

# Sistema de amostragem em imagem como base para estimativa de áreas de culturas agrícolas no município de Ipuã-SP

JOSÉ CARLOS NEVES EPIPHANIO<sup>1</sup>

OTON OSÓRIO DE BARROS NETO<sup>1</sup>

ALFREDO JOSÉ BARRETO LUIZ<sup>2</sup>

ANTONIO ROBERTO FORMAGGIO<sup>1</sup>

<sup>1</sup>INPE - Instituto Nacional de Pesquisas Espaciais  
Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brasil  
{epiphani, on, formag}@ltid.inpe.br

<sup>2</sup>Embrapa Meio Ambiente (doutorando no INPE)  
barreto@ltid.inpe.br

**Abstract** Remote sensing-based techniques can be used as a complementary tool for crop area estimating. However, the right balance between the extent of a crop in some region and the remote sensing data characteristics, such as spatial resolution, must be observed. In this paper it is presented the application of a method for crop area estimation based on data sampling considering the discrete and regular distribution of pixels in remote sensing images. Results show that it is possible to have differences around 4% between the simulated area of 110 soybean and 12 corn fields and the estimated area using the proposed method. Sample size impact on data variability is discussed as well as the confidence intervals for population estimation based on a sample and on a hypergeometric distribution.

**Keywords:** crop area, agricultural survey, sampling techniques, remote sensing image.

## 1 Introdução

A estimativa da área de culturas agrícolas é uma importante variável no estabelecimento de instrumentos de políticas e negócios agrícolas. Muitas decisões sobre importações ou exportações, manutenção de estoques reguladores, estratégias para os diversos insumos, alocação de créditos e outros aspectos são estabelecidas com base nas estatísticas agrícolas (Pino, 1999). Porém, dada a natureza fragmentada da distribuição dos talhões, o levantamento preciso da área de uma determinada cultura é dificultado.

Há diversos métodos para estimar a área de culturas agrícolas. No Brasil, o método mais usual é o Levantamento Sistemático da Produção Agropecuária (LSPA), realizado pelo IBGE (Instituto Brasileiro de Geografia e Estatística). Este método baseia-se em informações dos agentes ligados ao meio rural, como pessoal ligado a bancos de crédito agrícola, à assistência técnica, à comercialização de insumos. Parte dos níveis mais básicos, como municípios ou pequenas regiões, e vai sendo generalizado para os níveis regionais, estaduais e, finalmente, nacional. Tem a vantagem de, no nível mais básico, coletar as informações com agentes que estão próximos ao meio no qual está sendo gerada a produção agrícola. A principal desvantagem é que não permite uma aferição mensurável das áreas. Com isso, não são possíveis análises

quantitativas dos erros e incertezas envolvidos (Ippoliti-Ramilo et al., 1999).

No próprio meio governamental tem havido um esforço para definir mecanismos alternativos que venham apoiar o atual sistema de levantamento agrícola. A FAO (Food and Agriculture Organization of the United Nations) preconiza métodos baseados em modelos amostrais para os levantamentos agrícolas. Nesses métodos é possível definir um sistema objetivo de levantamento que possa ser quantitativamente aferido. No caso do Brasil, o IBGE vem aplicando um método amostral para o levantamento agrícola no Estado do Paraná (FAO, 1998). O método, denominado PREVS – Sistema Amostral de Levantamento Agrícola, baseia-se em uma estratificação do Estado segundo o padrão de uso, e na definição de um número de segmentos a serem visitados em campo. Os produtos de sensoriamento remoto, como imagens de satélite e fotografias aéreas, são usados nas fases de estratificação e de levantamento dos segmentos no campo. Cada segmento pode ser composto de diversos talhões, com múltiplos usos. Os dados dos segmentos, que em geral têm um delineamento com limites físicos bem definidos, são tratados estatisticamente, gerando uma previsão de safras para o Estado. Os segmentos têm tamanhos variáveis em função do estrato, indo de 1 a 8 km<sup>2</sup> em média. Para o Estado do Paraná são definidos em torno de 600 segmentos, abrangendo uma área de aproximadamente 0,5% da superfície estadual.

O sistema PREVS, sendo um sistema amostral, permite a geração de estatísticas de aferição da previsão feita (FAO, 1998). De fato, um sistema que se baseia em estatísticas derivadas de levantamentos de campo oferece padrões mais quantitativos de avaliação do que os subjetivos (FAO, 1996). No sistema em apreço a definição do painel amostral é uma etapa crucial, pois é a partir daí que são definidas as unidades menores subsequentes até o detalhamento do segmento a ser efetivamente levantado em campo. Neste processo, os produtos de sensoriamento remoto desempenham importante papel. Num estágio inicial, as imagens de satélite são utilizadas para a estratificação estadual. Tais produtos são especialmente apropriados em função de sua abrangência geográfica e características físicas, como resolução espectral e espacial. Num estágio final, o uso de fotografias aéreas em escalas grandes (tipicamente 1:10.000) têm sido utilizadas, a partir de ampliações de fotografias aéreas obtidas originalmente em escalas menores (por exemplo 1:40.000). Esta última fase do processo apresenta o inconveniente temporal, ou seja, em geral há uma defasagem entre a aquisição da fotografia aérea e os trabalhos de campo. Em geral, a atualização fotográfica esbarra em problemas logísticos e de custos. A aplicação da metodologia pelo IBGE conta com um painel amostral fixo definido sobre imagens de satélite, e com fotografias aéreas defasadas de vários anos.

A definição de um painel amostral onde o desenho dos segmentos e seu posterior detalhamento em campo é feito em produtos de sensoriamento remoto, seja em imagens de satélite ou com o uso de uma fotografia aérea, é um processo trabalhoso, particularmente em terrenos de topografia mais irregular. Uma possibilidade alternativa ao delineamento dos segmentos é assumir uma imagem de satélite como um universo amostral regular, com os pixels da imagem servindo como elementos amostrais regularmente espaçados, de área fixa, passíveis de identificação e de localização. Neste caso, o número de pixels pertencentes a certa classe determina a proporcionalidade da área ocupada por aquela classe em relação ao universo amostral.

Dada a disponibilidade de um conjunto de imagens de satélite, é possível, pelo menos teoricamente, atingir um razoável nível de precisão de interpretação do uso do solo. Como um pixel amostral agrícola não ocorre isoladamente, mas sim dentro de um talhão constituído de muitos pixels, com uma conformação tipicamente agrícola, há uma maior facilidade para sua

identificação. Além disso, outros pontos amostrais servem de apoio à interpretação de um dado pixel ou talhão.

Outro auxiliar importante na caracterização de pontos amostrais é que os sistemas de GPS permitem uma vistoria de campo com maior rapidez e precisão. Tais pontos de caracterização de campo servem de padrões para a interpretação das imagens, num processo de retro-alimentação.

Uma das hipóteses básicas do presente trabalho é que é possível ter uma boa avaliação da área agrícola municipal para um certo número de culturas. Para isso é necessário que sejam compatibilizadas as características dos talhões agrícolas com as características dos produtos de sensoriamento remoto utilizados, particularmente em relação à resolução espacial do sistema sensor. Assim, deverão ser ajustadas as culturas de interesse aos sistemas sensores utilizados. A compatibilização é mais apropriada entre culturas que ocupem talhões com dimensões de pelo menos alguns hectares, a fim de que a incidência de pixels mistos ou de bordas seja minimizada. Esse é um ponto fundamental para a aplicação de sensoriamento remoto à avaliação de áreas agrícolas.

Um ponto que deve ser considerado na caracterização dos pixels amostrais é que o uso de imagens multitemporais tanto intra como inter-anuais deve ser explorado ao máximo. Isso é particularmente importante, ao se considerar que numa certa região, ao menos num intervalo de tempo de alguns anos, há uma manutenção dos cultivos (Ortiz et al., 1997). Por exemplo, uma região sojicultora tende a se manter como tal ao longo de vários anos. Isso facilita e dá maior segurança na interpretação de certos pontos amostrais. Assim, para a agricultura, não se deve menosprezar a disponibilidade de dados multitemporais.

Assim, o objetivo deste trabalho é fazer uma avaliação do sistema amostral proposto e baseado na própria conformação da imagem de satélite para um município do Estado de São Paulo, a partir da simulação com talhões de forma e distribuição espacial como aparecem no campo.

## 2 Material e Métodos

A área de estudo localiza-se no norte do Estado de São Paulo e refere-se ao município de Ipuã (coordenadas W 47° S 22°). Esse município, que fica na EDR (Escritório de Desenvolvimento Regional) de Orlandia, conforme a **Figura 1**, tem uma área de 466,5 km<sup>2</sup> e é caracterizado por ser eminentemente agrícola. Há um predomínio de solos férteis (Latosolos Vermelhos), com relevo suave ondulado a plano. O clima é classificado como Cw. A agricultura é tecnificada, havendo várias áreas com pivô central. As principais culturas são a soja, o milho e a cana-de-açúcar. É comum a prática de dois ciclos agrícolas por ano, com milho e sorgo após o período normal de safra. Também é possível observar a ocorrência da rotação entre cana-de-açúcar e uma cultura anual por ocasião da reforma do canal em algumas talhões.

No presente trabalho, a ênfase concentra-se nas culturas de verão expressivas, particularmente a soja e o milho. Assumindo um universo constituído por uma certa região, por exemplo, um município, sabe-se com precisão o número de pixels que constitui aquele universo. A partir daí é possível amostrar um número de pixels a fim de estimar a área de certa classe, por exemplo soja, naquele universo. O interesse é estimar a área de uma classe C num município de área Y constituído de N pixels a partir de uma probabilidade p determinada a partir de uma amostra de n pixels extraída de N (Luiz e Epiphanyo, 2001).



**Figura 1** – Representação da localização da região de estudo.

Uma das premissas básicas permitida quase que exclusivamente pelos dados de sensoriamento remoto é que há uma regularidade na distribuição dos pixels numa imagem de satélite. Isso permite que um sistema amostral possa ser definido de tal forma que toda e qualquer unidade de imagem constituída por um pixel possa ter a mesma chance de ser amostrada. Outra característica básica de um sistema amostral baseado nas imagens de satélite é que, embora a unidade amostral seja o pixel, a base para a interpretação e conseqüente atribuição de uma classe a um dado pixel é fundada no talhão a que pertence aquele dado pixel. Para a identificação do talhão, todas as características do sensoriamento remoto que possam auxiliar na interpretação agrícola podem ser utilizadas, como por exemplo, análises multitemporais, conhecimentos prévios, informações de calendário agrícola, processamentos diversos. Para os talhões que eventualmente não puderem ser identificados com segurança, recorre-se a trabalhos de campo.

Utilizou-se uma imagem TM/Landsat-5, cuja órbita/ponto é 220/74, de 8/1/2000, que serviu como base para os trabalhos de campo e para a localização e interpretação das amostras selecionadas. Embora o formato da imagem já produzisse razoável nível de georreferenciação, a imagem foi georreferenciada com auxílio de 8 pontos de controle obtidos com GPS no campo. Para a delimitação do município sobre a imagem, utilizou-se a grade municipal do IBGE, com posterior ajuste dos limites através de cartas 1:50.000. Com auxílio de trabalhos de campo

realizados na região em abril e julho de 2000, realizou-se uma interpretação de talhões típicos de soja e de milho ocorrentes na data da passagem do satélite. É importante salientar que o objetivo, no caso específico deste trabalho, não foi o de obter uma fidelidade entre o total de área dessas duas culturas no campo e a interpretação das imagens. O objetivo desta fase foi o de derivar uma simulação com distribuição real e interpretável de talhões dessas duas culturas sobre aquela imagem de satélite específica. Desta forma, o mapa derivado da interpretação obrigatoriamente subestimou a área real dessas culturas existente no campo. Ou seja, a realidade retratada na interpretação foi apenas a daqueles talhões que apresentassem comportamento espectral típico dessas duas culturas e que estivessem em pleno vigor vegetativo naquela data da passagem do satélite. Assim, áreas de plantio mais recente, em que a resposta espectral em início de janeiro ainda não era típica de soja ou de milho não foram consideradas.

No final do processo de interpretação dispunha-se de um polígono correspondente ao limite municipal, e de diversos polígonos de soja e de milho. Com isso, foi possível derivar a área poligonal do município, e uma simulação em termos de formato e distribuição de talhões de soja, de milho, e da área que não correspondia a soja ou milho.

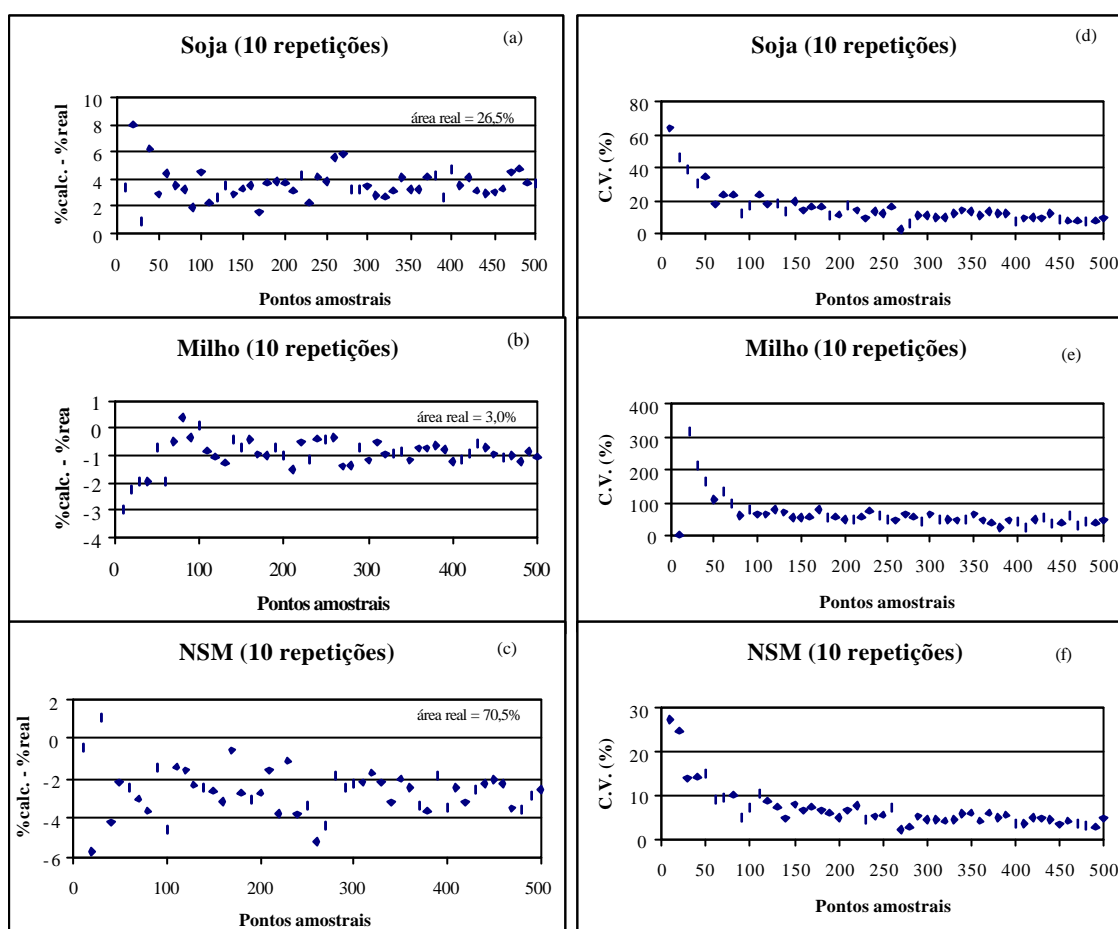
Foi selecionada uma seqüência de 500 pontos com coordenadas geográficas definidas com base em números aleatórios gerados através do SAS (SAS, 1988). Verificou-se a que classe pertencia cada ponto. O critério de pertinência foi atribuir o pixel que contivesse o ponto à classe soja ou milho sempre que este pixel tocasse o limite do polígono de uma dessas classes. Ou seja, os pontos de borda foram definidos como pertencentes às classes de milho ou soja. Esse critério conduz a uma superestimação da área de culturas, uma vez que mesmo os pontos que caíssem fora do limite do polígono das culturas, mas que estivessem em pixels que as tocassem, levava o pixel a ser considerado como uma das culturas. Uma vez conhecido o número de pontos amostrais em cada classe, foi possível derivar a proporção de cada classe no município. A área derivada pelo método de amostragem foi então comparada com a área interpretada.

A seguir, a partir da amostra total foram extraídos subconjuntos amostrais aleatórios de 10 a 500 pontos amostrais, variando de 10 em 10, com 10 repetições cada, a fim de verificar a variabilidade estatística desses esquemas amostrais. Finalmente, e como complemento da análise, fez-se uma avaliação da estimativa da área das culturas tendo por base uma amostra genérica de 500 pontos, e adotando por critério o nível de significância de 0,05.

### 3 Resultados e Discussão

O procedimento de georreferenciação da imagem com os pontos de controle forneceu um erro de 0,65 pixels, quando foram utilizados 8 pontos de controle. A poligonalização das áreas agrícolas forneceu um total de 110 talhões de soja e 12 de milho. As áreas do município ficaram divididas da seguinte forma: área municipal igual a 46.645,4 ha, sendo 12.338,4 ha de soja, 1.384,8 ha de milho e 32.922,2 ha referentes a outros usos.

A Figura 1 apresenta a diferença entre a percentagem de área calculada segundo o método amostral e a quantificada pela interpretação (referência). Estão representadas as médias (**Figura 2a-c**) e os coeficientes de variação (**Figura 2d-f**) de 10 amostras de 10 a 500 pontos amostrais, para soja, milho e área que não é nenhuma dessas duas culturas (NMS).



**Figura 2** – Média de 10 amostras para diferença entre as percentagens de área calculada e quantificada para soja (a), milho (b) e não-milho ou não-soja (c) (NMS); e coeficientes de variação para 10 amostras com número variável de pontos amostrais para soja (d), milho (e) e não-milho ou não-soja (f) (NMS).

No caso da soja (**Figura 2a**) observa-se que até amostras de aproximadamente 50 pontos amostrais a diferença entre a percentagem de área calculada e a real (para fins desse trabalho) sofre acentuadas oscilações, de até 8%. Após 50 pontos amostrais há queda nas oscilações e a diferença cai para valores ao redor de 4% e mantém-se assim, mesmo que haja aumento do número de pontos amostrais por amostra. Efeitos semelhantes são observados para o milho (**Figura 1b**), porém, com valores de diferença oscilando ao redor de -1%. A área de não-soja ou não-milho (**Figura 1c**) apresentou o mesmo comportamento, mas com valores de diferença oscilando ao redor de -3%, pois é, na verdade, o complemento da área de milho e soja. Esses resultados mostram uma superestimativa da área de soja da ordem de 3%. Isso está basicamente relacionado ao método de atribuição do ponto amostral à classe de cultivo (soja ou milho). O critério foi o de que se o pixel que contivesse o ponto amostral tocasse o polígono da classe de cultivo, então tal elemento de imagem seria classificado como aquela classe. Ou seja, todos os pixels de borda com associação amostral foram atribuídos às classes agrícolas, causando uma superestimativa para o caso da soja. No caso do milho, dado que sua área física no universo amostral era pequena, tal efeito de atribuição de pontos amostrais de borda à classe agrícola,

provocador de superestimativas, não foi suficiente para ultrapassar o efeito de sub-amostragem causado pela pouca área de milho no universo em estudo.

Quanto aos coeficientes de variação para 10 amostras (**Figura 2**), observa-se que à medida que aumenta o número de pontos amostrais nas amostras, diminui a variabilidade inter-amostras. Essa queda do coeficiente de variação é muito acentuada até amostras de aproximadamente 50 pontos amostrais. A partir daí, embora o coeficiente de variação continue em queda com o aumento dos pontos amostrais, a taxa de queda diminui. Porém, o coeficiente de variação para um mesmo número de pontos amostrais varia em função do percentual de área ocupada pelas culturas. Por exemplo, para a soja (aproximadamente 26,5% da área municipal), quando há 200 pontos amostrais, o coeficiente de variação de 10 amostras é de aproximadamente 10% (**Figura 2d**), enquanto que para o milho (3,0% da área municipal) o coeficiente de variação é de 50% (**Figura 2e**) e, para a área de não-milho ou não-soja (70,6% do município) é de 5% (**Figura 2f**).

Essa discussão anterior embute o desvio que ocorre devido ao método de atribuição do ponto amostral à classe agrícola devida (milho ou soja). Há processos que podem ser aplicados à correção desses desvios, como por exemplo, o uso de segmentação de imagens. Neste caso, os limites traçados obedeceriam aos limites dos elementos de imagem e, portanto, é evitado o viés da posição do ponto amostral.

Uma outra abordagem de análise necessária e complementar é a que leva em conta os dados da amostra para o cálculo do intervalo de confiança da estimativa da população. Neste caso, onde se tomou uma amostra de 500 pontos amostrais, obteve-se, no nível de significância de 5%, um intervalo de 1,4% a 4,0% para a área de milho, para uma média amostral de 2,60%; e um intervalo de 26,2% a 34,2% para a área de soja, para uma média amostral de 30,2%. Esses desvios podem ser considerados muito elevados para o caso do milho; neste caso para ser atingido um nível maior de precisão seriam necessárias muito mais amostras. Para o caso da soja, os desvios são, em pontos percentuais, maiores; mas sendo a área de soja bem maior que a área de milho, os erros são menos acentuados. Além disso, há que se considerar que para um nível de significância tão restrito, a variação da estimativa foi bastante satisfatória.

De um ponto de vista prático, o método aqui exposto pode ser aplicado a grandes áreas, com um nível de esforço de interpretação relativamente baixo. O uso de conhecimento de comportamento espectral de alvos agrícolas, aliado a um banco de imagens multitemporais e a um volume de dados de campo pareados com imagens assegura a interpretação de culturas específicas em um tempo relativamente curto. Eventuais pontos de dúvida podem facilmente ser sanados em campo com auxílio de GPS. Deve ser ressaltado que o uso de GPS para atribuição do ponto amostral a uma classe é muito facilitado em vista de que tal ponto amostral pertence a um talhão muito maior do que a unidade amostral. Isso é baseado num dos pressupostos básicos do sensoriamento remoto agrícola, ou seja, o de que um pixel agrícola é cercado por muitos pixels da mesma natureza.

#### **4 Conclusões**

O método apresentado para estimativa da área agrícola de soja, a partir de uma simulação tomando situações reais de distribuição e formato de talhões, mostrou-se muito eficiente. Para a área de milho, uma vez que sua proporção no universo amostral era baixa, o erro de estimativa ficou proporcionalmente elevado. O método tem a vantagem de usar o próprio universo amostral regular e padronizado fornecido pela imagem de satélite. Os erros são minimizados em função do tamanho da amostra e há indicativos de que o custo amostral pode ser otimizado através de melhorias no procedimento metodológico. Uma das vantagens desse método em relação aos

tradicionais é que há um critério estatístico mensurável para os desvios. Alguns erros observados têm causa conhecida e podem ser minimizados através de técnicas de processamento dos dados já implementadas em sistemas de processamento de imagem.

O método aqui apresentado ainda precisa ser avaliado para outras condições, como por exemplo, para regiões de diferentes características agrícolas, culturas em proporções variadas, situações de variabilidade de extensão. Não obstante suas limitações apontadas, são grandes as perspectivas de utilizar os dados de sensoriamento remoto de uma maneira mais intensa para estimativa de áreas agrícolas das grandes culturas.

### **Agradecimentos**

À Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo) pelo financiamento do projeto de número 99/09485-8, do qual se originou este trabalho.

### **Referências**

- FAO (Food and Agriculture Organization of the United Nations). **Multiple frame agricultural surveys, v.1**. FAO, Rome, 1996. 119p.
- FAO (Food and Agriculture Organization of the United Nations). **Multiple frame agricultural surveys, v.2**. FAO, Rome, 1998. 242p.
- Ippoliti-Ramilo, G.A.; Epiphanyo, J.C.N.; Shimabukuro, Y.E.; Formaggio, A.R. Sensoriamento remoto orbital como meio auxiliar na previsão de safras. **Agricultura em São Paulo**, **46**(1):89-101, 1999.
- Luiz, A.J.B.; Epiphanyo, J.C.N. **Amostragem por pontos em imagens de sensoriamento remoto para estimativa de área plantada por município**. X Simpósio Brasileiro de Sensoriamento Remoto. INPE, Foz do Iguaçu, 21-26/4/2000. 8p.
- Ortiz, M.J.; Formaggio, A.R.; Epiphanyo, A.R. Classification of croplands through integration of remote sensing, GIS and historical database. **International Journal of Remote Sensing**, **18**(1):95-105, 1997.
- Pino, F.A. Estatísticas agrícolas para o século XXI. **Agricultura em São Paulo**, **46**(2):71-105, 1999.
- SAS INSTITUTE. **SAS procedures guide**, release 6.03 edition. Cary, NC, 1988. 441p.