

Documentos

ISSN 1516-4691
Dezembro, 2015

100

Método para estratificação em levantamentos agrícolas com mais de uma variável

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Meio Ambiente
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 100

Método para estratificação em levantamentos agrícolas com mais de uma variável

Alfredo José Barreto Luiz
Bruno Schultz
Kleber Trabaquini
Isaque Daniel Rocha Eberhardt
Antonio Roberto Formaggio

Embrapa Meio Ambiente
Jaguariúna, SP
2015

Exemplares desta publicação podem ser adquiridos na:

Embrapa Meio Ambiente

Rodovia SP-340, Km 127,5, Tanquinho Velho
Caixa Postal 69, CEP: 13820-000, Jaguariúna, SP
Fone: + 55 (19) 3311-2700
Fax: + 55 (19) 3311-2640
<https://www.embrapa.br/meio-ambiente/>
SAC: <https://www.embrapa.br/fale-conosco/sac/>

Comitê de Publicações da Unidade

Presidente: *Ladislau Araújo Skorupa*
Secretária-Executiva: *Vera Lúcia S. S. de Castro*
Secretária: *Cristina Tiemi Shoyama*
Bibliotecário: *Victor Paulo Marques Simão*
Membros: *Marília Ieda da Silveira Folegatti Matsuura, Elisabeth Francisconi Fay, Nilce Chaves Gattaz, Joel Leandro de Queiroga, Daniel Terao (suplente), Lauro Charlet Pereira (suplente) e Maria Lúcia Zuccari (suplente).*
Normalização bibliográfica: *Victor Paulo Marques Simão*
Edição eletrônica: *Gabriel Pupo Nogueira*

1ª edição eletrônica (2015)

Todos os direitos reservados

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

**Dados Internacionais de Catalogação na Publicação (CIP)
Embrapa Meio Ambiente**

Método para estratificação em levantamentos agrícolas com mais de uma variável / Alfredo José Barreto Luiz... [et al.]. – Jaguariúna: Embrapa Meio Ambiente, 2015.
39 p. -- (Embrapa Meio Ambiente. Documentos ; 100).

1. Estatística agrícola 2. Amostragem. 3. Sensoriamento remoto.
I. Luiz, Alfredo José Barreto. II. Título. III. Série.

CDD 519.5

Sumário

Introdução.....	07
O caso de São Paulo.....	13
O caso do Paraná	27
O caso do Rio grande do Sul.....	33
Conclusões.....	37
Referências	38

Método para estratificação em levantamentos agrícolas com mais de uma variável

Alfredo José Barreto Luiz¹, Bruno Schultz², Kleber Trabaquini³, Isaque Daniel Rocha Eberhardt⁴, Antonio Roberto Formaggio²

Introdução

A estratificação é uma técnica comum em amostragem e existem muitas razões para isso. Entre essas razões pode-se destacar que quando se deseja conhecer o grau de precisão dos dados para certas subdivisões da população é conveniente tratar cada uma delas como uma “população” em separado, ou seja, cada subpopulação equivale a um estrato.

Outra forte motivação é a existência de conveniências administrativas; por exemplo, a agência que conduz a coleta de dados pode ter vários escritórios, cada um com sua área de abrangência pré-definida segundo critérios muitas vezes alheios aos objetivos de um levantamento específico.

Outro motivo pode ser percebido no caso de uma amostragem por produtores agrícolas de determinada cultura, quando só se possui uma lista das grandes propriedades, as quais são colocadas em um estrato separado,

¹Embrapa Meio Ambiente

²INPE

³EPAGRI

⁴UnB

enquanto as propriedades menores, não listadas, são divididas em outros estratos, nos quais algum tipo de amostragem por área pode ser aplicado (COCHRAN, 1977; FAO, 1996).

A estratificação pode ainda produzir um ganho na precisão das estimativas das características da população completa, pois é possível dividir uma população heterogênea em subpopulações, cada uma das quais sendo internamente homogênea. Isso é sugerido pelo nome “estrato”, com sua implicação de uma divisão em camadas. Se cada um dos estratos é homogêneo, de modo que as medidas variam pouco entre as unidades que os compõem, uma estimativa precisa da média de cada estrato pode ser obtida de uma pequena amostra naquele estrato. As estimativas dos estratos podem, então, ser combinadas em uma única estimativa precisa de toda a população (COCHRAN, 1977).

A estratificação pode ser do tipo espacial, sendo especialmente adequada quando o uso do solo varia muito de uma região para outra dentro dos limites estaduais. Vários são os motivos que podem levar a essa situação. O uso agrícola é fortemente afetado pelo clima, tipo de solo e pela topografia; por exemplo, o arroz irrigado por inundação, nas condições brasileiras de manejo, ocorre quase que exclusivamente em planícies hidromórficas; já as grandes culturas com intensa mecanização, ocupam prioritariamente terrenos com pequena declividade; e as culturas perenes e semiperenes podem ocorrer em áreas de relevo mais acidentado, sendo a escolha da sua localização mais condicionada pelo clima.

Outro fator a afetar o uso do solo são os fatores antropomórficos, entre eles estão principalmente a estrutura fundiária, o preço das terras e a distribuição de infraestrutura de transporte, transformação e armazenamento dos produtos agrícolas. Neste aspecto, é relativamente fácil imaginar que, dentro de um mesmo estado, uma região de difícil acesso – por ausência de estradas ou pontes – será preterida na ocupação em relação a uma outra região melhor servida nesses aspectos; ou então, que as culturas adequadas às regiões divididas em minifúndios serão diferentes daquelas

presentes nas áreas onde os latifúndios são mais frequentes. É importante observar que, embora a estrutura fundiária e as condições da infraestrutura sejam geralmente correlacionadas com as variáveis ambientais de solo, relevo e clima, isso nem sempre é válido (LUIZ, 2009).

Mais um item determinante do tipo de uso do solo em cada região do estado é o componente legal ou jurídico. O estabelecimento de leis de preservação ambiental pode transformar toda uma área em reserva, abolindo qualquer forma de uso agrícola; por outro lado leis que determinem limites para alguma atividade agrícola podem forçar um rearranjo espacial da produção, como no caso das leis relativas à declividade máxima passível de mecanização, ou a que determina a distância mínima em relação a centros urbanos a partir da qual se pode realizar a queima da palhada de cana-de-açúcar antes da colheita (GIANNOTTI, 2001).

Do ponto de vista da estimativa da área ocupada pelas atividades agrícolas, alguns ou todos os fatores citados acima podem agir de forma a promover, ou não, a concentração espacial de uma determinada atividade dentro de um estado. Os fatores e sua importância variam de uma atividade para outra, podendo haver, em um mesmo território, um exemplo cuja distribuição seja homogênea e outro que apresente alta concentração espacial.

A opção pela estratificação, portanto, dependerá mais uma vez do objetivo do estudo. Se for possível, deve ser escolhido o método que privilegia a estimativa da variável de maior interesse. Em outras palavras, dependendo da atividade de interesse, em um mesmo estado pode-se optar por fazer, ou não, a estratificação; e mais, mesmo que se decida pela estratificação, a alocação ótima do tamanho da amostra pelos estratos também variará em função da atividade e da variável de interesse. Por exemplo, no caso de São Paulo, um estado ocupado principalmente pelas culturas da cana-de-açúcar, da soja e do milho, a distribuição espacial da cana será fortemente influenciada pela distância às Usinas, enquanto as culturas anuais dependerão principalmente de outras características mais difusas para definir sua localização. O número de pontos em cada estrato – supondo que

já foi definido o número total – variará de acordo com as características dos estratos estabelecidos.

Antes de prosseguir, é preciso explicar o que se quer dizer com variável e diferenciar variável de atividade. A atividade pode ser, por exemplo, o cultivo anual de soja, enquanto a variável pode ser, para essa atividade, a área ocupada em hectares, ou a produção em toneladas, ou a quantidade de agrotóxicos utilizada, ou ainda o número de empregos gerados por 1.000 hectares. Assim, o planejamento amostral é influenciado pelo tipo de atividade que será investigada e, da mesma forma, pela variável que será mensurada. A estratificação pode ser combinada com a amostragem aleatória baseada em informações obtidas a partir de imagens de sensoriamento remoto, por exemplo, visando dividir um território municipal em áreas internamente homogêneas e diferentes entre si quanto ao uso predominante do solo.

Um conjunto de municípios pertencentes a um Estado pode ser estratificado de forma a se obterem subpopulações homogêneas quanto à variável de interesse (LUIZ, 2009). Isso é relativamente simples quando se trata de uma única variável e de uma determinada atividade, como, por exemplo, a área plantada com determinada cultura. Entretanto, quando, em um levantamento amostral, o interesse é simultâneo por mais de um item, a solução não é trivial, tanto na definição dos limites dos estratos e do tamanho da amostra, como na alocação dos pontos amostrais entre os estratos, sem falar da estimativa desses valores em face de uma semi-amplitude e de um grau de significância desejado para o valor da média. Uma abordagem possível é detalhada a seguir.

O método da amostragem aleatória simples (AAS) consiste na seleção de uma amostra de n elementos de um total populacional de N , de tal forma que qualquer possível amostra tenha a mesma probabilidade de ser escolhida. Daí a importância fundamental do uso do conceito de pixel, extraído das imagens de sensoriamento remoto, para possibilitar a amostragem aleatória.

Como não existe uma lista permanentemente atualizada e precisa de todas as áreas utilizadas na produção agrícola, não é possível realizar

uma amostragem estatística da população de “áreas de produção” com base em títulos de propriedade, nomes de proprietários ou outra lista qualquer. Entretanto, se dividirmos toda a superfície da região de estudo em quadrículas ou elementos de cena (os pixels), eles são adequados para a realização da amostragem aleatória. Então, para extrair uma AAS de n pixels de uma imagem digital de uma região, enumera-se de 1 até N os elementos de cena que compõem a imagem e, através de um algoritmo de geração de números aleatórios, por exemplo, selecionam-se os n elementos da amostra.

Ao assumir que o maior interesse é estimar a área ocupada pelas culturas agrícolas e que a unidade amostral a ser utilizada será o pixel, que tem área conhecida e constante (por exemplo, os pixels de imagens dos sensores a bordo dos satélites da família Landsat tem pixels de 30 por 30 metros), a proporção de ocorrência das classes de interesse na amostra equivale à proporção de área amostral ocupada pelas culturas, o que, por expansão, permite estimar as áreas totais em uma região, como um estado, por exemplo.

O método amostral é especialmente adequado às aplicações nas estimativas agrícolas por meio de sensoriamento remoto em razão da dificuldade de se obter imagens totalmente isentas de nuvens sobre grandes extensões contínuas, nas regiões tropicais e subtropicais, e nas épocas apropriadas aos principais usos agrícolas. Se fosse necessária a obtenção de imagens livres de nuvens, em um curto período de tempo, compatível com a permanência da cultura em pleno desenvolvimento vegetativo, para a confecção de um mapa completo de ocupação do solo de um estado, dificilmente a estimativa resultante poderia ser incorporada em um método rotineiro de previsão de safras, em razão do risco de não haver disponibilidade de informação no prazo necessário.

O problema da alocação da amostra entre estratos quando o levantamento visa estimar mais de uma variável.

Este tema é tratado detalhadamente nas páginas 119 a 121, tópico 5A.3, do livro de Cochran (1977), de onde se fez a adaptação a seguir.

Se um levantamento tem como objetivo estimar o valor de mais de uma variável ao mesmo tempo, utilizando o mesmo planejamento amostral estratificado, o problema da alocação da amostra pelos estratos deixa de ser simples como no caso do estudo de uma única variável ou item. Em geral, se calcular a alocação ótima para um item, o resultado obtido quase nunca será ótimo, no sentido de mínimo coeficiente de variação (CV) para a estimativa, e poderá ser muito ruim para os outros itens do levantamento. Por isso, alguns ajustes devem ser feitos em um levantamento com numerosos itens. Então, o primeiro passo é reduzir o máximo possível os itens considerados na alocação, com objetivo de trabalhar com um conjunto relativamente pouco numeroso, restrito àqueles considerados mais importantes. Ou seja, um levantamento que pretenda estimar o valor de muitas variáveis ao mesmo tempo, dificilmente conseguirá obter baixos valores de coeficiente de variação das estimativas de todas elas. A sugestão é que se trabalhe com um pequeno número de variáveis de cada vez, ou, se isso for impossível, que se escolha apenas algumas para controlar o valor do CV.

Se dados prévios de boa qualidade estiverem disponíveis, pode-se, então, calcular a alocação ótima para cada item em separado e avaliar o quanto isso gera de discordância. Em um levantamento de um tipo especializado, em que as correlações entre os itens sejam altas, as distribuições podem diferir relativamente pouco. Entretanto, em alguns casos os itens podem ser não correlacionados ou até possuírem correlação negativa. É o caso, por exemplo, do levantamento das áreas plantadas com várias culturas, simultaneamente, em determinada região (ex. estado) dividida em sub-regiões (ex. mesorregiões). Nesse caso, como o espaço em cada sub-região, ou estrato, é finito, as culturas tendem a competir por área, podendo resultar em correlação negativa. Ou, no caso de ocuparem relativamente uma pequena fração do espaço disponível em cada estrato, pode não haver correlação.

No presente caso, em que o objetivo é propor um plano amostral para estimar a área ocupada com as culturas de cana-de-açúcar, milho e soja nos estados de São Paulo (SP), Paraná (PR) e Rio Grande do Sul (RS), as mesorregiões foram escolhidas como estratos. Isso se deu por vários motivos, entre eles: as mesorregiões já existirem como subdivisões utilizadas pelo órgão oficial responsável pelas estatísticas agrícolas brasileiras, o IBGE; as mesorregiões serem porções de território contínuas geograficamente, o que permite, com certa facilidade, a delimitação de conjuntos de quadrículas (pixels) contíguas, todas pertencentes a uma única mesorregião e somente a ela; as mesorregiões possuem extensão de área relativamente semelhantes; e, finalmente, dificuldade de estabelecer estratos, principalmente com as características das mesorregiões, por critérios estatísticos que considerem simultaneamente três ou mais *variáveis* (“variáveis” aqui entendidas como culturas agrícolas). A figura 1 mostra uma representação esquemática dos estados de São Paulo, Paraná e Rio Grande do Sul, divididos em mesorregiões, com uma das possíveis realizações de uma distribuição aleatória de pontos amostrais no estado de São Paulo, para serem classificados em imagens de satélite.

A seguir são apresentados os cálculos necessários para a escolha do melhor método de alocação de amostras estratificadas nos estados de São Paulo, Paraná e Rio Grande do Sul. Os dados, a priori, são oriundos do IBGE, da safra mais recente disponível.

O caso de São Paulo

Dados de 2011 da produção agrícola nos municípios do estado de São Paulo (IBGE, 2014) foram aqui utilizados para ilustrar um levantamento amostral agrícola deste tipo. O estado de São Paulo é dividido em 15 mesorregiões criadas pelo IBGE, que são utilizadas apenas para fins estatísticos e não se constituem em entidades político-administrativas autônomas. Segundo esse órgão, mesorregião é uma área individualizada do estado que apresenta

formas de organização do espaço geográfico definidas pelas seguintes dimensões: o processo social, como determinante; o quadro natural, como condicionante; e a rede de comunicação e de lugares, como elemento da articulação espacial. Estas dimensões possibilitam que o espaço delimitado como mesorregião tenha uma identidade regional, definida pela realidade construída ao longo do tempo pela sociedade que ali se formou (IBGE, 2013).

Os três itens escolhidos como sendo de maior interesse no presente estudo são as áreas anuais cultivadas com as culturas da cana-de-açúcar, soja e milho (Tabela 1). Estas culturas, além de estarem entre os dez produtos alimentícios produzidos em maior volume em todo o mundo (FAO, 2014), ocupam as maiores áreas plantadas no Brasil e em São Paulo (IBGE, 2014). Aplicando, então, a ideia da imagem de sensoriamento remoto e do pixel, podemos dividir toda a área de estudo em quadriculas do mesmo tamanho (= pixels).

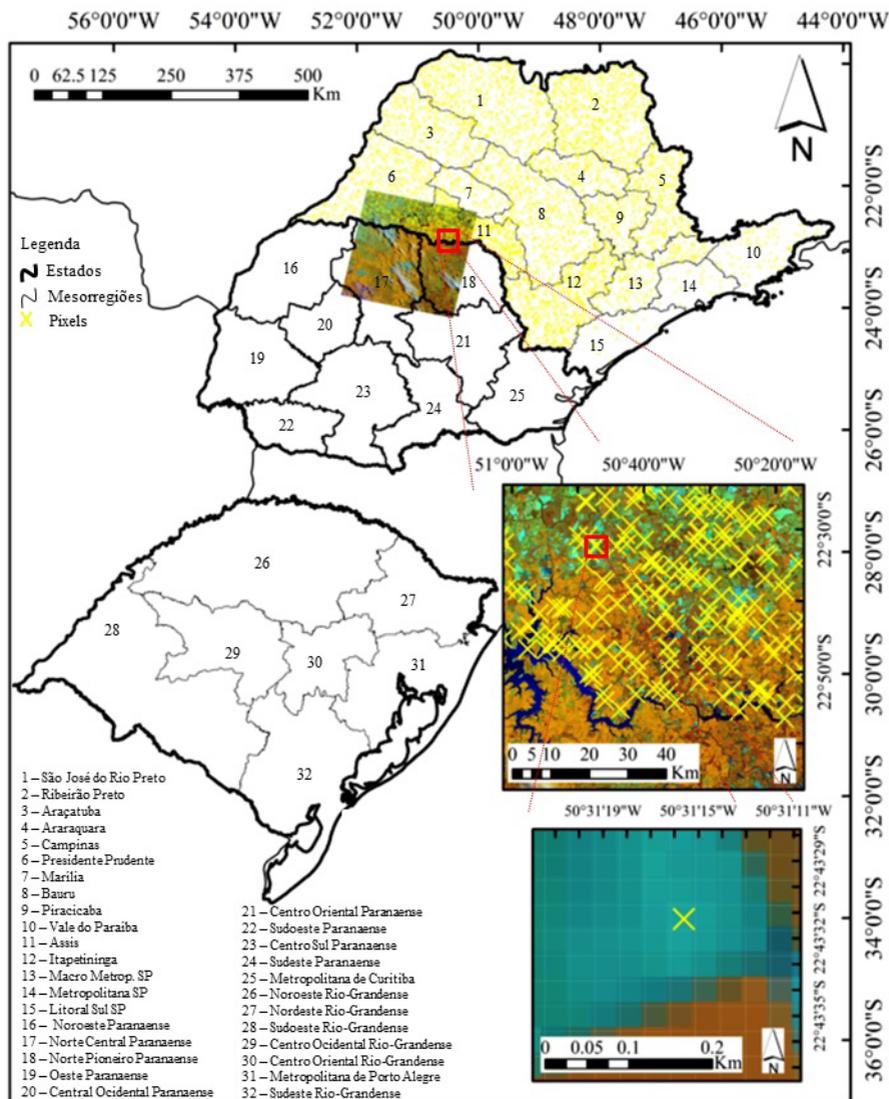


Figura 1. Representação esquemática das mesorregiões dos estados de São Paulo, Paraná e Rio Grande do Sul, com exemplo de pontos distribuídos de forma aleatória sobre SP e detalhes da localização dos pontos em imagem de satélite até o nível de pixel.

Desta forma, ao se traçar um quadriculado imaginário sobre o estado de São Paulo e se contar o número de quadriculas em toda a área e em cada subárea (mesorregião), considerando-os como a população e as subpopulações, pode-se extrair uma *amostra* aleatória ou uma amostra aleatória estratificada. Ao atribuir para cada quadricula uma única classe de ocupação agrícola, pode-se considerar que a probabilidade de ocorrência de cada classe segue uma distribuição hipergeométrica (JOHNSON; KOTZ, 1969) e, portanto, é possível calcular as estatísticas próprias dessa distribuição.

Tabela 1. Área total do território e áreas colhidas (ha) de cana-de-açúcar, milho e soja, e número teórico de pixels de 30 x 30 m, nas mesorregiões e no estado de São Paulo, em 2011 (fonte: IBGE, 2014).

Estrato (mesorregião)	Território	Cana	Milho	Soja	Número teórico de pixels
Araçatuba	1.676.321,1	545.853	32.373	26.909	18.625.790
Araraquara	945.529,9	368.476	12.250	3.260	10.505.888
Assis	1.271.021,0	346.194	198.982	168.409	14.122.456
Bauru	2.672.325,7	598.601	51.431	23.855	29.692.508
Campinas	1.422.612,8	238.300	89.995	6.415	15.806.809
Itapetininga	2.017.216,4	67.919	174.299	78.100	22.413.516
Litoral Sul Paulista	1.320.089,3	9	562	0	14.667.659
Macro Metropolitana Paulista	1.230.960,0	24.814	39.102	3.082	13.677.333
Marília	717.017,1	67.482	11.531	1.794	7.966.857
Metropolitana de São Paulo	929.819,2	0	597	0	10.331.324
Piracicaba	904.563,0	310.428	20.542	2.711	10.050.700
Presidente Prudente	2.403.430,0	455.276	32.014	22.310	26.704.778
Ribeirão Preto	2.753.223,0	1.333.636	64.241	107.058	30.591.367
São José do Rio Preto	2.939.470,0	847.958	64.943	17.202	32.660.778
Vale do Paraíba Paulista	1.617.994,7	895	6.168	200	17.977.719
Total	24.821.593,2	5.205.841	799.030	461.305	275.795.480

A proporção (p_i) de cada cultura é calculada simplesmente dividindo-se a área plantada pela área total no estado, conforme demonstrado na Equação 1.

$$p_i = AC_i / A \quad \text{sendo } i = 1, 2 \text{ ou } 3 \quad [1]$$

onde A é a área total do estado, em hectares (ha); e AC_i é a área do estado ocupada com a cultura i (1 = cana-de-açúcar; 2 = milho; e 3 = soja). Se ordenar as mesorregiões de São Paulo em ordem alfabética e chamar cada uma de estrato h e numerando-os de 1 até 15, os valores da proporção (p_{ih}), de cada cultura i em cada estrato h , podem ser obtidos de forma semelhante, como expresso na Equação 2.

$$p_{ih} = AC_{ih} / A_h \quad \text{sendo } h = 1, 2 \dots 15 \quad [2]$$

onde A_h é a área total do estrato h , em hectares (ha); e AC_{ih} é a área do estrato h ocupada com a cultura i .

No início, se considerar cada cultura separadamente, pode-se supor que cada unidade na população (cada pixel, no presente caso) cai dentro de uma de duas classes C_i - significando que o pixel é ocupado pela cultura i , e C_i' - significando que o pixel não é ocupado pela cultura. Por enquanto não se discute a maneira de decidir pela classificação de cada pixel, apenas assume-se que ela é possível e existe uma única classe para cada pixel. Nesse caso, conforme detalhado em Luiz (2009) e com base nas descrições dos cálculos envolvidos nas estimativas em amostragem de proporções, feitas por Cochran (1977), é possível afirmar que o desvio padrão (s_i), de cada cultura i no estado e o desvio padrão (s_{ih}), de cada cultura i em cada estrato h , podem ser obtidos pelas Equações 3 e 4, respectivamente.

$$s_i = \sqrt{[(p_i \cdot q_i)/n_i]}$$

sendo n_i o tamanho total da amostra e $q_i = (1-p_i)$ [3]

$$s_{ih} = \sqrt{[(p_{ih} \cdot q_{ih})/n_{ih}]}$$

sendo n_{ih} o tamanho da amostra no estrato e $q_{ih} = (1-p_{ih})$ [4]

Ao se dividir a população em estratos, é comum definir como peso do estrato nas estimativas populacionais a proporção entre ele e a população, tomando como base os tamanhos respectivos. Nesse caso, o tamanho é expresso pela área e para calcular os pesos de cada estrato (W_h), basta dividir a área de cada um pela área total do estado, conforme a Equação 5.

$$W_h = A_h / A \quad [5]$$

Para a definição da intensidade amostral por estrato de forma mais vantajosa para o estudo e curso, foi utilizado o conhecimento prévio do comportamento das variáveis a serem observadas, ou seja, as áreas de cana-de-açúcar, milho e soja, no passado recente (Tabela 1). Com base nestes dados, os valores de p_{ih} , s_{ih} e W_h foram calculados e estão expostos na Tabela 2. A partir dos dados da Tabela 2 é possível fazer o cálculo do tamanho de amostra em cada estrato, segundo os critérios da proporcionalidade ao peso W_h e da otimização por cultura em função de p_{ih} e s_{ih} . No caso da alocação proporcional, o tamanho da amostra nos estratos (nr_h) é igual para qualquer cultura, pois só depende do tamanho do estrato em relação ao estado. Para o seu cálculo, basta multiplicar o peso do estrato (W_h) pelo tamanho total da amostra (n_i), que nesse caso foi definido como igual para todas as culturas ($n_1 = n_2 = n_3 = 3.000$). Para a alocação ótima, o tamanho da amostra em cada estrato (no_{ih}) é definido levando-se em conta também o desvio padrão da cultura i em cada caso, como expresso na Equação 6.

$$no_{ih} = n_i \cdot (W_h \cdot s_{ih}) / \sum (W_h \cdot s_{ih}) \quad [6]$$

No exemplo usado por Cochran (1977) e já citado acima, além dos tamanhos de amostra obtidos da alocação ótima individual, foi proposto o cálculo da sua média como forma de atender simultaneamente os requisitos de cada variável. Entretanto, no caso analisado por ele, as variáveis eram mais correlacionadas e apresentavam um padrão de variabilidade menos heterogêneo. No presente caso, para tentar contornar essa variabilidade maior entre as culturas, além de calcular a média, propõe-se também o cálculo de uma alocação pelo máximo entre os tamanhos de amostra obtidos na alocação ótima individual.

Tabela 2. Proporção de área (p_{ih}) e respectivo desvio padrão (s_{ih}) para as culturas de cana, soja e milho, e peso (W_h) de cada estrato no estado de São Paulo, 2011.

Estrato (h)	Mesorregião	p_{ih}			s_{ih}^*			W_h
		Cana	Milho	Soja	Cana	Milho	Soja	
1	Araçatuba	0,3256	0,0193	0,0161	0,0331	0,0097	0,0089	0,0675
2	Araraquara	0,3897	0,0130	0,0034	0,0345	0,0080	0,0041	0,0381
3	Assis	0,2724	0,1566	0,1325	0,0315	0,0257	0,0240	0,0512
4	Bauru	0,2240	0,0192	0,0089	0,0295	0,0097	0,0067	0,1077
5	Campinas	0,1675	0,0633	0,0045	0,0264	0,0172	0,0047	0,0573
6	Itapetininga	0,0337	0,0864	0,0387	0,0128	0,0199	0,0136	0,0813
7	Litoral Sul Paulista	0,0000	0,0004	0,0000	0,0002	0,0015	0,0000	0,0532
8	Macro Metropolitana Paulista	0,0202	0,0318	0,0025	0,0099	0,0124	0,0035	0,0496
9	Marília	0,0941	0,0161	0,0025	0,0206	0,0089	0,0035	0,0289
10	Metropolitana de São Paulo	0,0000	0,0006	0,0000	0,0000	0,0018	0,0000	0,0375
11	Piracicaba	0,3432	0,0227	0,0030	0,0336	0,0105	0,0039	0,0364
12	Presidente Prudente	0,1894	0,0133	0,0093	0,0277	0,0081	0,0068	0,0968
13	Ribeirão Preto	0,4844	0,0233	0,0389	0,0353	0,0107	0,0137	0,1109
14	São José do Rio Preto	0,2885	0,0221	0,0059	0,0320	0,0104	0,0054	0,1184
15	Vale do Paraíba Paulista	0,0006	0,0038	0,0001	0,0017	0,0044	0,0008	0,0652
Total no Estado de SP**		0,2097	0,0322	0,0186	0,0288	0,0125	0,0025	-

* o cálculo supõe amostra aleatória simples tamanho $n_{ih} = 200$ para cada estrato e $n = 3.000$ para o estado.

** os valores se referem à p_i e s_i .

Para obter-se o tamanho de amostra em um estrato pela alocação média (nm_h) e pela alocação máxima (nx_h), basta calcular a média e o máximo, respectivamente, entre os tamanhos de amostra ótimos das culturas naquele estrato (no_{ih}). Entretanto, ao adotar-se o maior valor entre as três culturas (nx_h), ou o valor ótimo para o pior caso, foi encontrado um tamanho final da amostra $nx = 3.998$, o que é 33,3% maior que o tamanho inicial proposto. Para manter o tamanho total na dimensão pré-estabelecida, é proposta a alocação máxima corrigida (nc_h), que é obtida da aplicação da Equação 7.

$$nc_h = nx_h \cdot n / \sum nx_h \quad [7]$$

Tabela 3. Tamanhos de amostra calculados para cada tipo de alocação em cada estrato.

Estrato (mesorregião)	Alocação						
	proporcional (nr_h)	ótima (no_{ih})			média (nm_h)	máxima (nx_h)	corrigida (nc_h)
		cana	milho	soja			
Araçatuba	203	284	186	245	238	284	213
Araraquara	114	167	86	64	106	167	125
Assis	154	205	372	500	359	500	375
Bauru	323	403	296	292	330	403	302
Campinas	172	192	279	111	194	279	209
Itapetininga	244	132	457	452	347	457	343
Litoral Sul Paulista	160	2* (1)	22	2* (0)	8	22	17
Macro Metropolitana Paulista	149	63	174	71	103	174	131
Marília	87	76	73	42	64	76	57
Metropolitana de São Paulo	112	2* (0)	19	2* (0)	6	19	14
Piracicaba	109	155	109	57	107	155	116
Presidente Prudente	290	341	222	268	277	341	256
Ribeirão Preto	333	498	335	618	484	618	464
São José do Rio Preto	355	482	348	260	363	482	362
Vale do Paraíba Paulista	196	4	21	21	15	21	16
Total	3.001	3.006	2.999	3.005	3.001	3.998	3.000

* Segundo Cochran (1977), deve haver pelo menos duas unidades amostrais em cada estrato.

Como se pode ver na Tabela 3, onde aparecem todos esses valores calculados para um tamanho de amostra total proposto de $n = 3.000$, a alocação individual ótima difere bastante de uma cultura para outra e mais ainda quando comparadas à alocação proporcional. Observa-se ainda que, devido ao arredondamento para valores inteiros e o estabelecimento do mínimo da amostra de uma cultura em cada estrato $n_{ih} = 2$, o tamanho total das amostras pode diferir ligeiramente do planejado.

A partir dos tamanhos de amostra distribuídos pelos estratos, podem ser calculados os valores do desvio padrão amostral esperado para cada cultura (Tabela 4).

As Equações 8, 9, 10 e 11 foram utilizadas no cálculo do desvio padrão para cada tipo de alocação: ótima, pela média, pelo máximo e proporcional, respectivamente.

$$s_{otim} = \sqrt{(\sum W_h s_h)^2 / n}$$

[8]

$$s_{med} = \sqrt{\sum [(W_h s_h)^2 / n m_h]}$$

[9]

$$s_{cor} = \sqrt{\sum [(W_h s_h)^2 / n c_h]}$$

[10]

$$s_{prop} = \sqrt{(\sum W_h s_h^2) / n}$$

[11]

Tabela 4. Desvio padrão amostral esperado das médias estimadas.

Alocação	Cana	Milho	Soja
Ótima	0,000432	0,000194	0,000134
Média	0,000486	0,000202	0,000140
Corrigida	0,000462	0,000198	0,000142
Proporcional	0,000486	0,000222	0,000170

A alocação ótima produz os resultados mais precisos, como era de se esperar, entretanto, gera uma alocação para cada cultura, o que não satisfaz a demanda por um único plano amostral para as três culturas a serem estudadas.

Tanto a alocação pela média como pelo máximo (corrigida) resultaram em desvios menores do que aqueles encontrados para a alocação proporcional. Entre elas, a alocação pelo máximo representou menor variação nas estimativas para cana e milho e praticamente se igualou à alocação pela média no caso da soja. Considerando todo o exposto, é preciso lembrar que os valores da Tabela 4 superestimam a precisão das alocações ótima, pela média e pelo máximo, pois elas foram feitas para a variância estimada. Supondo a alocação de uma amostra aleatória simples, não estratificada, de tamanho $n = 3.000$ no estado de São Paulo, apresenta-se na Tabela 5 os valores, em percentagem, do coeficiente de variação para cada uma das culturas i (CV_i) de interesse. A fórmula usada para o cálculo do CV_i , conforme Luiz (2009) é exposta como a Equação 12.

$$CV_i = 100 \cdot \sqrt{[q_i / (p_i \cdot n)]} \quad [1 2]$$

Para o cálculo dos valores do CV_{ir} , de cada cultura i , supondo uma amostra aleatória estratificada com o mesmo tamanho total $n = 3.000$, alocados nos estratos h de forma proporcional ao seu tamanho, utilizou-se primeiramente a equação 13, para o cálculo.

$$CV_{ir} = [100 \cdot \sqrt{\sum (W_h^2 \cdot p_{ih} \cdot q_{ih} / n_{ih})}] / p_i \quad [13]$$

No caso da alocação ser feita pela média ou pelo máximo corrigido dos valores ótimos, conforme já descrito, o cálculo dos CV_{im} e CV_{ic} é praticamente o mesmo, mudando apenas o tamanho da amostra em cada estrato, conforme pode ser observado nas Equações 14 e 15.

$$CV_{im} = [100 \cdot \sqrt{\sum (W_h^2 \cdot p_{ih} \cdot q_{ih} / nm_h)}] / p_i \quad [14]$$

$$CV_{ic} = [100 \cdot \sqrt{\sum (W_h^2 \cdot p_{ih} \cdot q_{ih} / nc_h)}] / p_i \quad [15]$$

Todos os resultados são apresentados na Tabela 5, onde pode ser confirmada a hipótese de que a estratificação melhora a precisão das estimativas amostrais quando baseada em dados anteriores correlacionados com os que se pretende estimar.

Tabela 5. Coeficiente de variação (*CV*), em porcentagem, das estimativas das áreas com cana, milho e soja, em São Paulo, considerando as proporções observadas em 2011 e uma amostra aleatória simples de tamanho $n = 3.000$.

Tipo da amostra	Alocação nos estratos	Cana	Milho	Soja
aleatória simples	-	3,54	10,01	13,27
estratificada	proporcional	3,28	9,79	12,93
estratificada	média	3,09	9,37	10,67
estratificada	máxima corrigida	2,70	7,98	9,34

Com base nos resultados, pode-se afirmar que o método de alocação proposto, proporcional ao valor máximo entre os valores obtidos pela alocação ótima para cada uma das culturas: cana-de-açúcar, milho e soja, é o que produz coeficientes de variação menores, segundo os dados disponíveis de área cultivada em 2011, no estado de São Paulo. É possível até expressar o tamanho da amostra em cada estrato como uma proporção do tamanho total da amostra (R_h), conforme apresentado na Tabela 6. Para o cálculo de R_h , é utilizada a Equação 16.

$$R_h = nc_h / \sum nc_h \quad [16]$$

Dessa maneira, para um estudo que tenha o interesse principal nestas culturas, dado um tamanho de amostra n , condizente com a capacidade de realização dos responsáveis pelo levantamento, basta multiplicar o valor total n pela proporção em cada estrato (R_h), conforme a Equação 17.

$$n_h = R_h \cdot n \quad [17]$$

Isto permite a obtenção do número de pontos amostrais que deve ser alocado em cada estrato (n_h) para obter o coeficiente de variação mínimo para cada uma das culturas. O exemplo para uma amostra de tamanho total $n=5.000$ é apresentado na Tabela 6. Por causa da aproximação nos cálculos, o tamanho total acabou sendo de 5.001 pontos.

Como forma de avaliar como seria o desempenho dessa estratificação, caso fosse utilizada na estimativa de outras culturas que não aquelas utilizadas no planejamento, foram escolhidas a cultura de laranja e as áreas com pastagens, eucalipto e florestas plantadas. Essas atividades foram escolhidas por representarem importantes atividades rurais no estado de São Paulo e no Brasil e ocuparem áreas representativas.

Tabela 6. Proporção da amostra alocada em cada estrato (R_h) calculada em relação ao tamanho máximo entre os valores de alocação ótima para as culturas de cana, milho e soja, e exemplo dos valores alocados em cada estrato (n_h) de uma amostra tamanho $n = 5.000$.

Mesorregião	Estrato (h)	R_h	n_h (exemplo para $n=5.000$)
Araçatuba	1	0,071036	355
Araraquara	2	0,041771	209
Assis	3	0,125063	625
Bauru	4	0,100800	504
Campinas	5	0,069785	349
Itapetininga	6	0,114307	572
Litoral Sul Paulista	7	0,005503	28
Macro Metropolitana Paulista	8	0,043522	218
Marília	9	0,019010	95
Metropolitana de São Paulo	10	0,004752	24
Piracicaba	11	0,038769	194
Presidente Prudente	12	0,085293	426
Ribeirão Preto	13	0,154577	773
São José do Rio Preto	14	0,120560	603
Vale do Paraíba Paulista	15	0,005253	26
Total		1,0	5.001

Além disso, os dados de silvicultura e pecuária provêm de bases diferentes dos dados das culturas agrícolas, o que permite inferir sua independência

quanto à forma de aquisição e, com isso, testar a influência do método de forma mais ampla. No caso da laranja, a base de dados é a mesma, ou seja, a PAM/IBGE, com dados de área colhida em 2011.

Para as áreas ocupadas com florestas plantadas e pastagens (que englobam pastagens naturais e plantadas), foram usados os dados do Censo Agropecuário de 2006. Os dados de eucalipto (ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS, 2009) referem-se ao ano de 2008. A Tabela 7 apresenta os dados de área utilizados.

Tabela 7. Área (ha) colhida de laranja, em 2011, ocupada por eucalipto, em 2008, e ocupada por pastagens e florestas plantadas, em 2005, nas mesorregiões e no estado de São Paulo (IBGE, 2014; ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS, 2009).

Mesorregião	Laranja	Pastagens	Eucalipto	Florestas plantadas
Araçatuba	2.679	683.067	985	2.846
Araraquara	107.839	98.844	43.225	9.892
Assis	4.510	333.458	2.011	8.631
Bauru	83.231	872.324	102.324	87.042
Campinas	95.337	256.642	170.320	13.671
Itapetininga	24.227	487.927	100.012	91.440
Litoral Sul Paulista	0	118.129	4.122	1.716
Macro Metropolitana Paulista	4.693	193.390	208.320	26.605
Marília	1.833	435.015	5.335	3.305
Metropolitana de São Paulo	28	15.611	57.307	8.438
Piracicaba	54.934	139.661	30.632	16.721
Presidente Prudente	205	1.441.404	3.004	7.825
Ribeirão Preto	77.653	317.825	93.640	44.503
São José do Rio Preto	106.103	1.161.837	3.120	3.161
Vale do Paraíba Paulista	680	421.105	110.003	46.768
Total	563.952	6.976.239	934.360	372.564

As áreas ocupadas com cada uso (Tabela 7) foram obtidas de forma totalmente independente do cálculo realizado para a distribuição da amostra entre os estratos, e os usos escolhidos apresentam uma

distribuição proporcional entre as mesorregiões que resultam em correlações maiores ou menores, positivas ou negativas, entre si e com as culturas escolhidas para basear a distribuição da amostra (Tabela 8).

Tabela 8. Coeficiente de correlação entre a área (ha) colhida de cana-de-açúcar, laranja, milho e soja, 2011, ocupada por eucalipto, 2008, e por pastagens e florestas plantadas, 2005, nas mesorregiões e no estado de São Paulo (IBGE, 2014; ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS, 2009).

	Laranja	Pastagens	Eucalipto	Floresta	Soja	Milho
Cana	0,57	0,40	-0,16	0,08	0,40	0,13
Milho	0,07	0,09	0,09	0,35	0,82	
Soja	-0,04	0,03	-0,13	0,22		
Floresta	0,15	0,09	0,49			
Eucalipto	0,17	-0,30				
Pastagens	0,07					

Podemos, então, estimar os coeficientes de variação esperados para cada um destes tipos de ocupação do solo, caso fosse aplicada a estratificação em mesorregiões e utilizada uma amostra de tamanho 5.000, distribuída nos mesmos moldes que para cana-de-açúcar, milho e soja, como apresentado anteriormente (Tabela 5). Os resultados estão apresentados na Tabela 9. Para a cultura da laranja, o padrão se repete e a melhor forma de distribuir uma amostra de pixels tamanho 3.000 no estado de São Paulo, quando o objetivo é diminuir o CV da estimativa de área, é a chamada amostra aleatória estratificada, na qual os estratos são as mesorregiões e a alocação dos pontos amostrais pelos estratos se faz pelo método da máxima corrigida.

Para os outros três usos estudados, entretanto, os resultados mostram que a alocação dos pontos amostrais nos estratos tanto pelo método da média, recomendado por Cochran (1977), quanto pelo da máxima corrigida, proposto aqui, resultaram em valores de CV maiores que os produzidos pela amostra aleatória simples (não estratificada) e pela amostra estratificada com alocação dos pontos pelo método proporcional.

Isso pode ser explicado pela correlação baixa ou até negativa da área ocupada com esses tipos de uso em relação àquelas das culturas originalmente utilizadas para o cálculo das alocações pela média ou pelo máximo (Tabela 8).

Tabela 9. Coeficiente de variação (*CV*), em porcentagem, das estimativas das áreas com laranja, eucalipto, pastagens e florestas plantadas, em São Paulo, considerando uma amostra aleatória simples de tamanho $n = 3.000$.

Tipo da amostra	Alocação nos estratos	Laranja	Eucalipto	Pastagens	Florestas
aleatória simples	-	11,97	9,23	2,92	14,79
estratificada	proporcional	11,78	8,99	2,72	14,69
estratificada	média	11,46	16,80	4,14	24,82
estratificada	máxima corrigida	9,76	13,00	3,34	20,26

Esses resultados permitem expandir a afirmação de Cochran (1977), que disse que “a melhor alocação para um item, geralmente não será a melhor para outro”, e reescrevê-la da seguinte forma: a melhor alocação para um grupo de itens, geralmente não será a melhor para outros. Adicionalmente, ressalta-se que a alocação baseada em um grupo de itens terá mais chances de se mostrar eficiente para outros itens, à medida em que os últimos apresentem maior correlação positiva com os primeiros.

O caso do Paraná

Dados de 2012 da produção agrícola nos municípios do estado do Paraná (IBGE, 2014) foram utilizados para ilustrar um levantamento amostral agrícola deste tipo. O estado do Paraná é dividido em 10 mesorregiões (IBGE, 2013). Da mesma maneira que se adotou para São Paulo, os três itens escolhidos como de maior interesse neste estudo são as áreas anuais cultivadas com as culturas da cana-de-açúcar, soja e milho (Tabela 1). No Paraná, as culturas temporárias com as maiores áreas colhidas em 2012 foram, pela ordem, soja, milho, trigo e cana-de-açúcar (IBGE, 2014), também demonstrando a importância estadual das três escolhidas

nesse estudo.

Tabela 10. Área (ha) total do território e colhida de cana-de-açúcar, milho e soja, e número teórico de pixels de 30 x 30 m, nas mesorregiões e no estado do Paraná, em 2012 (fonte: IBGE, 2014).

Estrato (mesorregião)	Território	Cana	Milho	Soja	Número teórico de pixels
Centro Ocidental Paranaense	1.193.730	29.592	374.751	601.683	13.263.667
Centro Oriental Paranaense	2.185.050	323	147.150	451.420	24.278.333
Centro-Sul Paranaense	2.640.700	617	186.123	444.975	29.341.111
Metropolitana de Curitiba	2.283.100	851	135.915	73.068	25.367.778
Noroeste Paranaense	2.448.870	335.382	133.612	169.997	27.209.667
Norte Central Paranaense	2.455.280	179.505	576.128	751.222	27.280.889
Norte Pioneiro Paranaense	1.572.710	100.007	337.891	397.400	17.474.556
Oeste Paranaense	2.285.360	6.629	809.729	952.900	25.392.889
Sudeste Paranaense	1.702.540	283	134.860	238.220	18.917.111
Sudoeste Paranaense	1.165.010	2.320	175.571	375.920	12.944.556
Total	19.932.350	655.509	3.011.730	4.456.805	221.470.556

Da mesma forma que no caso anterior, para SP, a proporção (p_i) de cada cultura é calculada simplesmente dividindo-se a área plantada pela área total no estado, conforme demonstrado na Equação 1.

Se ordenar as mesorregiões do Paraná em ordem alfabética e chamar cada uma de estrato h e numerando-os de 1 até 10, os valores da proporção (p_{ih}), de cada cultura i em cada estrato h , podem ser obtidos de forma semelhante, como expresso na Equação 2. Como já foi visto, é possível afirmar que o desvio padrão (s_i), de cada cultura i no estado e o desvio padrão (s_{ih}), de cada cultura i em cada estrato h , podem ser obtidos pelas Equações 3 e 4, respectivamente. Para calcular os pesos de cada estrato (W_h), basta dividir a área de cada um pela área total do estado, conforme a Equação 5. Para a definição da intensidade amostral por estrato de forma mais vantajosa para o estudo e curso, foi utilizado o conhecimento prévio do comportamento das variáveis a serem observadas, ou seja, as áreas de cana-de-açúcar, milho e soja, no passado recente (Tabela 10). Com base nestes dados, os valores de p_{ih} , s_{ih} e W_h foram calculados e

estão expostos na Tabela 11.

Tabela 11. Proporção de área (p_{ih}) e respectivo desvio padrão (s_{ih}) para as culturas de cana, soja e milho, e peso (W_h) de cada estrato no estado do Paraná, 2012.

Estrato (h)	Mesorregião	p_{ih}			s_{ih}^*			W_h
		Cana	Milho	Soja	Cana	Milho	Soja	
1	Centro Ocidental Paranaense	0,0248	0,3139	0,5040	0,0070	0,0208	0,0224	0,0599
2	Centro Oriental Paranaense	0,0001	0,0673	0,2066	0,0005	0,0112	0,0181	0,1096
3	Centro-Sul Paranaense	0,0002	0,0705	0,1685	0,0007	0,0114	0,0167	0,1325
4	Metropolitana de Curitiba	0,0004	0,0595	0,0320	0,0009	0,0106	0,0079	0,1145
5	Noroeste Paranaense	0,1370	0,0546	0,0694	0,0154	0,0102	0,0114	0,1229
6	Norte Central Paranaense	0,0731	0,2346	0,3060	0,0116	0,0190	0,0206	0,1232
7	Norte Pioneiro Paranaense	0,0636	0,2148	0,2527	0,0109	0,0184	0,0194	0,0789
8	Oeste Paranaense	0,0029	0,3543	0,4170	0,0024	0,0214	0,0221	0,1147
9	Sudeste Paranaense	0,0002	0,0792	0,1399	0,0006	0,0121	0,0155	0,0854
10	Sudoeste Paranaense	0,0020	0,1507	0,3227	0,0020	0,0160	0,0209	0,0584
Total no estado do PR**		0,2097	0,0322	0,0186	0,0025	0,0051	0,0059	-

* o cálculo supõe amostra aleatória simples tamanho $n_{ih} = 500$ para cada estrato e $n = 5.000$ para o estado.

** os valores se referem à p_i e s_i .

A partir dos dados da Tabela 11 foi possível fazer o cálculo do tamanho de amostra em cada estrato, segundo os critérios da proporcionalidade ao peso W_h e da otimização por cultura em função de p_{ih} e s_{ih} . No caso da alocação proporcional, o tamanho da amostra nos estratos (nr_h) é igual para qualquer cultura, pois só depende do tamanho do estrato em relação ao estado. Para o seu cálculo, basta multiplicar o peso do estrato (W_h) pelo tamanho total da amostra (n_i), que nesse caso foi definido como igual para todas as culturas ($n_1 = n_2 = n_3 = 5.000$).

Para a alocação ótima, o tamanho da amostra em cada estrato (no_{ih}) é definido levando-se em conta também o desvio padrão da cultura i em cada caso, como expresso na Equação 6. Novamente, para obter-se o tamanho de amostra em um estrato pela alocação média (nm_h) e pela alocação máxima (nx_h), basta calcular a média e o máximo,

respectivamente, entre os tamanhos de amostra ótimos das culturas naquele estrato (no_{ih}).

Entretanto, ao adotar-se o maior valor entre as três culturas (nx_h), ou o valor ótimo para o pior caso, foi encontrado um tamanho final da amostra $nx = 7.616$, o que é 52,3% maior que o tamanho inicial proposto. Para manter o tamanho total na dimensão pré-estabelecida, propõem-se a alocação máxima corrigida (nc_h), que é obtida da aplicação da Equação 7.

Tabela 12. Tamanhos de amostra calculados para cada tipo de alocação em cada estrato.

Estrato (mesorregião)	Alocação						
	proporcional (nr_h)	ótima (no_{ih})			média (nm_h)	máxima (nx_h)	corrigida (nc_h)
		cana	milho	soja			
Centro Ocidental Paranaense	299	394	424	394	404	424	279
Centro Oriental Paranaense	548	56	419	584	353	584	384
Centro-Sul Paranaense	662	86	518	653	419	653	429
Metropolitana de Curitiba	573	93	414	265	257	414	272
Noroeste Paranaense	614	1.785	426	411	874	1.785	1.172
Norte Central Paranaense	616	1.355	797	747	966	1.355	890
Norte Pioneiro Paranaense	395	814	495	451	587	814	534
Oeste Paranaense	573	261	837	744	614	837	550
Sudeste Paranaense	427	47	352	390	263	390	256
Sudoeste Paranaense	292	110	319	360	263	360	236
Total	4.999	5.001	5.001	4.999	5.000	7.616	5.002

Como se pode ver na Tabela 12, onde aparecem todos esses valores calculados para um tamanho de amostra total proposto de $n = 5.000$, a alocação individual ótima difere bastante de uma cultura para outra, e mais ainda quando comparadas à alocação proporcional. Observa-se ainda que, devido ao arredondamento para valores inteiros, o tamanho total das amostras pode diferir ligeiramente do planejado.

A partir dos tamanhos de amostra distribuídos pelos estratos, podem ser calculados os valores do desvio padrão amostral esperado para

cada cultura (Tabela 13). As Equações 8, 9, 10 e 11 foram utilizadas no cálculo do desvio padrão para cada tipo de alocação: ótima, pela média, pelo máximo e proporcional, respectivamente.

Tabela 13. Desvio padrão amostral esperado das médias estimadas.

Alocação	Cana	Milho	Soja
Não estratificada	0,002522	0,005065	0,005892
Proporcional	0,002432	0,004829	0,005568
Média	0,002019	0,004872	0,005752
Máxima corrigida	0,001954	0,005022	0,005887
Ótima	0,001673	0,004633	0,005372

Todas as formas de estratificação produzem desvios menores que a forma não estratificada. A alocação ótima produz os resultados mais precisos, como era de se esperar, entretanto, gera uma alocação para cada cultura, o que não satisfaz a demanda por um único plano amostral para as três culturas a serem estudadas.

A alocação pelo máximo (corrigida) resulta em desvio menor do que aquele encontrado para as demais formas de alocação, exceto a ótima, para a cana-de-açúcar. Já para milho e soja, cuja distribuição espacial entre as mesorregiões é mais homogênea, os menores valores de desvio são oriundos da alocação proporcional. Considerando todo o exposto, é preciso lembrar que os valores da Tabela 13 superestimam a precisão das alocações ótima, pela média e pelo máximo (corrigida), pois elas foram feitas para a variância estimada e utilizaram dados do passado.

Supondo a alocação de uma amostra aleatória simples, não estratificada, de tamanho $n=5.000$ no estado do Paraná, apresenta-se na Tabela 5 os valores, em percentagem, do coeficiente de variação para cada uma das culturas i (CV_i) de interesse. A fórmula usada para o cálculo do CV_i , conforme Luiz (2009), foi exposta pela Equação 12.

Para o cálculo dos valores do $CV_{i,r}$, de cada cultura i , supondo uma amostra aleatória estratificada com o mesmo tamanho total $n=5.000$,

alocados nos estratos h de forma proporcional ao seu tamanho, utilizou-se primeiramente a equação 13, para o cálculo. No caso da alocação ser feita pela média ou pelo máximo corrigido dos valores ótimos, conforme já descrito, o cálculo dos CV_{im} e CV_{ic} é praticamente o mesmo, mudando apenas o tamanho da amostra em cada estrato, conforme foi observado nas Equações 14 e 15.

Todos os resultados são apresentados na Tabela 14, onde pode ser confirmada a hipótese de que a estratificação melhora a precisão das estimativas amostrais quando baseada em dados anteriores correlacionados com os que se pretende estimar.

Tabela 14. Coeficiente de variação (CV), em porcentagem, das estimativas das áreas com cana, milho e soja, no Paraná, considerando as proporções observadas em 2012 e uma amostra aleatória simples de tamanho $n = 5.000$.

Tipo da amostra	Alocação nos estratos	Cana	Milho	Soja
aleatória simples	-	7,6691	3,3521	2,6353
estratificada	proporcional	7,3946	3,1962	2,4902
estratificada	média	6,1406	3,2245	2,5725
estratificada	máxima corrigida	5,9419	3,3238	2,6327

Com base nos resultados, pode-se afirmar que o método de alocação proposto, proporcional ao valor máximo entre os valores obtidos pela alocação ótima para a cultura da cana-de-açúcar é o que produz coeficiente de variação menor, já para milho e soja, os menores valores de CV são provenientes da alocação proporcional, segundo os dados disponíveis de área cultivada em 2012, no estado do Paraná. Como a cana-de-açúcar é a cultura com maior CV , optou-se por adotar a forma de alocação que minimiza esse coeficiente nesse caso, que é o do máximo corrigido. Os valores a serem adotados são os nc_h da Tabela 12.

O caso do Rio Grande do Sul

Dados de 2012 da produção agrícola nos municípios do estado do Rio Grande do Sul (IBGE, 2014), foram utilizados para elaborar um

levantamento amostral agrícola deste tipo. O estado do RS é dividido em sete mesorregiões (IBGE, 2013). Diferentemente do que foi adotado para os estados de São Paulo e Paraná, foram escolhidas duas culturas anuais de maior interesse para este estudo: a soja e o milho (Tabela 15). No RS a cana-de-açúcar ocupa uma área pouco expressiva em comparação às demais culturas temporárias. Em 2012, as culturas com as maiores áreas colhidas no RS foram, pela ordem: soja, milho, arroz e trigo (IBGE, 2014). Embora ligeiramente diferente dos casos de SP e PR, os dados demonstram a importância estadual de duas das culturas escolhidas nesse estudo.

Tabela 15. Área (ha) total do território e colhida de milho e soja, e número teórico de pixels de 30 x 30 m, nas mesorregiões e no estado do Rio Grande do Sul, em 2012 (fonte: IBGE, 2014).

Estrato (mesorregião)	Território	Milho	Soja	Número teórico de pixels
Noroeste rio-grandense	6.494.076	610.442	2.747.600	72.156.400
Nordeste rio-grandense	2.586.751	161.780	218.636	28.741.678
Centro Ocidental rio-grandense	2.595.575	54.845	564.075	28.839.722
Centro Oriental rio-grandense	1.721.175	120.920	196.094	19.124.167
Metropolitana de Porto Alegre	2.986.062	59.795	42.994	33.178.467
Sudoeste rio-grandense	6.269.998	38.700	354.448	69.666.644
Sudeste rio-grandense	4.234.859	72.738	145.400	47.053.989
Total do estado	26.888.498	1.119.220	4.269.247	298.761.089

Da mesma forma que nos casos de São Paulo e Paraná, a proporção (p_i) de cada cultura é calculada simplesmente dividindo-se a área plantada pela área total no estado, conforme demonstrado na Equação 1. Se as mesorregiões do RS forem colocadas em ordem alfabética e denominar-se cada uma de estrato h e numerando-os de 1 até 10, os valores da proporção (p_{ih}), de cada cultura i em cada estrato h , podem ser obtidos de forma semelhante, como expresso na Equação 2. Como já demonstrado anteriormente, é possível determinar que o desvio padrão (s_i), de cada cultura i no estado e o desvio padrão (s_{ih}), de cada cultura i em cada estrato h , podem ser obtidos pelas Equações 3 e 4, respectivamente. Para calcular os pesos de cada estrato (W_h), é necessário somente dividir

a área de cada um pela área total do estado, conforme a Equação 5.

A definição de qual intensidade amostral por estrato é mais vantajosa para o estudo foi a utilização do conhecimento prévio do comportamento das variáveis a serem observadas, ou seja, as áreas de milho e soja, no passado recente (Tabela 16). Com base nestes dados, os valores de p_{ih} , s_{ih} e W_h foram calculados e estão expostos na Tabela 16.

Tabela 16. Proporção de área (p_{ih}) e respectivo desvio padrão (s_{ih}) para as culturas de soja e milho, e peso (W_h) de cada estrato no Estado do Rio Grande do Sul, 2012.

Estrato (h)	Mesorregião	p_{ih}		s_{ih}^*		W_h
		Milho	Soja	Milho	Soja	
1	Noroeste rio-grandense	0,0940	0,4231	0,0109	0,0185	0,2415
2	Nordeste rio-grandense	0,0625	0,0845	0,0091	0,0104	0,0962
3	Centro Ocidental rio-grandense	0,0211	0,2173	0,0054	0,0154	0,0965
4	Centro Oriental rio-grandense	0,0703	0,1139	0,0096	0,0119	0,0640
5	Metropolitana de Porto Alegre	0,0200	0,0144	0,0052	0,0045	0,1111
6	Sudoeste rio-grandense	0,0062	0,0565	0,0029	0,0086	0,2332
7	Sudeste rio-grandense	0,0172	0,0343	0,0049	0,0068	0,1575

* o cálculo supõe amostra aleatória simples tamanho $n_{ih} = 714$ para cada estrato e $n = 5.000$ para o estado.

A partir dos dados da Tabela 16 foi possível fazer o cálculo do tamanho de amostra em cada estrato, segundo os critérios da proporcionalidade ao peso W_h e da otimização por cultura em função de p_{ih} e s_{ih} . No caso da alocação proporcional, o tamanho da amostra nos estratos (nr_h) é igual para qualquer cultura, pois só depende do tamanho do estrato em relação ao estado. Para o seu cálculo, basta multiplicar o peso do estrato (W_h) pelo tamanho total da amostra (n_i), que nesse caso foi definido como igual para todas as culturas ($n_1 = n_2 = 5.000$). Para a alocação ótima, o tamanho da amostra em cada estrato (no_{ih}) é definido levando-se em conta também o desvio padrão da cultura i em cada caso, como expresso na Equação 6.

Novamente, para obter-se o tamanho de amostra em um estrato pela alocação média (nm_h) e pela alocação máxima (nx_h), basta calcular a média e o máximo, respectivamente, entre os tamanhos de amostra ótimos das culturas naquele estrato (no_{ih}). Entretanto, ao adotar-se o maior valor entre as três culturas (nx_h), ou o valor ótimo para o pior caso, foi encontrado um tamanho final da amostra $nx = 7.616$, o que é 52,3% maior que o tamanho inicial proposto. Para manter o tamanho total na dimensão pré-estabelecida, é proposta a alocação máxima corrigida (nc_h), que é obtida da aplicação da Equação 7.

Tabela 17. Tamanhos de amostra calculados para cada tipo de alocação em cada estrato.

Estrato (mesorregião)	Alocação					
	proporcional (nr_h)	ótima (no_{ih})		média (nm_h)	máxima (nx_h)	corrigida (nc_h)
		milho	soja			
Noroeste rio-grandense	1.208	1.976	1.975	1.976	1.976	1.749
Nordeste rio-grandense	481	653	443	548	653	578
Centro Ocidental rio-grandense	483	389	659	524	659	583
Centro Oriental rio-grandense	319	459	337	398	459	406
Metropolitana de Porto Alegre	555	436	219	328	436	386
Sudoeste Rio-grandense	1.166	512	891	702	891	789
Sudeste Rio-grandense	787	574	475	525	574	508
Total	4.999	4.999	4.999	5.001	5.648	4.999

Como se pode ver na Tabela 17, onde aparecem todos esses valores calculados para um tamanho de amostra total proposto de $n = 5.000$, a alocação individual ótima difere bastante de uma cultura para outra, e mais ainda quando comparadas à alocação proporcional. Observa-se ainda que, devido ao arredondamento para valores inteiros, o tamanho total das amostras pode diferir ligeiramente do planejado.

A partir dos tamanhos de amostra distribuídos pelos estratos, podem ser calculados os valores do desvio padrão amostral esperado para cada cultura (Tabela 18). As Equações 8, 9, 10 e 11 foram utilizadas no

cálculo do desvio padrão para cada tipo de alocação: ótima, pela média, pelo máximo e proporcional, respectivamente.

Tabela 18. Desvio padrão amostral esperado das médias estimadas.

Alocação	Milho	Soja
não estratificada	0,087606	0,042024
média	0,061509	0,027314
máxima corrigida	0,061920	0,027356

Todas as formas de estratificação produzem desvios menores que a forma não estratificada. A alocação ótima produz os resultados mais precisos, como era de se esperar; entretanto, gera uma alocação para cada cultura, o que não satisfaz a demanda por um único plano amostral para as três culturas a serem estudadas. A alocação pelo máximo (corrigida) resulta em desvio menor do que aquele encontrado para as demais formas de alocação, exceto a ótima, para a cana-de-açúcar. Já para milho e soja, cuja distribuição espacial entre as mesorregiões é mais homogênea, os menores valores de desvio são oriundos da alocação proporcional.

Supondo-se a alocação de uma amostra aleatória simples, não estratificada, de tamanho $n=5.000$ no estado do Rio Grande do Sul, apresentam-se na Tabela 14 os valores, em percentagem, do coeficiente de variação (CV_i) para cada uma das culturas i de interesse. A fórmula usada para o cálculo do CV , conforme Luiz (2009), foi exposta pela Equação 12.

Para o cálculo dos valores do $CV_{i,r}$, de cada cultura i , supondo uma amostra aleatória estratificada com o mesmo tamanho total $n=5.000$, alocados nos estratos h de forma proporcional ao seu tamanho, utilizou-se primeiramente a equação 13 para o cálculo.

No caso da alocação ser feita pela média ou pelo máximo corrigido dos valores ótimos, conforme já descrito, o cálculo dos CV_{im} e CV_{ic} é

praticamente o mesmo, mudando apenas o tamanho da amostra em cada estrato, conforme foi observado nas Equações 14 e 15.

Todos os resultados são apresentados na Tabela 19, onde pode ser confirmada a hipótese de que a estratificação melhora a precisão das estimativas amostrais quando baseada em dados anteriores correlacionados com os que se pretende estimar.

Tabela 19. Coeficiente de variação (CV), em porcentagem, das estimativas das áreas com milho e soja, no Rio Grande do Sul, considerando as proporções observadas em 2012 e uma amostra aleatória simples de tamanho $n = 5.000$.

Tipo da amostra	Alocação nos estratos	Milho	Soja
aleatória simples	-	8,84%	4,32%
estratificada	média	7,94%	3,53%
estratificada	máxima corrigida	7,99%	3,53%

Com base nos resultados, pode-se afirmar que o método da alocação pela média das alocações ótimas para as culturas do milho e da soja é o que produziu o menor coeficiente de variação para ambos cultivos, segundo os dados disponíveis de área colhida em 2012 no estado do RS.

Conclusões

As técnicas estatísticas de estratificação demonstraram proporcionar ganhos de precisão nas estimativas de área ocupada por múltiplas culturas de interesse em levantamentos agrícolas por amostragem auxiliados por sensoriamento remoto.

O presente trabalho disponibiliza as bases teóricas e práticas para metodologias voltadas à obtenção de estimativas de áreas agrícolas, utilizando técnicas estatísticas de amostragem, conjugadas com estratificação e auxiliadas por sensoriamento remoto, quando se tem três

culturas de interesse no mesmo levantamento.

Bancos de dados como os do IBGE são imprescindíveis para a consecução de objetivos como os dos levantamentos amostrais de áreas com culturas agrícolas via técnicas de estratificação, como exemplificadas no presente estudo.

O tipo de alocação por estratos que resulta em menores valores conjuntos de CV em levantamentos em que se busca estimar ao mesmo tempo as áreas ocupadas por mais de uma cultura, varia em função da proporção do território ocupado com cada cultura em cada estado.

Em São Paulo, o método de alocação proposto do valor máximo corrigido foi o que resultou no menor CV para as três culturas. No Paraná, o método de alocação proporcional foi a melhor opção para soja e milho, enquanto isso, a cana, de menor expressão nesse estado, teve menor CV quando a alocação se deu pelo método do máximo corrigido. No caso do Rio Grande do Sul, onde a cana-de-açúcar não foi considerada nos cálculos para a alocação, o método que resultou no menor CV tanto para soja como para milho foi o da média entre os valores ótimos para cada cultura.

Referências

ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS. **Anuário estatístico da ABRAF**: ano base 2008. Brasília, DF: 2009. 120 p.

IBGE. **Nomes Geográficos do Brasil - Divisão Político-Administrativa**. Rio de Janeiro, 2013. Disponível em: <<http://www.ngb.ibge.gov.br/Default.aspx?pagina=divisao>>. Acesso em: 31 mai. 2013.

IBGE. **Sistema IBGE de Recuperação Eletrônica (SIDRA)**. Rio de Janeiro, 2014. Disponível em: <<http://www.sidra.ibge.gov.br>>. Acesso em: 29 out. 2014.

COCHRAN, W. G. **Sampling techniques**. 3ed. New York: J. Wiley, 1977. 428p.

FAO. FAOSTAT database. Disponível em: <http://faostat3.fao.org/browse/rankings/commodities_by_regions/E>. Acesso em: 05 nov. 2014.

FAO. **Multiple frame agricultural surveys**: volume 1 current survey based on area and list sampling methods. Rome, 1996. 119 p. (FAO Statistical Development Series, 7).

GIANNOTTI, M. A. **Geotecnologias na análise de impactos sócios-ambientais: o caso da queima da cana-de-açúcar na região de Piracicaba**. 2001. 147 f. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. Disponível em: <<http://mtc-m12.sid.inpe.br/col/sid.inpe.br/jeferson/2003/07.17.08.20/doc/publicacao.pdf>>. Acesso em: 10 nov. 2014.

JOHNSON, N. L.; KOTZ, S. **Discrete distributions**. New York: J. Wiley, 1969. 328 p.

LUIZ, A. J. B. **Estatísticas agrícolas por amostragem auxiliadas pelo sensoriamento remoto**. 2009. 112 f. Tese (Doutorado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. Disponível em: <<http://mtc-m18.sid.inpe.br/col/sid.inpe.br/jeferson/2004/02.04.16.08/doc/publicacao.pdf>>. Acesso em: 12 nov. 2014.

Embrapa

Meio Ambiente

Ministério da
**Agricultura, Pecuária
e Abastecimento**

GOVERNO FEDERAL
BRASIL
PÁTRIA EDUCADORA