente da fração amostral aplicada ao estrato não agrícola. Além disso, O impacto do tamanho dos segmentos sobre a dispersão dos estimadores não pôde ser avaliada isoladamente por se confundir com o efeito do aumento do tamanho da amostra, 3 e 6 vezes maior se comparada a amostra com segmentos de arestas iguais a 700 e 1000 metros, respectivamente. Os resultados obtidos para a variância dos estimadores 2 e 3 estão em consonância com o que diz (NEALON, 1984), a estimação por segmento ponderado reduz a variância das estimativas.

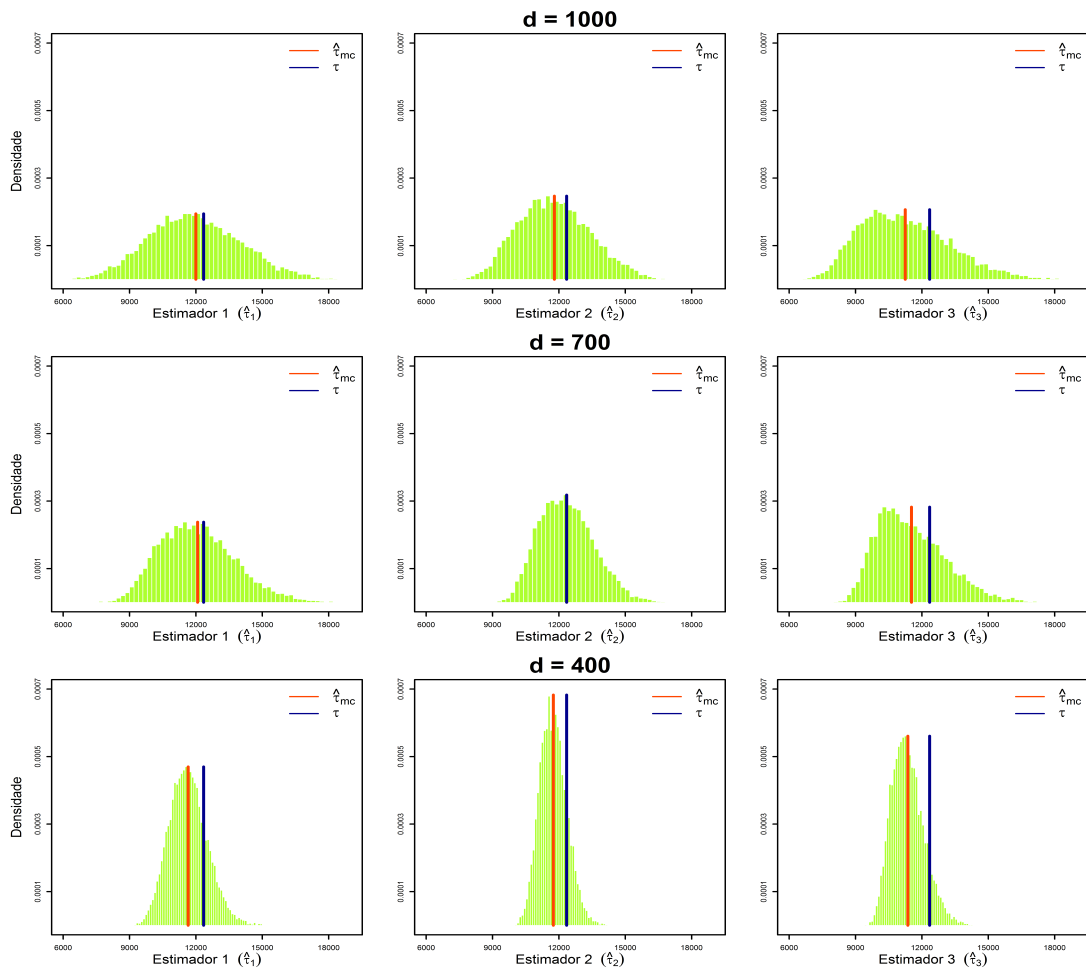
O total populacional de área cultivada com a cultura I foi 12.344,07ha. Essa foi a cultura mais predominante, sessenta e seis propriedades tinham áreas com a cultura I. A tabela 13 mostra que o Estimador 2 apresentou os melhores resultados em termos de EQM. Chama atenção que todos estimadores apresentaram viés negativo, com exceção do Estimador 2 para $d = 700$. Para os estimadores 2 e 3, este último fato pode estar relacionado com a assimetria das variável resposta. Todos estimadores apresentaram aumento de viés para $d = 400$.

A figura 17 mostra a densidade empírica dos estimadores da variância para a cultura I, onde as colunas detalham a densidade por estimador e as linhas detalham por tamanho de aresta dos segmentos. A densidade numérica do Estimador 3 mostra assimetria positiva acentuada para $d = 1000$ e $d = 700$, mas com o aumento do tamanho da amostra (para $d = 400$) houve redução dessa assimetria. Outro efeito observado para $d = 400$ foi a redução da dispersão das estimativas em torno de $\hat{\tau}_{mc}$. Contudo, a média de todas as densidades se distanciou do total populacional,

Tabela 13 – Estimativas para o total de área com a cultura I, cenário 1.

	$\hat{\tau}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	11.955,19	-388,88	-3,15	2.056,94	4.382.235,03
Estimador 2	11.791,51	-552,56	-4,48	1.602,96	2.874.817,10
Estimador 3	11.242,83	-1101,24	-8,92	2.010,03	5.252.954,00
$d = 700$					
Estimador 1	12.072,31	-271,76	-2,20	1.692,69	2.939.067,72
Estimador 2	12.349,49	5,41	0,04	1.236,10	1.527.977,68
Estimador 3	11.524,48	-819,59	-6,64	1.543,62	3.054.500,73
$d = 400$					
Estimador 1	11.649,72	-694,35	-5,62	864,65	1.229.739,88
Estimador 2	11.745,12	-598,95	-4,85	610,62	731604,33
Estimador 3	11.359,47	-984,60	-7,98	734,90	1.509.516,15

causando um aumento no viés numérico. Também pode ser visto que o Estimador 1 apresenta a maior dispersão entre os estimadores

Figura 17 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura I, cenário 1.

A tabela 14 mostra as estimativas das variâncias dos estimadores, expressas em (2.2), (2.5) e (2.7). Como visto na figura 17, o Estimador 1 apresentou a maior dispersão das

estimativas de total com a cultura I, o que foi comprovado através das suas estimativas de variância. O estimador para a variância do Estimador 2 apresentou menores, se comparado ao Estimador 1, com vieses positivos para todos os valores de aresta. O estimador para a variância do Estimador 3 apresentou o menores resultados para $d = 100$ e $d = 700$, mas nestes casos os vieses (negativos) foram muito grandes, i. e., o estimador teve sua variância subestimada.

Tabela 14 – Estimativas para a variância do total estimado de área com a cultura I, cenário 1.

	$\hat{v}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	$4,2 \times 10^6$	$-7,3 \times 10^4$	-1,72	$1,3 \times 10^{13}$	$1,3 \times 10^{13}$
Estimador 2	$2,7 \times 10^6$	$8,4 \times 10^4$	3,26	$3,7 \times 10^{12}$	$3,7 \times 10^{12}$
Estimador 3	$1,4 \times 10^6$	$-2,6 \times 10^6$	-64,64	$1,1 \times 10^{13}$	$1,7 \times 10^{13}$
$d = 700$					
Estimador 1	$2,8 \times 10^6$	$-7,9 \times 10^4$	-2,76	$5,0 \times 10^{12}$	$5,0 \times 10^{12}$
Estimador 2	$1,5 \times 10^6$	$1,8 \times 10^4$	1,17	$9,7 \times 10^{11}$	$9,7 \times 10^{11}$
Estimador 3	$8,5 \times 10^5$	$-1,5 \times 10^6$	-64,25	$4,4 \times 10^{12}$	$6,7 \times 10^{12}$
$d = 400$					
Estimador 1	$1,1 \times 10^6$	$3,1 \times 10^5$	41,84	$2,6 \times 10^{11}$	$3,5 \times 10^{11}$
Estimador 2	$4,2 \times 10^5$	$4,2 \times 10^4$	11,31	$3,2 \times 10^{10}$	$3,4 \times 10^{10}$
Estimador 3	$5,3 \times 10^5$	$-1,3 \times 10^4$	-2,46	$4,2 \times 10^{11}$	$4,2 \times 10^{11}$

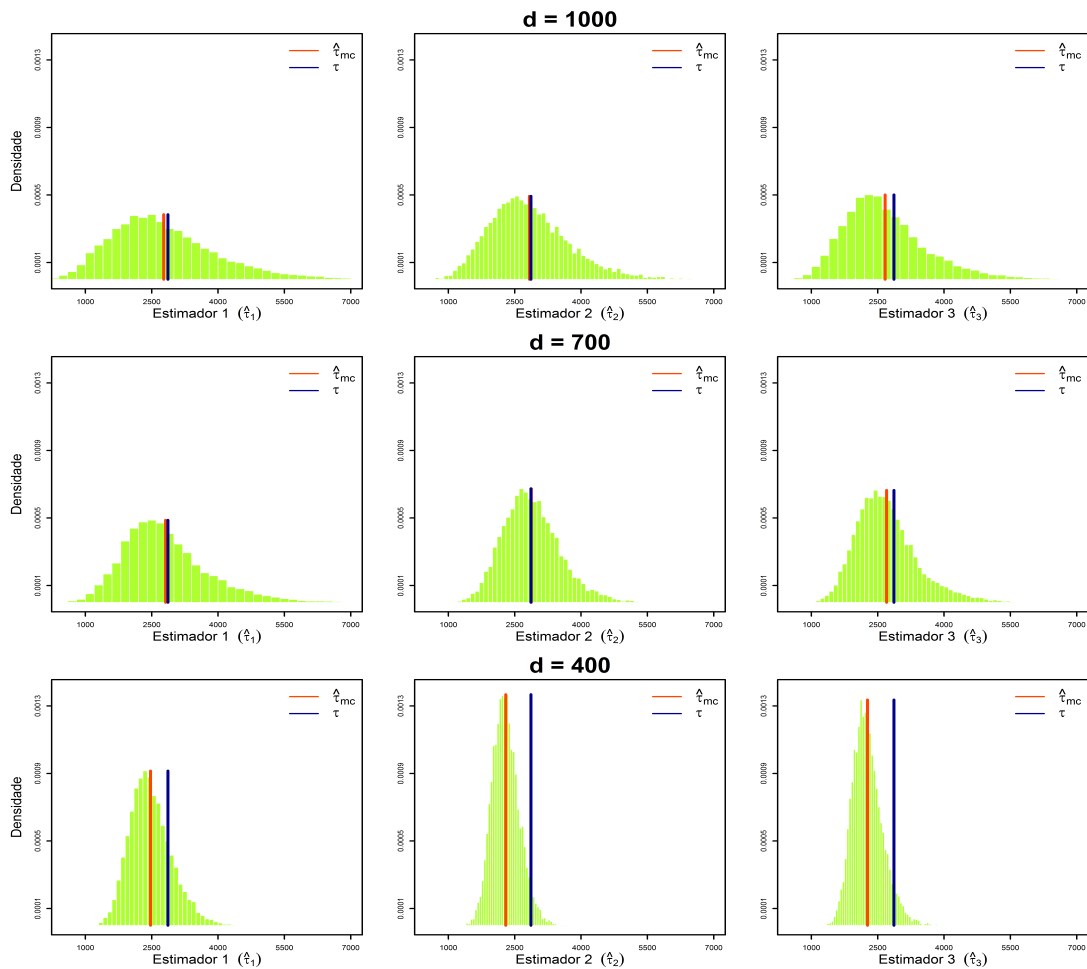
O total populacional de área cultivada com a cultura II foi de 2.866,84h. Vinte e cinco propriedades apresentaram áreas com essa cultura. A tabela 15 mostra que o Estimador 2 teve melhor desempenho, tanto em termos de viés quanto dispersão, com $d = 1000$ e $d = 700$. A diminuição da dimensão dos segmentos, $d = 400$, resultou em uma subestimação da área total com a cultura II na ordem de 13,73% e 19,92% para os estimadores 1 e 2, respectivamente. Para este tamanho de aresta, o Estimador 1 apresentou os menores viés e EQM.

Tabela 15 – Estimativas para o total de área com a cultura II, cenário 1.

	$\hat{t}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	2.770,11	-96,73	-3,37	1.204,93	1.461.223,35
Estimador 2	2.830,51	-36,32	-1,27	903,38	817.424,56
Estimador 3	2.667,05	-199,79	-6,97	941,26	925.879,35
$d = 700$					
Estimador 1	2.812,42	-54,42	-1,90	969,87	943.613,77
Estimador 2	2.874,30	7,46	0,26	641,59	411.700,23
Estimador 3	2.701,21	-165,63	-5,78	702,68	521.193,91
$d = 400$					
Estimador 1	2.473,26	-393,58	-13,73	474,19	379.767,19
Estimador 2	2.295,85	-570,98	-19,92	306,51	419.972,96
Estimador 3	2.269,43	-597,41	-20,84	343,69	475.014,70

A figura 18 mostra as densidades empíricas dos estimadores para o total de área com a cultura II. Assim como é visto na tabela 15 o Estimador 1 apresenta a maior variância de estimativas entre os estimadores. Novamente, com o aumento do número de segmentos selecionados ($d = 400$) e dispersão das estimativas em torno das médias das densidades diminuiu, mas os vieses aumentaram.

Figura 18 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura II, cenário 1.



A tabela 16 mostra os números para as estimativas das variâncias dos estimadores para a cultura II. O Estimador 1 foi o que apresentou maior estimativa de variância, para todos valores de aresta. O Estimador 2 foi o único que teve apenas vieses positivos (superestimação), além de menores estimativas de variância, se comparado ao Estimador 1. As estimativas da variância do Estimador 3 foram as que apresentaram maiores vieses. Além disso, percebe-se também um aumento nos vieses de todos os estimadores para $d = 400$.

Tabela 16 – Estimativas para a variância do total estimado de área com a cultura II, cenário 1.

	$\hat{v}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	$1,4 \times 10^6$	$-2,2 \times 10^4$	-1,48	$2,7 \times 10^{12}$	$2,7 \times 10^{12}$
Estimador 2	$8,2 \times 10^5$	$5,9 \times 10^3$	0,72	$6,4 \times 10^{11}$	$6,4 \times 10^{11}$
Estimador 3	$6,2 \times 10^5$	$-2,6 \times 10^5$	-29,66	$7,0 \times 10^{11}$	$7,7 \times 10^{11}$
$d = 700$					
Estimador 1	$9,4 \times 10^5$	$1,7 \times 10^3$	0,18	$1,5 \times 10^{12}$	$1,5 \times 10^{12}$
Estimador 2	$4,1 \times 10^5$	$2,9 \times 10^3$	0,71	$1,0 \times 10^{11}$	$1,0 \times 10^{11}$
Estimador 3	$3,9 \times 10^5$	$-1,0 \times 10^5$	-20,27	$3,8 \times 10^{11}$	$3,9 \times 10^{11}$
$d = 400$					
Estimador 1	$4,5 \times 10^5$	$2,2 \times 10^5$	98,18	$9,3 \times 10^{10}$	$1,4 \times 10^{11}$
Estimador 2	$1,2 \times 10^5$	$2,6 \times 10^4$	27,80	$4,7 \times 10^9$	$5,3 \times 10^9$
Estimador 3	$2,4 \times 10^5$	$1,2 \times 10^5$	100,30	$1,3 \times 10^{11}$	$1,5 \times 10^{11}$

A tabela 17 mostra os resultados para as estimativas do total de área com a cultura III, cujo total populacional foi 1.403,09ha. Vinte e uma propriedades apresentaram áreas com essa cultura. O Estimador 1 apresentou menor viés em relação ao Estimador 2, porém este apresentou menor desvio padrão, i. e., apresentou menor sensibilidade à mudanças na amostra. O EQM mostra que o Estimador 2 foi melhor do que os outros estimadores, desde que um pouco mais de viés seja aceitável.

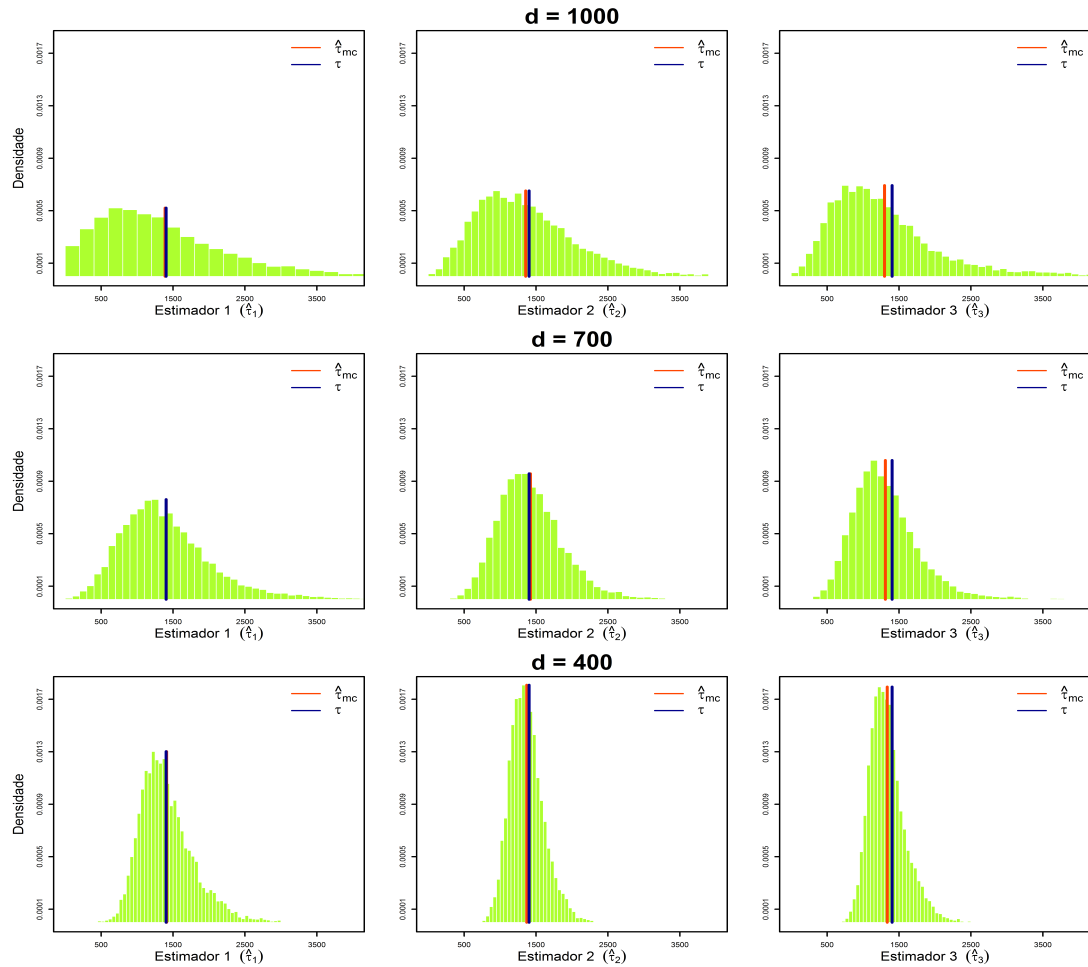
Tabela 17 – Estimativas para o total de área com a cultura III, cenário 1.

	$\hat{\tau}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	1.388,79	-14,30	-1,02	1.004,34	1.008.905,25
Estimador 2	1.356,13	-46,32	-3,35	689,68	477.869,61
Estimador 3	1.298,25	-104,85	-7,47	783,37	624.559,07
$d = 700$					
Estimador 1	1.404,67	1,57	0,11	646,00	417.313,42
Estimador 2	1.421,52	18,42	1,31	453,39	205.898,86
Estimador 3	1.309,91	-93,18	-6,64	464,64	224.580,66
$d = 400$					
Estimador 1	1.409,85	6,75	0,48	374,57	140.349,69
Estimador 2	1.369,41	-33,68	-2,40	234,82	56.277,04
Estimador 3	1.335,42	-67,67	-4,82	250,46	67.308,43

Na figura 19 estão as densidades numéricas dos estimadores para o total de área com a cultura III. A dispersão entre as estimativas reduziu a medida que o número de segmentos selecionados aumentou. Porém, com o aumento do tamanho da amostra, a densidade exibiu um padrão assimétrico à esquerda. O Estimador 1 apresentou a maior dispersão entre os estimadores.

A tabela 18 mostra as estimativas para as variâncias dos estimadores do total de área com a cultura III. Mais uma vez, as estimativas de variância para o Estimador 1 foram as maiores

Figura 19 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura III, cenário 1.



entre os três estimadores. Para o Estimador 2, as estimativas da variância foram menores se comparadas ao Estimador 1, além de apresentar os menores vieses. Além do mais, com o tamanho de aresta $d = 400$ houve um aumento de viés dos estimadores de variância.

Tabela 18 – Estimativas para a variância do total estimado de área com a cultura III, cenário 1.

	$\hat{V}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	$9,8 \times 10^5$	$-2,5 \times 10^4$	-2,48	$4,1 \times 10^{12}$	$4,1 \times 10^{12}$
Estimador 2	$4,8 \times 10^5$	$-3,7 \times 10^2$	-0,08	$3,0 \times 10^{11}$	$3,0 \times 10^{11}$
Estimador 3	$3,8 \times 10^5$	$-2,3 \times 10^5$	-37,33	$6,6 \times 10^{11}$	$7,1 \times 10^{11}$
$d = 700$					
Estimador 1	$4,2 \times 10^5$	$8,1 \times 10^2$	0,19	$4,7 \times 10^{11}$	$4,7 \times 10^{11}$
Estimador 2	$2,1 \times 10^5$	$1,2 \times 10^2$	0,06	$6,0 \times 10^{10}$	$6,0 \times 10^{10}$
Estimador 3	$1,9 \times 10^5$	$-2,6 \times 10^4$	-12,06	$1,5 \times 10^{11}$	$1,5 \times 10^{11}$
$d = 400$					
Estimador 1	$2,3 \times 10^5$	$8,8 \times 10^4$	63,04	$4,3 \times 10^{10}$	$5,1 \times 10^{10}$
Estimador 2	$6,6 \times 10^4$	$1,1 \times 10^4$	19,13	$2,7 \times 10^9$	$2,8 \times 10^9$
Estimador 3	$1,3 \times 10^5$	$6,6 \times 10^4$	104,50	$5,8 \times 10^{10}$	$6,2 \times 10^{10}$

A cultura IV foi a cultura com menor área de plantio na população, apenas 171,35ha, observada em seis propriedades. A tabela 19 mostra que os estimadores 2 e 3 tiveram os menores EQM's, com estimador 2 apresentando menor viés e dispersão para $d = 1000$ e $d = 700$. Novamente, houve uma redução na dispersão das estimativas do total para $d = 400$, mas também houve um aumento nos vieses.

Tabela 19 – Estimativas para o total de área com a cultura IV, cenário 1.

	$\hat{\tau}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	172,59	1,06	0,61	276,27	76.325,25
Estimador 2	171,51	-0,03	-0,02	179,53	32.232,93
Estimador 3	160,89	-10,66	-6,21	179,92	32.485,15
$d = 700$					
Estimador 1	172,69	1,16	0,67	221,93	49.225,35
Estimador 2	172,29	0,76	0,44	140,71	19.799,17
Estimador 3	159,78	-11,76	-6,85	152,78	23.481,14
$d = 400$					
Estimador 1	188,15	16,61	9,68	94,65	9.233,77
Estimador 2	203,32	31,79	18,53	70,95	6.044,96
Estimador 3	194,38	22,84	13,32	70,27	5.459,36

A figura 20 mostra as distribuições de densidade empíricas dos estimadores de variância para a cultura IV. Por essa cultura não ser frequente, em muitas réplicas as estimativas foram zero, para $d = 1000$ e $d = 700$. Para o maior tamanho de amostra ($d = 400$) observou-se uma redução da dispersão das estimativas em torno de sua média e redução das estimativas de valor nulo. Houve também aumento nos vieses dos estimadores para este valor de aresta, mas agora positivos. Novamente, o Estimador 1 apresentou maior dispersão entre todos os estimadores.

A tabela 20 mostra os números para as estimativas das variâncias dos estimadores para a cultura IV. Como pôde também ser visto na figura 20, a variância para o Estimador 1 foi a maior entre os estimadores. Esse comportamento também foi observado para as outras culturas. As estimativas de variância do Estimador 2 foram os menores valores, se comparado apenas ao Estimador 1, além de apresentar os menores vieses.

4.3.2 Cenário 2

O total de área agrícola coberta pelo cadastro de segmentos foi semelhante ao percen-

Figura 20 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura IV, cenário 1.

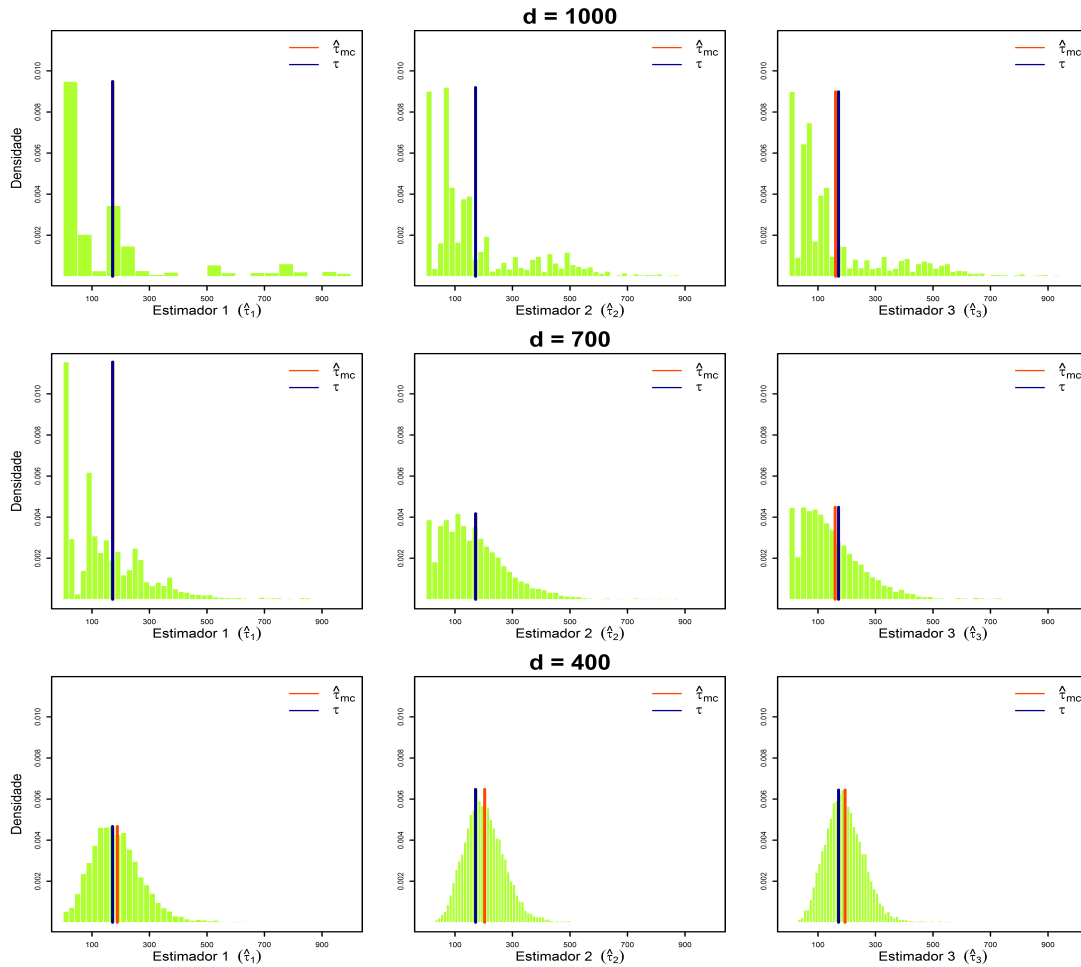


Tabela 20 – Estimativas para a variância do total estimado de área com a cultura IV, cenário 1.

	$\hat{V}_{mc}$	Viés	Viés relativo (%)	DP	EQM
$d = 1000$					
Estimador 1	$7,5 \times 10^4$	$-1,6 \times 10^3$	-2,06	$3,8 \times 10^{10}$	$3,8 \times 10^{10}$
Estimador 2	$3,3 \times 10^4$	$3,7 \times 10^2$	1,14	$3,3 \times 10^9$	$3,3 \times 10^9$
Estimador 3	$3,2 \times 10^4$	$-3,7 \times 10^2$	-1,15	$3,9 \times 10^9$	$3,9 \times 10^9$
$d = 700$					
Estimador 1	$4,9 \times 10^4$	$-5,4 \times 10^2$	-1,10	$5,1 \times 10^{10}$	$5,1 \times 10^{10}$
Estimador 2	$2,0 \times 10^4$	$-1,6 \times 10^2$	-0,83	$3,8 \times 10^9$	$3,8 \times 10^9$
Estimador 3	$2,0 \times 10^4$	$-3,0 \times 10^3$	-12,70	$1,1 \times 10^{10}$	$1,1 \times 10^{10}$
$d = 400$					
Estimador 1	$1,6 \times 10^4$	$7,4 \times 10^3$	82,65	$5,9 \times 10^8$	$6,5 \times 10^8$
Estimador 2	$4,9 \times 10^3$	$-9,6 \times 10^0$	-1,90	$4,5 \times 10^7$	$4,5 \times 10^7$
Estimador 3	$7,0 \times 10^3$	$2,0 \times 10^3$	41,06	$1,0 \times 10^9$	$1,0 \times 10^9$

tual apresentado no cenário 1. Porém, os estratos agrícola e altamente agrícola contiveram um percentual maior de área cultivada: 94,03% para $d = 1000$, 95,87% para $d = 700$ e 97,47% para $d = 400$. Isto indica que, além da homogeneidade da paisagem no solo, a acurácia da estratificação depende do tamanho do segmento. A tabela 21 mostra os resultados da estratificação para os diferentes valores de aresta.

Tabela 21 – Número de segmentos por estrato, cenário 2.

	Altamente agrícola	Agrícola	Não agrícola	Total
$d = 1000$	269	43	131	443
$d = 700$	566	67	284	917
$d = 400$	1789	121	886	2796

Os resultados mostram que o Estimador 2 teve melhores resultados na estimação da área total de plantio com a cultura I, o tipo de cultivo mais comum neste cenário (tabela 22). Para as culturas II e III os estimadores 1 e 2 demonstraram desempenhos semelhantes, com uma vantagem do Estimador 2 para $d = 700$ e vantagem do Estimador 1 para $d = 1000$. Apesar de apresentar o maior viés entre os estimadores, negativo em todos os casos, o Estimador 3 mostrou bons resultados para as culturas menos frequentes.

O aumento do número de segmentos selecionados, decorrente da diminuição do tamanho do segmento ($d = 400$), resultou em redução nos vieses das estimativas dos totais de área obtidas através dos estimadores 2 e 3, para todas as culturas. Para o Estimador 1, isto foi verificado apenas para as culturas I e III. Esse efeito oposto ao observado no cenário 1, para $d = 400$ houve aumento nos vieses observados, pode ter explicação no maior percentual de áreas agrícolas contidas nos estratos agrícola e altamente agrícola, visto que nos dois cenários foram aplicadas a mesma fração amostral no estrato não agrícola.

Nesse cenário, as culturas II e III tiveram áreas de cultivo próximas a área de cultivo com a cultura IV no cenário 1. A redução do número de propriedades resultou em vieses maiores para os estimadores 2 e 3 quando $d = 1000$ e $d = 700$. Adido a isto, a baixa incidência dessas culturas e as diferenças de tamanho das propriedades resultaram, para estes estimadores, em distribuições de densidades numéricas multimodais (figuras 22 e 23). O Estimador 1 demonstrou não ser afetado pela redução do número de propriedades.

As estimativas para a variância dos Estimador 1 foram as maiores entre os estimadores, exceto para a cultura II com $d = 1000$. O Estimador 3 apresentou as menores estimativas de variância. Porém, com exceção da cultura IV, estas estimativas apresentaram vieses negativos muito grandes, indicando que a real variância desse estimador tende a ser subestimada. As estimativas das variâncias do Estimador 2 foram, na grande maioria dos casos, menores se

comparado as variâncias do Estimador 1. Novamente, o efeito da diminuição do tamanho do segmento sobre a variância dos estimadores não pôde ser avaliada isoladamente por se confundir com o efeito do aumento do tamanho da amostra.

O total populacional de área cultivada com a cultura I no cenário 2 foi de 29.164,10ha. Essa cultura foi observada em vinte e duas propriedades. A tabela 22 mostra de o Estimador 2 apresentou melhores resultados em todos os casos, tanto menor viés quanto menor variância. O viés relativo do Estimador 1 foi aproximadamente o percentual de área cultivada na região não coberto pelo cadastro de segmentos, para todos valores de aresta. Adicionalmente, todos os estimadores demonstraram um viés sistemático negativo o que, para os estimadores 2 e 3, pode ser decorrente da assimetria das áreas das propriedades.

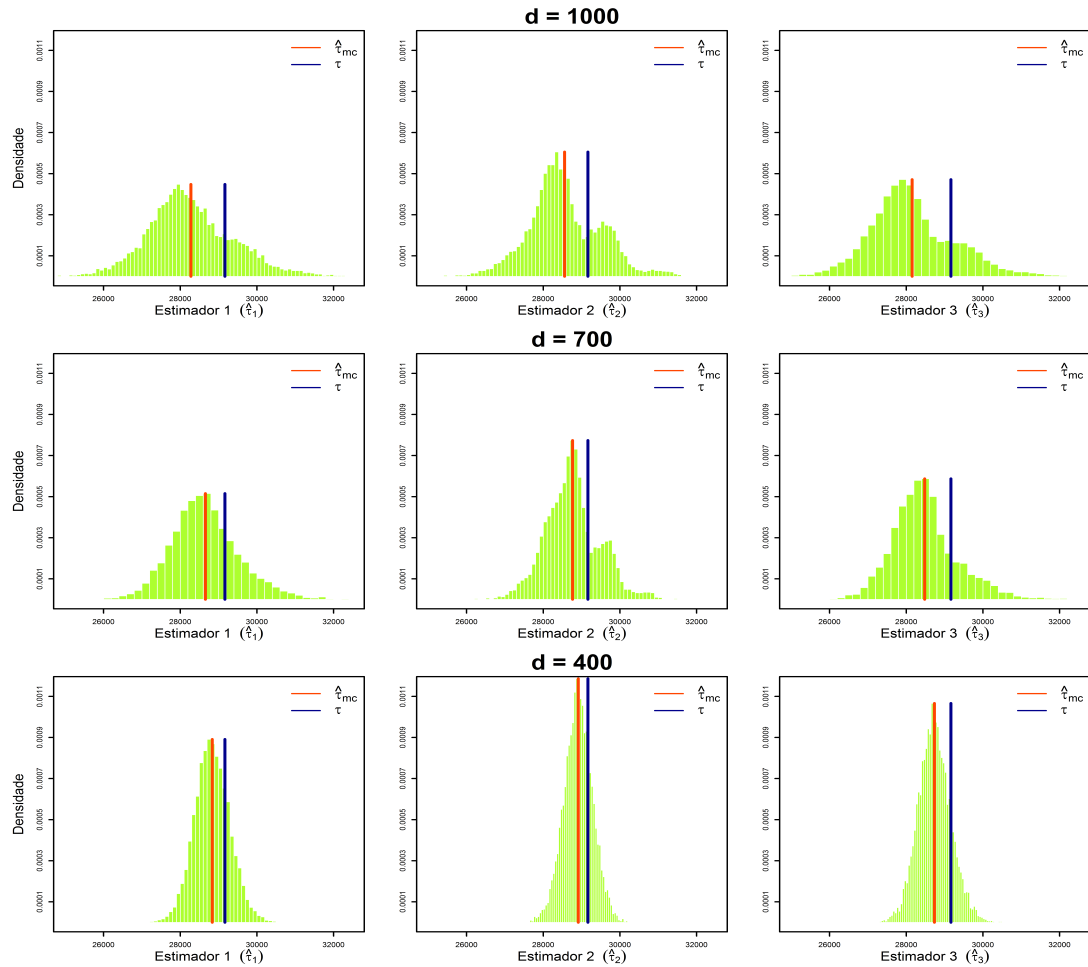
Tabela 22 – Estimativas para o total de área com cultura I, cenário 2.

	$\hat{\tau}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	28.277,93	-886,17	-3,04	1.119,94	2.039.546,81
Estimador 2	28.556,92	-607,18	-2,08	933,60	1.240.281,71
Estimador 3	28.152,22	-1.011,88	-3,47	1104,86	2.244.626,67
$d = 700$					
Estimador 1	28.656,65	-507,44	-1,74	890,56	1050.599,86
Estimador 2	28.765,25	-398,85	-1,37	712,55	666.813,09
Estimador 3	28.480,83	-683,27	-2,34	852,41	1.193.463,95
$d = 400$					
Estimador 1	28.833,75	-330,35	-1,13	462,38	322.923,14
Estimador 2	28.911,69	-252,41	-0,87	375,91	205.017,45
Estimador 3	28.735,06	-429,04	-1,47	429,90	368.888,63

Na figura 21 estão as densidades numéricas dos estimadores para o total de área com a cultura I, no cenário 2. Pode ser observada a influência das propriedades de maiores áreas nas distribuições das estimativas para $d = 1000$ e $d = 700$, principalmente sobre os estimadores 2 e 3; para o Estimador 2 observou-se distribuições bimodais. Com o aumento do tamanho da amostra ($d = 400$) essa influência foi reduzida. O Estimador 1 parece ser o estimador menos afetado por propriedades de grandes áreas, visto que utiliza apenas informações dos segmentos na construção de estimativas.

A tabela 23 mostra os resultados para as estimativas das variâncias dos estimadores para a cultura I. A estimativa da variância do Estimador 1 foi a maior entre os estimadores, para todos os valores de aresta. Ainda, os estimadores 1 e 3 apresentaram apenas vieses negativos,

Figura 21 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura I, cenário 2.



que reduziram com o aumento do tamanho da amostra, e o Estimador 2 apresentou apenas vieses positivos. Embora apresente as menores estimativas para a variância, o Estimador 3 apresenta vieses negativos muito grandes.

Tabela 23 – Estimativas para a variância do total estimado de área com a cultura I, cenário 2.

	$\hat{V}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	$1,2 \times 10^6$	$-4,7 \times 10^4$	-3,77	$1,7 \times 10^{12}$	$1,7 \times 10^{12}$
Estimador 2	$8,8 \times 10^5$	$8,5 \times 10^3$	0,98	$7,9 \times 10^{11}$	$7,9 \times 10^{11}$
Estimador 3	$4,9 \times 10^5$	$-7,3 \times 10^5$	-59,50	$8,8 \times 10^{11}$	$1,4 \times 10^{12}$
$d = 700$					
Estimador 1	$7,8 \times 10^5$	$-1,7 \times 10^4$	-2,12	$6,8 \times 10^{11}$	$6,8 \times 10^{11}$
Estimador 2	$5,3 \times 10^5$	$2,1 \times 10^4$	4,12	$2,2 \times 10^{11}$	$2,2 \times 10^{11}$
Estimador 3	$3,5 \times 10^5$	$-3,7 \times 10^5$	-51,57	$3,5 \times 10^{11}$	$4,9 \times 10^{11}$
$d = 400$					
Estimador 1	$2,1 \times 10^5$	$-1,7 \times 10^3$	-0,80	$1,4 \times 10^{10}$	$1,4 \times 10^{10}$
Estimador 2	$1,4 \times 10^5$	$2,5 \times 10^3$	1,76	$5,9 \times 10^9$	$6,0 \times 10^9$
Estimador 3	$1,5 \times 10^5$	$-3,0 \times 10^4$	-16,22	$3,6 \times 10^{10}$	$3,7 \times 10^{10}$

A tabela 24 mostra os resultados para as estimativas do total de área com cultura II, cujo total populacional foi 453,57ha. Essa cultura foi observada em oito propriedades. O Estimador 2 foi o único estimador que apresentou viés positivo. Para cada valor de aresta, um dos estimadores teve melhor precisão considerando o EQM: o Estimador 1 foi melhor com $d = 1000$, o Estimador 2 com $d = 700$ e o Estimador 3 com $d = 400$.

Tabela 24 – Estimativas para o total de área com cultura II, cenário 2.

	$\hat{\tau}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	453,34	-0,22	-0,05	455,62	207.587,20
Estimador 2	466,95	13,38	2,95	477,01	227.713,84
Estimador 3	400,51	-53,06	-11,70	382,18	148.879,98
$d = 700$					
Estimador 1	452,08	-1,49	-0,33	452,69	204.932,35
Estimador 2	458,67	5,10	1,12	374,80	140.501,63
Estimador 3	434,31	-19,26	-4,25	392,44	154.378,55
$d = 400$					
Estimador 1	451,58	-1,99	-0,44	258,26	66.700,29
Estimador 2	454,69	1,13	0,25	208,37	43.420,70
Estimador 3	432,96	-20,60	-4,54	199,90	40.384,88

A figura 22 mostra as distribuições de densidade empíricas dos estimadores de totais para a cultura II. Os estimadores apresentaram um grande volume de estimativas próximas de zero para $d = 1000$ e $d = 700$. Com o aumento do tamanho da amostra ($d = 400$), as estimativas se afastaram um pouco do zero. Contudo, por se tratar de uma cultura pouco frequente e por esse cenário ter um pequeno número de propriedades, os estimadores 2 e 3 apresentaram múltiplos picos em sua densidade empírica.

A tabela 25 mostra os resultados para as estimativas das variâncias dos estimadores para a cultura II. As estimativas da variância do Estimador 3 foram as menores entre os estimadores para $d = 1000$ e $d = 700$, mas com altos vieses negativos, i. e., houve subestimação. A estimativa para a variância do Estimador 2 foi a menor entre os estimadores para $d = 400$. As estimativas da variância do Estimador 1 foram as maiores entre os estimadores para $d = 700$ e $d = 400$.

A área com cultura III no cenário 2 foi de 127,41ha e foi observada em cinco propriedades. Os estimadores 1 e 2 tiveram resultados semelhantes em termos de EQM, o Estimador 1 com maior sensibilidade a mudanças na amostra e o Estimador 2 com maior viés.

Figura 22 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura II, cenário 2.

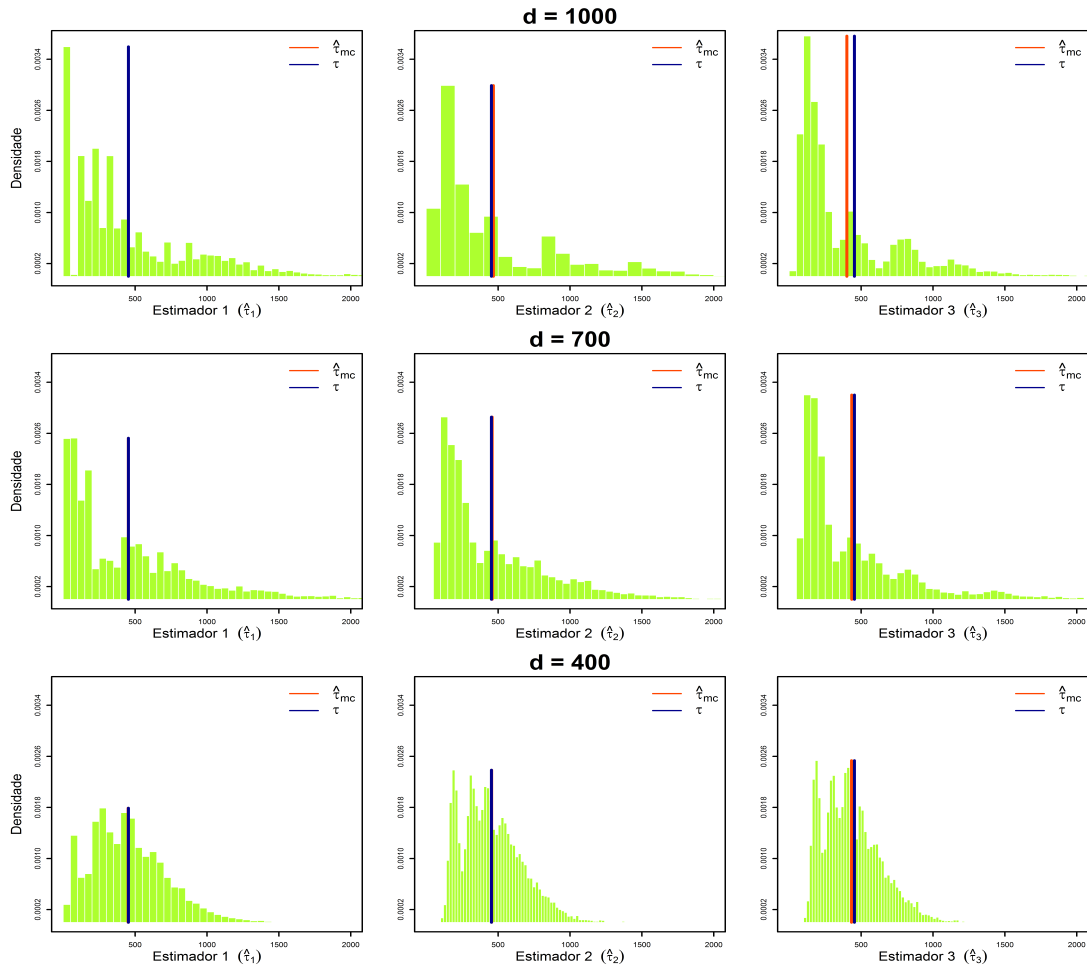


Tabela 25 – Estimativas para a variância do total estimado de área com a cultura II, cenário 2.

	$\hat{V}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	$2,1 \times 10^5$	$-1,7 \times 10^3$	-0,82	$1,3 \times 10^{11}$	$1,3 \times 10^{11}$
Estimador 2	$2,3 \times 10^5$	$-1,7 \times 10^3$	-0,73	$2,1 \times 10^{11}$	$2,1 \times 10^{11}$
Estimador 3	$9,0 \times 10^4$	$-5,6 \times 10^4$	-38,60	$3,7 \times 10^{10}$	$4,0 \times 10^{10}$
$d = 700$					
Estimador 1	$2,1 \times 10^5$	$2,0 \times 10^3$	0,99	$1,4 \times 10^{11}$	$1,4 \times 10^{11}$
Estimador 2	$1,4 \times 10^5$	$2,2 \times 10^3$	1,59	$5,8 \times 10^{10}$	$5,8 \times 10^{10}$
Estimador 3	$9,8 \times 10^4$	$-5,6 \times 10^4$	-36,11	$7,5 \times 10^{10}$	$7,8 \times 10^{10}$
$d = 400$					
Estimador 1	$6,6 \times 10^4$	$-2,7 \times 10^2$	-0,40	$3,1 \times 10^9$	$3,1 \times 10^9$
Estimador 2	$4,4 \times 10^4$	$9,6 \times 10^2$	2,22	$1,5 \times 10^9$	$1,5 \times 10^9$
Estimador 3	$5,2 \times 10^4$	$1,2 \times 10^4$	29,07	$9,7 \times 10^9$	$9,8 \times 10^9$

Com a diminuição do comprimento da aresta os estimadores 1 e 2 passaram a apresentar vieses positivos. O Estimador 3 apresentou os maiores vieses (negativos) entre os estimadores, mas com os menores desvios-padrão, o que resultou em menores EQM's, para $d = 700$ e $d = 400$.

A figura 23 mostra as densidades empíricas dos estimadores de totais para a cultura

Tabela 26 – Estimativas para o total de área com cultura III, cenário 2.

	$\hat{\tau}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	125,96	-1,45	-1,14	182,52	33.314,69
Estimador 2	99,07	-28,34	-22,24	175,97	31.769,95
Estimador 3	93,93	-33,48	-26,28	179,23	33.243,21
$d = 700$					
Estimador 1	128,26	0,85	0,67	161,16	25.971,87
Estimador 2	134,10	6,69	5,25	149,27	22.327,38
Estimador 3	117,76	-9,65	-7,57	134,16	18.092,20
$d = 400$					
Estimador 1	127,82	0,41	0,32	108,17	11.701,01
Estimador 2	132,96	5,55	4,36	100,47	10.125,89
Estimador 3	121,36	-6,05	-4,74	93,54	8.785,41

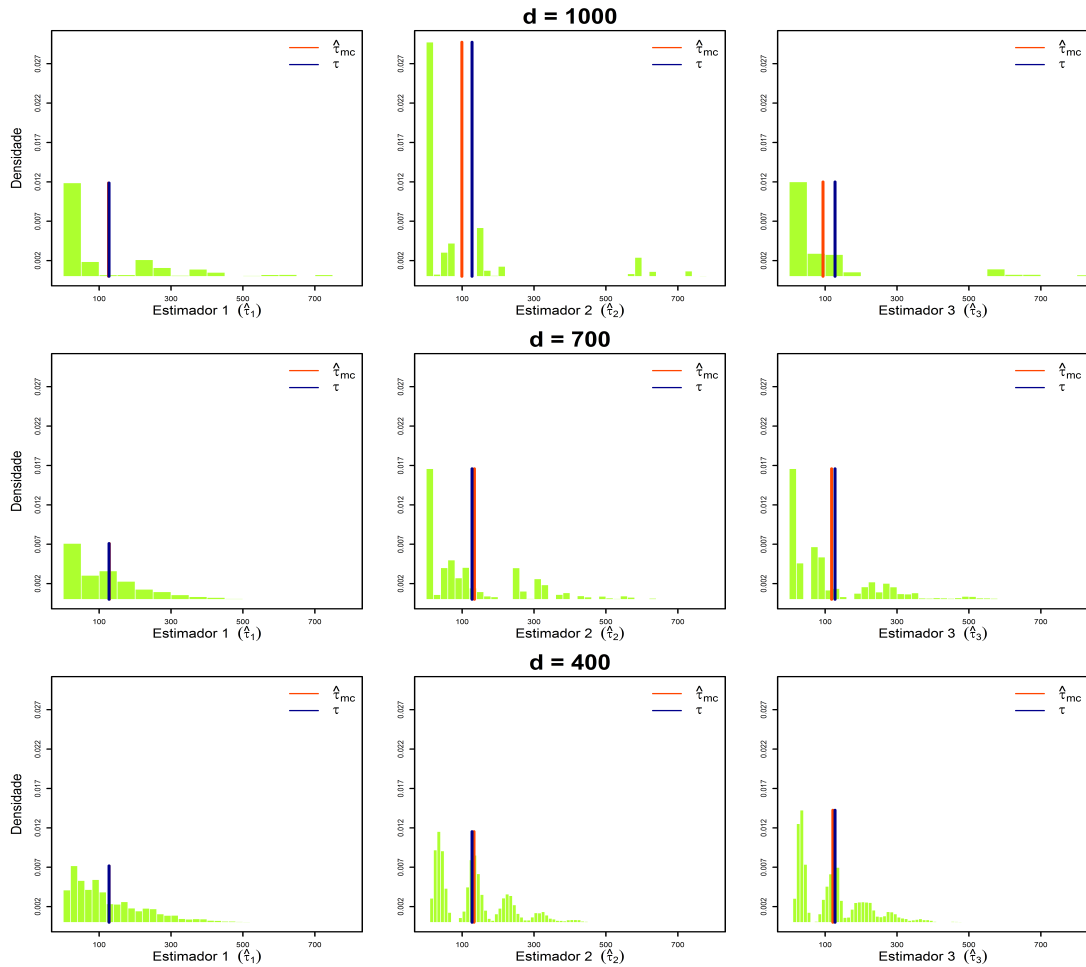
III. As densidades mostram um comportamento semelhante ao observado para a cultura II. Há uma concentração de estimativas próximas de zero para $d = 1000$ e $d = 700$, principalmente para o Estimador 2. Com $d = 400$, as densidades dos estimadores 2 e 3 apresentaram um padrão multimodal decorrente do pequeno número de proprietários e da baixa incidência dessa cultura. Estes fatores não influenciaram o desempenho do Estimador 1 que apresentou apenas assimetria positiva, pois muitos segmentos não tinham áreas com a cultura III.

A tabela 27 mostra os resultados para as estimativas das variâncias dos estimadores de totais com a cultura III. As estimativas para a variância do Estimador 1 foram maiores entre os estimadores. A variância estimada para o Estimador 2 foi ligeiramente menor em relação ao Estimador 1, para $d = 700$ e $d = 400$. As estimativas da variância do Estimador 3 foram as menores entre os estimadores para $d = 1000$ e $d = 700$, mas também apresentaram vieses negativos, i. e., houve subestimação da real variância.

Tabela 27 – Estimativas para a variância do total estimado de área com a cultura III, cenário 2.

	$\hat{v}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	$3,3 \times 10^4$	$-5,5 \times 10^2$	-1,64	$6,4 \times 10^9$	$6,4 \times 10^9$
Estimador 2	$3,1 \times 10^4$	$-2,1 \times 10^2$	-0,69	$7,9 \times 10^9$	$7,9 \times 10^9$
Estimador 3	$2,4 \times 10^4$	$-7,7 \times 10^3$	-24,06	$6,6 \times 10^9$	$6,6 \times 10^9$
$d = 700$					
Estimador 1	$2,6 \times 10^4$	$4,7 \times 10^2$	1,82	$1,6 \times 10^{10}$	$1,6 \times 10^{10}$
Estimador 2	$2,3 \times 10^4$	$5,8 \times 10^2$	2,60	$1,5 \times 10^9$	$1,5 \times 10^9$
Estimador 3	$1,6 \times 10^4$	$-1,8 \times 10^3$	-9,95	$1,0 \times 10^9$	$1,0 \times 10^9$
$d = 400$					
Estimador 1	$1,2 \times 10^4$	$-3,1 \times 10^1$	-0,26	$4,0 \times 10^8$	$4,0 \times 10^8$
Estimador 2	$1,0 \times 10^4$	$-5,6 \times 10^1$	-0,55	$1,3 \times 10^8$	$1,3 \times 10^8$
Estimador 3	$1,1 \times 10^4$	$2,0 \times 10^3$	23,06	$7,4 \times 10^8$	$7,5 \times 10^8$

Figura 23 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura III, cenário 2.

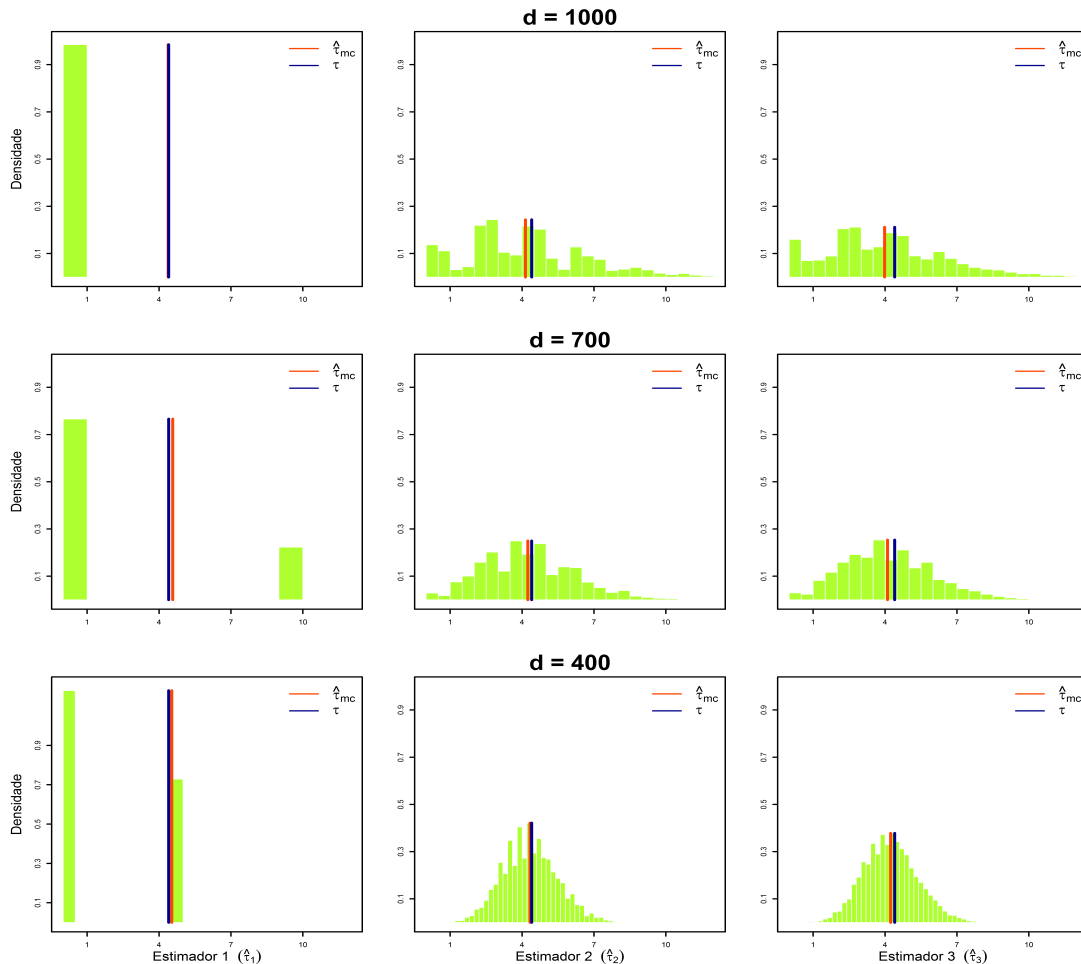


A cultura com menor área no cenário foi a cultura IV, com área de 4,40ha em uma única propriedade. Isto pode ser o motivo do Estimador 1 apresentar uma alta variabilidade nas estimativas entre as réplicas, como visto na tabela 28. Os estimadores em dois estágios, por outro lado, apresentaram bons desempenhos, pois a propriedade que apresentou essa cultura cobria uma ampla área e, então, um número razoável de segmentos foram associados indiretamente a esta cultura.

Na figura 24 estão as densidades numéricas dos estimadores para o total de área com a cultura IV, no cenário 2. O efeito da pequena área com essa cultura sobre a dispersão do Estimador 1 fica bem patente, mesmo com o aumento do tamanho amostral. Para os estimadores 2 e 3, a dispersão das estimativas em torno de sua média é menor devido ao número de segmentos através dos quais são obtidas informações acerca da cultura IV. Isso fica mais evidente com o aumento do tamanho da amostra ($d = 400$).

Tabela 28 – Estimativas para o total de área com cultura IV, cenário 2.

	$\hat{\tau}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	4,38	-0,02	-0,42	34,38	1.181,85
Estimador 2	4,14	-0,26	-5,94	2,56	6,62
Estimador 3	3,98	-0,41	-9,37	2,47	6,28
$d = 700$					
Estimador 1	4,57	0,17	3,87	22,64	512,40
Estimador 2	4,24	-0,16	-3,63	1,90	3,65
Estimador 3	4,10	-0,29	-6,71	1,85	3,52
$d = 400$					
Estimador 1	4,53	0,13	3,06	14,64	214,33
Estimador 2	4,33	-0,07	-1,59	1,15	1,32
Estimador 3	4,23	-0,17	-3,78	1,13	1,30

Figura 24 – Densidades empíricas dos estimadores, obtidas através das estimativas para o total de área com a cultura IV, cenário 2.

As estimativas para a variância do total estimado de área com cultura IV são exibidas na tabela 29. Assim como as estimativas do total, as estimativas da variância para o Estimador 1 não mostraram bons resultados, apresentando maiores estimativas de variâncias e desvios-padrão. As estimativas das variâncias para os estimadores 2 e 3 apresentaram resultados razoáveis, com estimativas ligeiramente menores para as variâncias do Estimador 3.

Tabela 29 – Estimativas para a variância do total estimado de área com a cultura IV, cenário 2.

	$\hat{V}_{mc}$	Viés	Viés relativo	DP	EQM
$d = 1000$					
Estimador 1	$1,2 \times 10^3$	-0,12	-0,01	$9,0 \times 10^7$	$9,0 \times 10^7$
Estimador 2	$6,7 \times 10^0$	0,16	2,47	$1,8 \times 10^1$	$1,8 \times 10^1$
Estimador 3	$5,9 \times 10^0$	-0,20	-3,28	$1,5 \times 10^1$	$1,5 \times 10^1$
$d = 700$					
Estimador 1	$5,1 \times 10^2$	0,54	0,11	$2,2 \times 10^7$	$2,2 \times 10^7$
Estimador 2	$3,6 \times 10^0$	-0,01	-0,40	$2,8 \times 10^0$	$2,8 \times 10^0$
Estimador 3	$3,3 \times 10^0$	-0,10	-2,97	$2,5 \times 10^0$	$2,5 \times 10^0$
$d = 400$					
Estimador 1	$2,2 \times 10^2$	2,00	0,93	$2,3 \times 10^6$	$2,3 \times 10^6$
Estimador 2	$1,3 \times 10^0$	-0,01	-0,49	$1,1 \times 10^{-1}$	$1,1 \times 10^{-1}$
Estimador 3	$1,3 \times 10^0$	0,03	2,53	$2,0 \times 10^{-1}$	$2,0 \times 10^{-1}$

4.4 DISCUSSÃO DOS RESULTADOS

O estudo de simulação mostrou que o Estimador 1 tem um viés próximo a diferença entre a área total de cultivo no cadastro de segmentos e a área de cultivo na região, se a cultura de interesse for predominante. Uma correção de viés pode ser aplicada para melhorar a precisão do estimador. Pode ainda ser implementado uma abordagem que considere apenas a interseção dos segmentos de borda com a região de interesse, evitando que segmentos sejam retirados do cadastro. Na prática, esse viés negativo deve ser menor devido à observação de campos dentro dos segmentos que estão localizados fora da região, a exemplo do experimento em Goiana. Isto pode funcionar como uma autocorreção de viés.

Apesar de algumas vezes o Estimador 2 exibir maiores vieses, este tem menor sensibilidade à mudanças na amostra por apresentar, no geral, os menores desvios-padrão em relação ao Estimador 1; isto resultou também em valores de EQM geralmente menores para o Estimador 2. Adicionado o fato que é facilmente implementado, se comparado ao Estimador 1, o Estimador 2 mostrou os resultados mais promissores, sobretudo para estimação de totais de área de culturas predominantes. A quantidade de propriedades demonstrou ser um importante fator para os desempenhos dos estimadores 2 e 3 na estimação de totais de área para culturas menos predominantes. Com a redução do número de propriedades no segundo cenário, esses estimadores demonstraram um aumento de viés e densidades numéricas multimodais.

Os resultados mostraram que a estimação por segmento ponderado consegue, de fato, reduzir a variância dos estimadores. O Estimador 1, construído sobre a estimação por segmento fechado, apresentou maiores estimativas de variância em relação aos estimadores 2 e 3, construídos sobre a estimação por segmento ponderado. As estimativas de variância para o Estimador 3 foram as menores entre os estimadores. Contudo, o Estimador 3 exibiu os maiores vieses que, no geral, foram negativos.

A estratificação por pontos consegue ser mais efetiva considerando pelo menos dois fatores: o tamanho do segmento e a configuração da paisagem agrícola no solo. A simulação mostrou que o percentual de área agrícola nos estratos agrícolas (agrícola e altamente agrícola) aumenta ao passo que o tamanho do segmento é reduzido. Em contra partida, a redução dos segmentos aumenta o número de pontos que devem ser foto-interpretados. Como esperado, quando a paisagem é composta por grandes campos de produção, a estratificação proposta consegue incorporar informações de áreas agrícolas em estratos agrícolas. Quando a paisagem apresenta muitos campos de cultivo, i. e., um paisagem heterogênea, o percentual de áreas agrícolas classificadas no estrato não agrícola é maior.

Os resultados também sugerem que a utilização de segmentos com pequenas dimensões ($d = 400$ nesse estudo), em cenários que exibam grande quantidade de campos produtivos (cenário 1), podem levar os estimadores a produzir estimativas mais enviesadas. Este fato pode estar relacionado com o percentual de área agrícola contida no estrato não agrícola, maior em paisagens heterogêneas. Para reduzir esses vieses, pode ser realizada a seleção de uma amostra maior nesse estrato.

5 CONSIDERAÇÕES FINAIS

Esse trabalho foi iniciado com o intuito de descrever e difundir técnicas de amostragem de área e, em particular, sua aplicabilidade em estudos agropecuários. Para atingir esses objetivos, foram apresentados os principais tipos de unidades amostrais de um cadastro de área, além de meios para a associação entre as unidades amostrais e observacionais. Ferramentas de baixo custo que podem ser utilizadas para construir e estratificar um cadastro de área também foram expostas. Neste sentido, o uso de segmentos com limites geométricos foi apontado como alternativa a segmentos com limites irregulares, por apresentar baixo custo de delineamento. Técnicas de estimação baseadas em cadastros de segmentos e pontos foram elucidadas, enfatizando as vantagens e desvantagens de cada método. A aplicação de três estimadores de área foi exemplificada através de um experimento prático, realizado no município de Goiana-PE, e seus desempenhos comparados através de simulação de Monte Carlo.

Entre os principais resultados deste trabalho, figuram comparações sistêmicas entre estimadores de área baseadas em fatores de aplicabilidade, eficiência na amostragem e suscetibilidade a erros não amostrais. No estudo de Goiana, as estimativas de área total com cana-de-açúcar foram semelhantes para os estimadores 1 e 2. Contudo, as estimativas para a quantidade de fertilizantes evidenciaram uma desvantagem do Estimador 1 que, mesmo aplicado através de uma adaptação, provavelmente produziu subestimativas. Além disso, a estimação por segmento fechado é de difícil implementação na prática e está sujeita a erros de mensuração, tanto devido a erros do pesquisador quanto ao uso de instrumentos auxiliares na geração das variáveis respostas. O Estimador 3 demonstrou ter a menor aplicabilidade, por requerer a mensuração de áreas nos segmentos e por ser afetado pela subjetividade das informações dos proprietários.

O comparativo numérico mostrou que, apesar de apresentar bons resultados para o viés, o Estimador 1 tem maior variabilidade. Isto somado a complexidade dos planos amostrais que utilizam esse estimador, faz com que a sua aplicação seja menos aconselhável. Nesse contexto, o Estimador 2 apresentou melhor desempenho, se comparado ao Estimador 1, por duas razões: (i) embora possa apresentar um viés um pouco maior, esse estimador é mais preciso por apresentar menor variância e (ii) o estimador exige planos amostrais menos complexos. A utilização do Estimador 3 não é recomendada devido a esse estimador ser mais impreciso, se

comparado ao Estimador 2, e por exigir planos amostrais que, dependendo da densidade de produtores, exige um grande esforço na fase do trabalho de campo. Os estimadores para as variâncias dos estimadores demonstraram que os estimadores 2 e 3, baseados na estimação por segmento ponderado, possuem menores variâncias, se comparado ao Estimador 1. Contudo, o estimador da variância do Estimador 3 apresentou tendência a subestimar a real variância.

O estudo de simulação aqui apresentado, e principal contribuição dessa dissertação, foi o ponto de partida para uma série de análises comparativas entre os diferentes estimadores de área. Em trabalhos futuros, podem ser realizados estudos mais detalhados sobre os seguintes tópicos:

- Incluir na avaliação numérica outros estimadores apresentados. Podem ser inclusos, por exemplo, o estimador por segmento aberto, o estimador por segmento ponderado (sem subamostragem de pontos) e estimadores baseados unicamente em pontos;
- Analisar como erros na estratificação afetam a precisão dos estimadores baseados em segmentos;
- Analisar como o tamanho dos segmentos e o número de pontos subamostrados afetam o estimador por segmento ponderado, com subamostragem de pontos;
- Avaliar separadamente o efeito do aumento do tamanho da amostra e das dimensões dos segmentos sobre a precisão dos estimadores do total e da variância do total;
- Avaliar a propriedade de consistência dos estimadores para a variância dos estimadores de totais.

REFERÊNCIAS

- BIVAND, R.; KEITT, T.; ROWLINGSON, B. **rgdal: Bindings for the 'Geospatial' Data Abstraction Library**. [S.l.], 2018. R package version 1.2-18. Disponível em: <<https://CRAN.R-project.org/package=rgdal>>.
- BIVAND, R.; RUNDEL, C. **rgeos: Interface to Geometry Engine - Open Source ('GEOS')**. [S.l.], 2017. R package version 0.3-26. Disponível em: <<https://CRAN.R-project.org/package=rgeos>>.
- BÍBLIA. **A Bíblia Sagrada**. Revista e corrigida. 1995. São Paulo: Sociedade Bíblica do Brasil, 1995. Traduzida em português por João Ferreira de Almeida.
- CARFAGNA, E. Using satellite imagery and geo-referencing technology for building a master sampling frame. In: **Proceedings of the 59th ISI World Statistics Congress, Hong Kong**. [S.l.: s.n.], 2013.
- CARFAGNA, E.; GALLEGÓ, F. J. Using remote sensing for agricultural statistics. **International Statistical Review / Revue Internationale de Statistique**, [Wiley, International Statistical Institute (ISI)], v. 73, n. 3, p. 389–404, 2005. Disponível em: <<http://www.jstor.org/stable/25472682>>.
- COCHRAN, W. G. **Sampling Techniques**. [S.l.]: 1. ed. Mumbai: John Wiley & Sons, 1977.
- COTTER, J.; NEALON, J. **Area frame design for agricultural surveys**. [S.l.]: US Department of Agriculture, National Agricultural Statistics Service Washington, DC, 1987.
- DAVIES, C. **Area frame design for agricultural surveys**. Washington, D.C., 2009.
- DELINCÉ, J. **Technical Report on Cost-Effectiveness of Remote Sensing for Agricultural Statistics in Developing and Emerging Economies**. [S.l.], 2015.
- DELINCÉ, J. **Master Sampling Frames: the field experiments conducted in Nepal**. [S.l.], 2017.
- EUROPEAN, C. **Building Agro Environmental Indicators: Focussing on the European area frame survey LUCAS**. [S.l.]: European Commission, Joint Research Centre, Institute for Environment and Sustainability, Land Management Unit, 2002.
- FAO. **Multiple frame agricultural surveys Surveys: Volume I: Current Surveys based on area and list sampling methods**. [S.l.]: FAO Statistical Development Series, No. 7. FAO Publication: Rome, 1996.
- FERRAZ, C. **Linking Area Frames and List Frames in Agricultural Surveys**. [S.l.], 2015.
- FERRAZ, C. **Global Strategy to improve Agricultural and Rural Statistics. Brazil's Master Sampling Frame Experiments**. [S.l.], 2018.
- FORD, B. L.; NEALON, J.; TORTORA, R. D. *et al.* Area frame estimators in agricultural surveys: sampling versus nonsampling errors. **Agricultural Economics Research**, United States Department of Agriculture, Economic Research Service, Washington, D.C., v. 38, n. 2, p. 1–9, 1986.
- GALLEGÓ, F. J.; DELINCÉ, J.; CARFAGNA, E. Two stage area frame sampling on square segments for farm surveys. **Survey Methodology**, v. 20, n. 2, p. 107–115, 1994.

GALLEGO, F. J. *et al.* **Sampling frames of square segments**. [S.l.]: Office for Official Publ. of the European Communities, 1995.

GOEBEL, J. J.; SCHREUDER, H. T.; HOUSE, C. C.; GEISSLER, P. H.; OLSEN, A. R.; WILLIAMS, W. A study on integrating surveys of terrestrial natural resources: The oregon demonstration project. **EPA/600/R-98/032**. Corvallis, OR: US Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory. 44 p., 1998.

GSARS – Global Strategy to improve Agricultural and Rural Statistics. **Master Sampling Frames for Agricultural Statistics Frame Development, Sample Design and Estimation**. [S.l.]: Global Strategy Handbook: Rome, 2015.

GSARS – Global Strategy to improve Agricultural and Rural Statistics. **Handbook on Remote Sensing for Agricultural Statistics**. [S.l.]: GSARS Handbook: Rome, 2017.

HENDRICKS, W. A.; SEARLS, D. T.; HORVITZ, D. G. A comparison of three rules for associating farms and farmland with sample area segments in agricultural surveys. 1965.

HOUSEMAN, E. E. **Area frame sampling in agriculture**. [S.l.]: Statistical Reporting Service, US Department of Agriculture Washington, DC, 1975.

LOHR, S. **Sampling: design and analysis**. Second edition. [S.l.]: 2. ed. Boston: Cengage Learning, 2009.

MARTINEZ, L. I. In: FOOD AND AGRICULTURE ORGANIZATION. **Improving the Use of GPS, GIS and RS for Setting Up a Master Sampling Frame**. [S.l.], 2013.

NEALON, J. P. **Review of the Multiple and Area Frame Estimators**. Washington, D.C., 1984.

PEBESMA, E. J.; BIVAND, R. S. Classes and methods for spatial data in R. **R News**, v. 5, n. 2, p. 9–13, November 2005. Disponível em: <<https://CRAN.R-project.org/doc/Rnews/>>.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>.

TAYLOR, J.; SANNIER, C.; DELINCÉ, J.; GALLEGGO, F.; JRC, I. Regional crop inventories in europe assisted by remote sensing. **Synthesis Report, Office for Publications of the EC**, 1997.

TILLÉ, Y.; MATEI, A. **sampling: Survey Sampling**. [S.l.], 2016. R package version 2.8. Disponível em: <<https://CRAN.R-project.org/package=sampling>>.

APÊNDICE A – CÓDIGOS COMPUTACIONAIS

Neste apêndice, são apresentados alguns códigos computacionais que foram implementados, utilizando a sintaxe do R, para obtenção dos resultados expostos nessa dissertação.

IMPLEMENTAÇÃO DO ESTIMADORES

Estimador 1

A seguinte função calcula os valores das estimativas para o Estimador 1 e sua respectiva variância, dados por

$$\hat{\tau}_1 = \sum_{h=1}^H \sum_{i \in S_h} \frac{y_i}{\pi_i} = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in S_h} y_i$$

e

$$\widehat{\text{Var}}(\hat{\tau}_1) = \sum_{h=1}^H \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_h^2.$$

```
##### Estimador 1 #####
#
# ----- Estimador para o total de área plantada e variância do ----- #
#           estimador na amostragem por segmento fechado                      #
#                                                                                   #
#####

# ===== #
#
#           DETALHES DA FUNÇÃO           #
#
# A função recebe 3 objetos de entrada:   #
#
# 1 - Matriz com informações sobre a variável resposta:   #
#           . Estrato                                     #
#           . Área cultivada pela propriedade no segmento i   #
#
# 2 - Vetor de tamanho da população por estratos         #
# 3 - Vetor de tamanho da amostra por estratos           #
```

```

#                                                                 #
#          ----- (NESSA ORDEM) -----                          #
#                                                                 #
#   Obs.: Essa função supõe que os dados são referentes a uma única   #
#           cultura de interesse                                       #
#                                                                 #
# ===== #

Estimador1 <- function(data, N, n){

  data <- data %>% as.data.frame
  colnames(data) <- c("estrato", "resposta")
  EST <- data.frame(estrato = c('Agrícola', 'Altamente Agrícola', 'Não Agrícola'))

  aux <- data %>% group_by(estrato) %>%
    summarise(total = sum(resposta),
              variancia = var(resposta),
              n.eff = length(resposta))

  aux <- left_join(EST, aux, by = 'estrato')
  for(i in 1:ncol(aux)) aux[aux[,i] %>% is.na,i] <- 0

  TOTAL <- aux$total*(N/n)
  TOTAL <- c(sum(TOTAL), TOTAL)

  k <- (N^2*n*(aux$n.eff - 1))/((n - 1)*aux$n.eff^2) * (1 - n/N)
  DP <- k*aux$variancia
  DP[is.na(DP)] <- 0
  DP <- c(sum(DP), DP) %>% sqrt

  CV <- DP/TOTAL
  N.EFF <- c(sum(aux$n.eff), aux$n.eff)

  RES <- rbind(TOTAL, DP, CV, N.EFF)
  row.names(RES) <- c("Total", "DP", "CV", "n efetivo")
  colnames(RES) <- c("Total", aux$estrato %>% levels)

  return(RES)
}

```

}

#####

Estimador 2

A seguinte função calcula os valores das estimativas para o Estimador 2 e sua respectiva variância, dados por

$$\hat{\tau}_2 = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{T_i}{P_i} \sum_{j=1}^{P_i} \frac{y_j}{A_j}.$$

e

$$\widehat{\text{Var}}(\hat{\tau}_2) = \sum_{h=1}^H \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_{Xh}^2 + \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{1}{P_i(P_i - 1)} \sum_{j=1}^{P_i} \left(\frac{y_j}{A_j} T_i - \hat{X}_i\right)^2$$

Estimador 2

#

----- Estimador para o total de área plantada e variância do -----

estimador na amostragem em dois estágios (por pontos)

#

#####

=====

#

DETALHES DA FUNÇÃO

#

A função recebe 5 objetos de entrada:

#

1 - Matriz com informações sobre a variável resposta:

. Estrato

. Segmento

. Proporção da propriedade com a cultura

2 - Vetor de tamanho da população por estratos

3 - Vetor de tamanho da amostra por estratos

4 - Área de um segmento

5 - Número de pontos dentro dos segmentos

#

----- (NESSA ORDEM) -----

#

```

# Obs.: Essa função supõe que os dados são referentes a uma única cultura de interesse
#
#
# =====
#

Estimador2 <- function(data, N, n, A.seg = 49, N.ponts = 5){

  data <- data %>% as.data.frame
  colnames(data) <- c("estrato", "segmento", "resposta")
  EST <- data.frame(estrato = c('Agrícola', 'Altamente Agrícola', 'Não Agrícola'))

  aux <- data %>% group_by(segmento) %>%
    summarise(X = sum(resposta*A.seg/N.ponts),
              V2 = sum((resposta*A.seg - X)^2))

  aux2 <- left_join(aux, data[!duplicated(data[, -3]), -3], "segmento")

  aux2 <- aux2 %>% group_by(estrato) %>%
    summarise(total = sum(X),
              V1 = var(X),
              V2 = sum(V2))

  aux2 <- left_join(EST, aux2, by = 'estrato')
  for(i in 1:ncol(aux2)) aux2[aux2[, i] %>% is.na, i] <- 0

  TOTAL <- aux2$total*N/n
  TOTAL <- c(sum(TOTAL), TOTAL)

  k1 <- (N^2/n)*(1 - n/N)
  k2 <- (N/n)/(N.ponts*(N.ponts - 1))
  DP <- k1*aux2$V1 + k2*aux2$V2
  DP[is.na(DP)] <- 0
  DP <- c(sum(DP), DP) %>% sqrt

  CV <- DP/TOTAL

  RES <- rbind(TOTAL, DP, CV)
  row.names(RES) <- c("Total", "DP", "CV")
  colnames(RES) <- c("Total", aux2$estrato %>% levels)

```

```

VAR <- rbind(c(sum(k1*aux2$V1), k1*aux2$V1),
             c(sum(k2*aux2$V2), k2*aux2$V2))
row.names(VAR) <- c("1° stage", "2° stage")
colnames(VAR) <- c("Total", aux2$estrato %>% levels)

return(list("Resultado" = RES, "Variancia" = VAR))

}

#####

```

Estimador 3

A seguinte função calcula os valores das estimativas para o Estimador 3 e sua respectiva variância, dados por

$$\hat{\tau}_3 = \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{G_i}{F_i} \sum_{j=1}^{F_i} \frac{y_j}{A_j}.$$

e

$$\widehat{\text{Var}}(\hat{\tau}_3) = \sum_{h=1}^H \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) S_{Xh}^2 + \sum_{h=1}^H \frac{N_h}{n_h} \sum_{i \in s_h} \frac{1}{F_i(F_i - 1)} \sum_{j=1}^{F_i} \left(\frac{y_j}{A_j} G_i - \hat{X}_i\right)^2$$

```

##### Estimador 3 #####
#
# ----- Estimador para o total de área plantada e variância do ----- #
#          estimador ponderado por pontos agrícolas no segmento          #
#
#####

# ===== #
#
#          DETALHES DA FUNÇÃO          #
#
# A função recebe 3 objetos de entrada: #
#
# 1 - Matriz com informações sobre a variável resposta: #
#          . Estrato #
#          . Segmento #

```

```

#           . Nome da propriedade                               #
#           . Proporção da propriedade com a cultura           #
#           . Área agrícola no segmento i                       #
#   2 - Vetor de tamanho da população por estratos           #
#   3 - Vetor de tamanho da amostra por estratos              #
#                                                                 #
#           ----- (NESSA ORDEM) -----                      #
#                                                                 #
#   Obs.: Essa função supõe que os dados são referentes a uma única cultura de interesse
#                                                                 #
#                                                                 #
# ===== #

Estimador3 <- function(data, N, n){

  data <- data %>% as.data.frame
  colnames(data) <- c("estrato", "segmento", "prop", "resposta", "AAU")

  data <- data[data$prop != 0,]
  EST <- data.frame(estrato = c('Agrícola', 'Altamente Agrícola', 'Não Agrícola'))

  aux <- data %>% group_by(segmento) %>%
    summarise(P = sum(prop != 0),
              X = sum(resposta*AAU/P),
              V2 = sum((resposta*AAU - X)^2),
              PP = P*(P - 1))

  aux$X[is.nan(aux$X)] <- 0
  aux$V2[is.nan(aux$V2)] <- 0
  aux$V2[aux$P == 0] <- 0
  aux$PP[aux$PP == 0] <- 1

  aux2 <- left_join(aux, data[!duplicated(data[, -c(3:5)]), -c(3:5)],
                    by = "segmento")

  aux2 <- aux2 %>% group_by(estrato) %>%
    summarise(total = sum(X),
              V1 = var(X),
              V2 = sum(V2/PP))

```

```

aux2 <- left_join(EST, aux2, by = 'estrato')
for(i in 1:ncol(aux2)) aux2[aux2[,i] %>% is.na,i] <- 0

TOTAL <- aux2$total*N/n
TOTAL <- c(sum(TOTAL), TOTAL)

DP <- N^2/n*(1 - n/N)*aux2$V1 + N/n*aux2$V2
DP[is.na(DP)] <- 0
DP <- c(sum(DP), DP) %>% sqrt

CV <- DP/TOTAL

RES <- rbind(TOTAL, DP, CV)
row.names(RES) <- c("Total", "DP", "CV")
colnames(RES) <- c("Total", aux2$estrato %>% levels)

VAR <- rbind(c(sum(N^2/n*(1-n/N)*aux2$V1), N^2/n*(1-n/N)*aux2$V1),
             c(sum(N/n*aux2$V2), N/n*aux2$V2))
row.names(VAR) <- c("1° stage", "2° stage")
colnames(VAR) <- c("Total", aux2$estrato %>% levels)

return(list("Resultado" = RES, "Variancia" = VAR))

}

#####

```

FUNÇÕES PARA A SIMULAÇÃO

Função Suplementar

Esta função faz a redução de escala em um conjunto de polígonos.

1. Recebe como entrada:

- uma lista de polígonos espaciais;
- um fator de redução de escala.

2. Tem como saída:

- uma lista de coordenadas de polígonos com a redução de escala;

- área dos polígonos espaciais.

```
#####

Reducao <- function(SPoly, reducao = 100){

  area <- numeric(0)
  fazendas <- list()

  for(i in 1:length(SPoly)){

    p <- SPoly@polygons[[i]]@Polygons[[1]]@coords
    pcentro <- matrix(coordinates(SPoly[i,]), ncol = 2,
                        nrow = nrow(p), byrow = TRUE)

    fazendas[[i]] <- (p - pcentro)/reducao
    area[i] <- SPoly@polygons[[i]]@area

  }

  return(list("Poligonos" = fazendas, "Areas" = area))

}

#####
```

Posicionamento dos Polígonos

Esta função faz o posicionamento dos polígonos em uma região, sem que haja interseção os polígonos dois a dois.

1. Recebe como entrada:

- um polígono representando a região de interesse;
- uma lista de polígonos espaciais;
- um fator de redução de escala.

2. Tem como saída:

- uma lista de polígonos espaciais posicionados sobre a região de interesse.

```
#####
```

```

Posionamento <- function(regiao, SPoly, reducao = 100){

  aux <- Reducao(SPoly, reducao)
  area <- aux$Areas
  poligonos <- aux$Poligonos

  a <- regiao@bbox[1,2] - regiao@bbox[1,1]
  b <- regiao@bbox[2,2] - regiao@bbox[2,1]
  sigma <- matrix(c(a,0,0,b), 2)/(reducao*4)
  mu <- spsample(regiao, n = 3, type = "random")@coords
  r <- quantile(area, probs = c(.45, .85)) %>% as.vector

  SPpoligonos = intersecao <- list()
  i <- 1
  ok = jump = rep <- 0

  while(i <= length(poligonos)){ # length(poligonos)

    if(area[i] <= r[1]){

      centro <- mvrnorm(mu = mu[1,], Sigma = sigma)
      deslocamento <- SpatialPoints(matrix(centro, byrow=TRUE,ncol=2,
                                              nrow=nrow(poligonos[[i]])),
                                      proj4string = CRS(proj4string(SPoly)))

      while(!gIntersects(regiao, deslocamento)){
        centro <- mvrnorm(mu = mu[1,], Sigma = sigma)
        deslocamento <- SpatialPoints(matrix(centro, byrow=TRUE,ncol=2,
                                              nrow=nrow(poligonos[[i]])),
                                          proj4string = CRS(proj4string(SPoly)))
      }
    }

    if(r[1] < area[i] && area[i] <= r[2]){

      centro <- mvrnorm(mu = mu[2,], Sigma = sigma)
      deslocamento <- SpatialPoints(matrix(centro, byrow=TRUE,ncol=2,
                                              nrow=nrow(poligonos[[i]])),
                                      proj4string = CRS(proj4string(SPoly)))
    }
  }
}

```

```

while(!gIntersects(regiao, deslocamento)){
  centro <- mvrnorm(mu = mu[2,], Sigma = sigma)
  deslocamento <- SpatialPoints(matrix(centro, byrow=TRUE,ncol=2,
                                         nrow=nrow(poligonos[[i]])),
                                proj4string = CRS(proj4string(SPoly)))
}
}

if(area[i] >= r[2]){

  centro <- mvrnorm(mu = mu[3,], Sigma = sigma)
  deslocamento <- SpatialPoints(matrix(centro, byrow=TRUE,ncol=2,
                                         nrow=nrow(poligonos[[i]])),
                                proj4string = CRS(proj4string(SPoly)))

  while(!gIntersects(regiao, deslocamento)){
    centro <- mvrnorm(mu = mu[3,], Sigma = sigma)
    deslocamento <- SpatialPoints(matrix(centro, byrow=TRUE,ncol=2,
                                           nrow=nrow(poligonos[[i]])),
                                   proj4string = CRS(proj4string(SPoly)))
  }
}

p <- Polygon(poligonos[[i+jamp]] + matrix(centro, byrow=TRUE,ncol=2,
                                           nrow=nrow(poligonos[[i+jamp]])))

SPpoligonos[[i]] <- Polygons(list(p), i)
intersecao[[i]] <- SpatialPolygons(list(Polygons(list(p), i)))

for(j in 1:(i-1)){

  if(i == 1) break

  if(gIntersects(intersecao[[i]], intersecao[[j]])){
    rep <- rep + 1
    i = i - 1
    break
  }
}

```

```

    }

    i <- i + 1
    ok <- ok + 1

    if(rep == 250){
        jump <- jump + 1
        rep <- 0
    }

    if(i + jump > length(poligonos)){
        cat("Posicionamento máximo de polígnos\n")
        break
    }

    if(ok == 10000){
        cat("Posicionamento de todos os polígnos falhou
            => Número de tentativas excedido\n")
        break
    }
}

rm(intersecao)
SPpoligonos <- SpatialPolygons(SPpoligonos,
                                proj4string = CRS(proj4string(SPoly)))

SP <- list()
for(i in 1:length(SPpoligonos)){
    aux <- gIntersection(SPpoligonos[i,], regioao)
    aux <- aux@polygons[[1]]@Polygons[[1]]
    SP[[i]] <- Polygons(list(aux), i)
}

rm(SPpoligonos)
SP <- SpatialPolygons(SP, proj4string = CRS(proj4string(SPoly)))

return(SP)

}

#####

```

Geração um Cadastro de Segmentos

Esta função gera um cadastro de segmentos sobre uma região.

1. Recebe como entrada:

- um polígono representando a região de interesse;
- o comprimento do retângulo no eixo x;
- o comprimento do retângulo no eixo y;
- projeção UTM da região de interesse.

2. Tem como saída:

- um cadastro de retângulos cobrindo quase toda a região de interesse.

```
#####
```

```
Grid.Segmentos <- function(Regiao, dist.x = 1000, dist.y = 1000,
                           projecao = "+proj=utm +zone=24 +south"){

  aux <- spTransform(Regiao, CRS(projecao))

  MIN <- aux@bbox[,1]
  MAX <- aux@bbox[,2]
  N.seg <- c(ceiling((MAX[1] - MIN[1])/ dist.x),
            ceiling((MAX[2] - MIN[2])/ dist.y))

  G.seg <- GridTopology(cellcentre.offset = as.vector(MIN),
                        cellsize = c(dist.x, dist.y),
                        cells.dim = as.vector(N.seg))

  G.seg <- as.SpatialPolygons.GridTopology(grd = G.seg,
                                           CRS(projecao))

  ADD <- logical(length(G.seg))
  ID <- 1

  for(i in 1: length(G.seg)){

    if(!gIntersects(G.seg[i,], aux)) next
```

```

aux2 <- gIntersection(G.seg[i,], aux)
seg.area <- G.seg@polygons[[1]]@area

if((aux2@polygons[[1]]@area/seg.area) >= .5){
  ADD[i] <- TRUE
  G.seg@polygons[[i]]@ID = ID %>% as.character
  ID <- ID + 1
}
}

G.seg <- G.seg[ADD,]
G.seg <- spTransform(G.seg, CRS(proj4string(Regiao)))
return(G.seg)

}

#####

```

Geração um Cadastro de Segmentos e Pontos

Esta função gera um cadastro de segmentos, com pontos dentro dos segmentos, sobre uma região.

1. Recebe como entrada:

- um cadastro de segmentos relativo à região de interesse;
- uma configuração de pontos centrada na origem;
- projeção UTM da região de interesse.

2. Tem como saída:

- um cadastro de pontos agrupados dentro de segmentos que cobre quase toda a região de interesse.

```

#####

Grid.SegPont <- function(Grid.Seg, shape.grid,
                          projecao = "+proj=utm +zone=24 +south"){

  shape.grid <- SpatialPoints(shape.grid, CRS(projecao))
  aux <- spTransform(Grid.Seg, CRS(projecao))

```

```

P <- list()
for(i in 1:length(Grid.Seg)){
  centro <- matrix((coordinates(aux[i,])), ncol = 2,
                    nrow = length(shape.grid), byrow = TRUE)
  P[[i]] <- coordinates(shape.grid@coords + centro) %>%
    SpatialPoints(CRS(projecao)) %>%
    spTransform(CRS(proj4string(Grid.Seg))) %>%
    coordinates
}

P <- SpatialMultiPointsDataFrame(P,data.frame(segmento=1:length(P)),
                                proj4string = CRS(proj4string(Grid.Seg)))

return(P)

}

#####

```

Estratificação um Cadastro de Segmentos

Esta função faz a estratificação de um cadastro de segmentos através da sobreposição de um cadastro de pontos agrupados e uma lista de polígonos espaciais à região de interesse.

1. Recebe como entrada:

- um cadastro de pontos agrupados relativo à região de interesse;
- uma lista de polígonos espaciais posicionados sobre a região de interesse;
- projeção UTM da região de interesse.

2. Tem como saída:

- uma matriz na qual a primeira coluna é o segmento e a segunda coluna é o estrato.

```
#####
```

```

Estratificacao <- function(MultPontGrid, SPoly, strat = c(2,4),
                           lab = c("Não Agrícola", "Agrícola",
                                   "Altamente Agrícola")){

  estrato <- NULL
  for(i in 1:length(MultPontGrid)){

```

```

AGRI <- 0
p <- MultiPontGrid[i,]@coords
pontos <- SpatialPoints(p, CRS(proj4string(SPoly)))

for(j in 1:length(pontos)){
  if(gIntersects(pontos[j,], SPoly)){
    AGRI <- AGRI + 1
  }
}

if(AGRI < strat[1]) estrato[i] <- lab[1]
if(AGRI >= strat[1] && AGRI < strat[2]) estrato[i] <- lab[2]
if(AGRI >= strat[2]) estrato[i] <- lab[3]
}

return(data.frame(MultiPontGrid@data, estrato))

}

#####

```