

## Aplicação de dados polarimétricos de radar aerotransportado (banda P) para estimativa de parâmetros estruturais de florestas tropicais.

Fábio Furlan Gama  
João Roberto dos Santos  
Corina da Costa Freitas  
José Cláudio Mura

Instituto Nacional de Pesquisas Espaciais – INPE  
Av. dos Astronautas, 1758 CEP:12.227-010 São José dos Campos – SP- Brasil  
{fabio, corina, mura}@dpi.inpe.br  
jroberto@ltd.inpe.br

**Abstract.** The main goal of this paper was look for the relationship between the forest dendrometric parameters and the radiometric image response obtained by a SAR sensor (Synthetic Aperture Radar) multipolarized, using multivariable regression techniques. The radar images were obtained during a surveying campaign in the Tapajos National Forest area (Pará state), with 1.7m resolution. The final results pointed out that the biomass variable and volume were better explained by the regression with the images in the HH polarization ( $\sigma_{HH}^{\circ}$ ), meanwhile the DBH presented better relationship with the combination of VV polarization and BMI index and the logarithm of biomass was better explained by the CSI index. The probable reason for this behaviour would be the fact that the vegetation under study had a complex structural arrangement, with great variability of individual density, and presence of many big horizontal branches, as well as distinct floristic composition.

**Palavras-chave:** SAR, P band, polarimetry, tropical forest, multiple regression, banda P, polarimetria, floresta tropical, regressão múltipla.

### 1. INTRODUÇÃO

A utilização de sistemas sensores por microondas tem-se mostrado uma ferramenta para estudo e caracterização dos mais variados tipos de alvos sobretudo das regiões tropicais, principalmente por apresentar independência das condições de iluminação solar e por proporcionar a obtenção de imagens sob as mais variadas condições meteorológicas, sofrendo pouca influencia da presença de nuvens ou neblina.

A utilização de radar para estimar determinados parâmetros biofísicos da cobertura florestal e, por conseguinte, obter um maior conhecimento da estrutura fisionômica da vegetação tem aumentado significativamente nos últimos anos, principalmente devido à interação do sinal de microondas do sensor radar com a vegetação em determinadas polarizações e faixas de espectrais, bem como aos avanços tecnológicos agora oferecidos, como a polarimetria, a interferometria e uma melhor qualidade/estabilidade radiométrica. Os estudos comparativos das bandas P, L e C, realizadas pelo JPL/NASA, durante a campanha Maestro-1, em 1994, em regiões de floresta de *pinus* (floresta de Landes -França), indicaram que a banda P foi a que apresentou uma melhor correlação em relação à sensibilidade para biomassa de florestas, de frequência de operação mais baixa (Beaudoin et al., 1994).

O INPE em parceria com o Exército Brasileiro (DSG-Brasília) e a empresa alemã Aerosensing Radarsysteme, realizou em outubro de 2000, uma missão de mapeamento na região Amazônica empregando um radar imageador interferométrico nas faixas de frequência P e X, obtendo uma resolução no terreno de 1,7 m, nas quatro polarizações da banda P, com a finalidade de avaliar a capacidade do sistema em detectar as características biofísicas dos diferentes alvos, como áreas de regeneração, florestas primárias, umidade do solo, solo nu, pastagens etc, bem como a capacidade de geração de cartas topográficas (Gama et. al., 2004). O

objetivo deste trabalho foi desenvolver um modelo de regressão multivariada para estimar a relação entre as variáveis biofísicas de biomassa, volume e DAP com os dados obtidos das imagens geradas pelo radar imageador empregado de alta penetração na vegetação.

## 2. ÁREA DE ESTUDO

A área de estudo se localiza na região da Floresta Nacional do Tapajós no estado do Pará, entre o rio Tapajós e a rodovia BR-167 (Cuiabá-Santarém), incluindo as vilas São Jorge e Tauari. A floresta dominante na região é classificada, segundo o projeto RADAM-Brasil (1976), como uma floresta Ombrófila Densa de terras baixas, verificando-se a existência de áreas de sucessão secundária em diferentes estágios, decorrente da prática da agricultura e pecuária; a **Figura 1** apresenta a área de estudo com os limites do mapeamento.

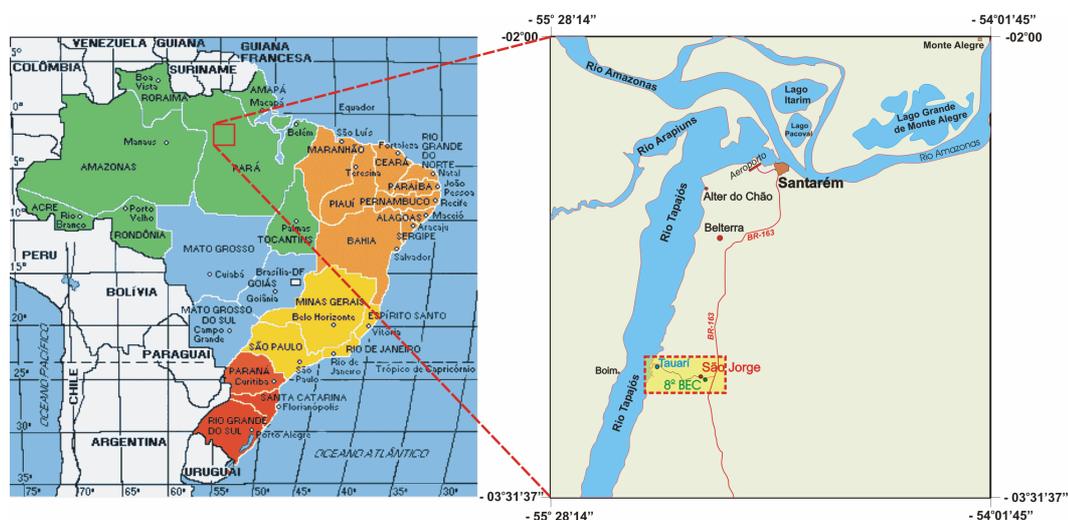


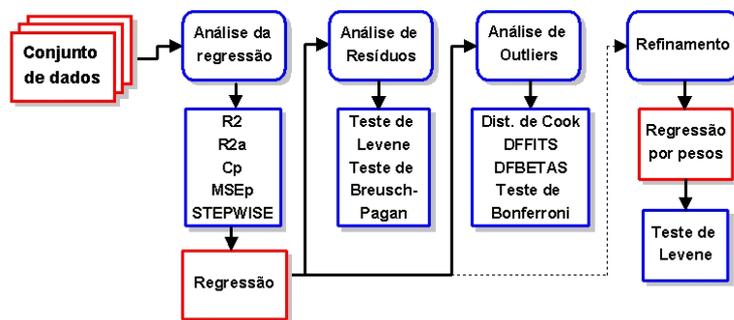
Figura 1. Área de estudo.

## 3. MATERIAL E MÉTODOS

Na região de estudo foram realizados inventários florestais em *transectos*, com medidas da altura das árvores e o seu diâmetro à altura do peito (DAP), em indivíduos cujo diâmetro fosse superior a 5,0 cm de diâmetro, além da identificação fisionômica. A biomassa da vegetação foi obtida através de equações alométricas, empregando as medidas do inventário, para suprir os modelos de estimativa (Santos et. al., 2003).

Além das respostas radiométricas polarimétricas calibradas, foram empregados índices desenvolvidos por Pope et al. (1994) derivados destes dados polarimétricos para realizar análises de ecossistemas agrícolas e florestais. Estes índices se relacionam basicamente com diferentes características da vegetação, o índice CSI corresponde à estrutura do dossel ( $CSI = (VV/VV+HH)$ ); o BMI corresponde à biomassa ( $BMI = (VV+HH)/2$ ), e o VSI, ao espalhamento volumétrico ( $VSI = (HV+VH)/potência\ total$ ); em que as respostas VV, HH, HV e VH correspondem à leitura de pixels de regiões da imagem por radar no modo linear.

O estudo das variáveis biofísicas ficou restrito à biomassa, ao volume e ao DAP, onde técnicas de regressão multivariada foram utilizadas para relacionar estas variáveis com os dados de radar e os seus índices derivados (BMI, CSI e VSI). O diagrama em blocos da **Figura 2** apresenta as etapas executadas, em que foram realizados vários testes na fase de “análise de regressão” para eliminação de variáveis, “análise dos resíduos” da regressão obtida bem como análise de *outliers*. Para as regressões que apresente uma não constância nos seus resíduos, emprega-se o refinamento da regressão por pesos para minimizar este efeito.



**Figura 2.** Diagrama das etapas de trabalho desenvolvidas.

Para a determinação do conjunto das variáveis que poderiam contribuir para a regressão, foram utilizados critérios de escolha de variáveis, como o índice de determinação ( $R^2$ ), o índice de determinação ajustado ( $R^2_a$ ), o critério  $C_p$ ,  $MSE_p$  (*mean square error*) e o critério F do método *Stepwise*. Segundo Neter et. al. (1996) um valor elevado de índice  $R^2$  não significa necessariamente que um modelo de regressão foi bem ajustado, pois depende do grau de independência entre as amostras utilizadas na regressão e do número de variáveis empregado. O índice  $R^2_a$  emprega uma ponderação para a medida de ajustamento da regressão sendo mais significativa que o índice  $R^2$ . O critério  $MSE_p$  verifica a redução na soma de quadrados dos erros quando uma ou mais variáveis preditoras são adicionadas no modelo de regressão, dado que outras variáveis preditoras já estão incluídas no mesmo, desta forma um valor baixo de  $MSE_p$  indica os melhores modelos, eliminando certas variáveis  $X$  do modelo final de regressão.

O critério  $C_p$  envolve o conceito do erro quadrático médio total de cada subconjunto de modelos de regressão ajustado, o qual considera o erro total em cada valor ajustado. Com o critério  $C_p$  pode-se identificar subconjuntos de variáveis  $X$ , nos quais o erro quadrático médio total é pequeno, ou seja, quando o valor de  $C_p$  for igual ou próximo ao número de parâmetros  $p$  e apresentar o menor valor corresponderá ao modelo com menor viés.

O procedimento *forward stepwise* consiste em iniciar a análise computacional somente com uma constante, sem nenhuma das variáveis de interesse. A cada passo, após acrescentar uma nova variável, retira-se do modelo aquela cuja contribuição parcial não foi considerada suficientemente significativa. Variáveis contidas no modelo, em um determinado passo, não necessariamente permanecerão até o final do processo. São definidos os limites de significância para entrada e permanência de variáveis no modelo de regressão.

Para a avaliação heterocedasticidade (constância da variância dos resíduos) foi empregado o teste de Levene, que realiza a comparação do teste  $t$  de dois sub-grupos de amostras do conjunto de dados, para determinar se a média dos desvios absolutos de um sub-grupo difere do outro. Outro método de avaliação da heterocedasticidade é o teste de Breusch-Pagan, que supõe que os erros são independentes e normalmente distribuídos. Este teste avalia a relação da soma dos quadrados dos erros quadráticos da regressão com a soma dos quadrados dos resíduos, caso seja menor ou igual ao teste Quiquadrado para um dado nível de significância e grau de liberdade igual a 1, a variância dos resíduos poderá ser considerada constante.

Uma forma de melhorar os modelos de regressão cujos resíduos decorrentes apresentam variâncias não constantes é realizar um refinamento da regressão através da aplicação de pesos, de forma a reduzir este efeito. A matriz de pesos é gerada a partir de uma regressão dos resíduos absolutos em relação a variável  $X$ , cujo inverso de seu valor ao quadrado para cada caso corresponderá aos pesos a serem aplicados.

Para a avaliação da existência de *outliers* no conjunto de dados, empregou-se o método de distância de Cook, DFFITS, DFBETAS e o teste de Bonferroni. O critério DFFITS analisa a influência do caso de *outlier*, através da diferença de ajuste do  $Y$  estimado para um  $ith$  caso quando todos os demais casos são empregados, e do  $Y$  estimado quando  $ith$  caso for omitido. O

teste da Distância de Cook considera a influência de um  $i_{th}$  caso em todos os  $n$  valores ajustados. Esta influência é medida pelo percentil de uma distribuição F, cujo limite aceitável de 20% para que um caso seja considerado um *outlier*. O critério DFBETAS avalia a ocorrência de *outliers* pela medida da influência do  $i_{th}$  caso pela diferença de cada coeficientes de regressão  $b_k$ , empregando todos os casos e quando o  $i_{th}$  caso é omitido. O Teste de Bonferroni pode identificar *outliers* em observações de Y, nos quais os resíduos padronizados excluídos são grandes e com grande influência no modelo, no qual os maiores resíduos padronizados excluídos seriam um *outlier*.

## 4. RESULTADOS

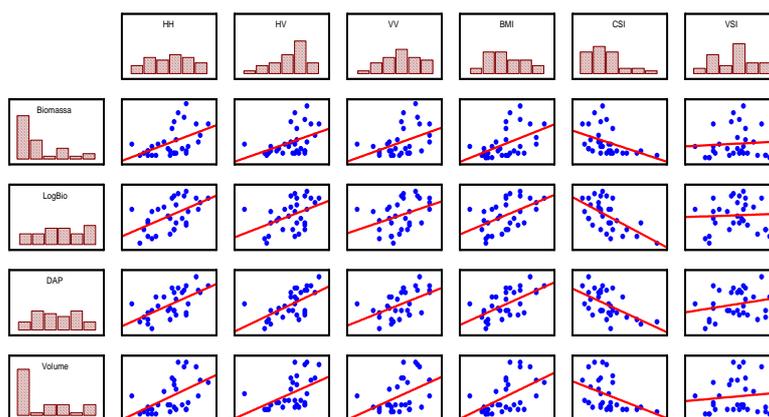
### 4.1. SELEÇÃO DE VARIÁVEIS

Analisando a matriz de correlação do conjunto de dados (**Tabela 1**) obtida pelo aplicativo Statistica 6.0, observou-se que a variável HH apresenta uma correlação acima de 50% (0,5) com todas as variáveis do inventário florestal (Biomassa, LogBio, DAP, Volume), sendo que, a variável HV apresentou uma correlação de 50% com o DAP, e a variável VV não apresentou correlações acima de 50%. Observa-se também na **Tabela 1**, que a variável VSI apresenta correlação praticamente nula em relação às variáveis do inventário florestal, a variável CSI apresentou uma correlação negativa acima de 50 % para as variáveis Biomassa, LogBio, DAP, e por sua vez a variável BMI apresentou uma correlação acima de 50% (0,5) com todas as variáveis do inventário florestal.

Variáveis	HH	HV	VV	BMI	CSI	VSI
Biomassa	0,550	0,450	0,371	0,517	-0,548	0,022
LogBio	0,550	0,450	0,371	0,517	-0,548	0,022
DAP	0,631	0,530	0,401	0,585	-0,616	0,090
Volume	0,534	0,485	0,451	0,540	-0,454	0,066

**Tabela 1.** Matriz de Correlação entre as variáveis.

Os resultados da matriz de correlação podem ser verificados no gráfico de espalhamento da **Figura 3**, em que a reta assinalada em vermelho representa o ajustamento da função de regressão linear com as respectivas variáveis.



**Figura 3.** Gráfico de espalhamento.

Analisando os gráficos da **Figura 3**, verifica-se que a variável VSI apresenta correlação praticamente nula em relação às variáveis do inventário florestal, por apresentar um forte espalhamento dos pontos indicando uma fraca contribuição ao modelo de regressão; por sua vez, as demais variáveis apresentam uma maior correlação indicando uma possível contribuição ao modelo de regressão. Não foram realizados estudos entre as variáveis do inventário florestal,



com maior identidade com o número de parâmetros, ocorreu com a variável HH. A análise do comportamento de  $R^2_a$  com os parâmetros ( $p$ ), mostrou que os valores iniciaram elevados para a variável HH, decaindo o valor de  $R^2_a$  com a inserção de mais variáveis ao modelo. A análise pelo critério  $MSE_p$  apresenta as variáveis HH-VV e VV-BMI como aquelas que apresentaram os menores valores, indicando que estas variáveis poderiam ser utilizadas no modelo de regressão.

Para a obtenção das funções de regressão empregou-se o método Stepwise, disponível no programa Statistica v. 6.0, que para a estimativa da Biomassa verificou-se que a variável HH foi a que melhor se ajustou, confirmando alguns dos testes anteriormente executados, indicando ser a variável independente que melhor explica a variável dependente (Biomassa). Para a variável LogBio verificou-se que a variável CSI foi a que apresentou melhor ajustamento, confirmando os testes anteriores executados. Para o DAP verificou-se que as variáveis que melhor se ajustaram foram VV e BMI, confirmando os testes anteriormente executados. A variável Volume apresentou um melhor ajustamento com HH. Os resultados das regressões obtidas e os respectivos índices de determinação ajustados, para cada variável do inventário florestal seguem na **Tabela 4**.

Variáveis	Regressão	$R^2_a$ (%)
Biomassa	233,64 + 24,51 HH	22,53
LogBio	8,56 – 15,95 CSI	34,68
DAP	–48,91 – 2,85 VV + 78,77 BMI	42,34
Volume	685,80 + 72,34 HH	25,94

**Tabela 4.** Resultados da regressão entre os parâmetros estruturais e os dados de radar.

Para a determinação do limite de seleção das variáveis, empregou-se um valor de  $F$  de entrada das variáveis no modelo de 4,2 e de 4,1 para a saída, para determinação deste valor foi considerado um  $\alpha$  igual a 5%,  $n$  igual a 31, e um grau de liberdade.

## 4.2. RESÍDUOS DA REGRESSÃO

Verificando a normalidade dos resíduos da regressão, foi constatado que os resíduos se alternam sobre a reta de indicação de normalidade, em que nos seus extremos há uma tendência a não normalidade, não permitindo afirmar que os resíduos apresentaram uma distribuição normal, uma vez que o histograma apresentava uma “cauda” na distribuição à direita do seu gráfico.

Foram realizados testes de Levene e Breusch–Pagan para os quatro modelos, para avaliação da heterocedasticidade (**Tabela 5**), em que verificou-se que o único modelo que apresentou constância em seus resíduos foi o DAP, os demais não apresentaram constância do seus resíduos e em alguns casos, apresentando uma assimetria em torno do zero, com a forma de “megafone” que segundo Neter et. al. (1996) necessita de refinamento do modelo de regressão através de pesos, para amenizar o efeito e melhorar a regressão.

Variáveis	Levene ( $t_L^*$ )	$t_{0,975;29}$	Resíduo	Breusch – Pagan ( $\chi^2_{BP}$ )	$X^2(0,95;1)$	Resíduo
Biomassa	2,63	2,04	-	3,20	3,84	Cte
LogBio	3,06		-	15,43		-
DAP	1,96 (VV); 0,25 (BMI)		Cte	0,39 (VV); 0,40 (BMI)		Cte
Volume	4,10		-	14,98		-

**Tabela 5.** Resultados dos testes de heterocedasticidade.

## 4.3. ANÁLISE DE OUTLIERS:

As análises iniciais dos dados indicaram a possibilidade de existência de *outliers*, uma vez que alguns pontos nos gráficos de espalhamento se mostraram mais afastados da nuvem de pontos obtidos, necessitando assim de testes objetivos da existência desses *outliers*.

Testes da Distância de Cook, DFFITS, DFBETAS e o Teste de Bonferroni foram realizados empregando o programa Statistica v. 6.0, que para a regressão da Biomassa, verificou-se que a terceira amostra (caso 3) das 31 realizadas no inventário, foi o caso que mais se destacou do conjunto, cuja distância de Cook ( $D_i$ ) correspondeu a um percentil de 17,21 % de uma distribuição  $F$ , para 2 parâmetros, e 31 casos, indicando assim uma baixa influência no modelo (limite aceitável de 20%) não sendo considerado um *outlier*. A análise de *outliers* pelo critério DFFITS para o caso 3, indicou não se tratar de um *outlier* pois o valor de  $DFFITS_3$  obtido foi de 0,0028 sendo muito menor que a referência (1,0). Analisando os valores obtidos do critério de DFBETAS para os diferentes  $b_k$  ( $b_0$ : 0,54;  $b_1$ : 0,43), com a inserção e eliminação do caso 3, os mesmos não ultrapassaram o limite de referência 1,0, indicando desta forma que o caso 3 não se tratava de um *outlier*. O teste de Teste de Bonferroni, para verificação de *outlying* em  $Y$ , apresentou valores muito inferiores ao teste  $t$  (3,21 para:  $\alpha = 0,1$ ;  $n = 31$  e  $p=2$ ), conclui-se por este teste que o conjunto de dados não contem *outliers*.

Para o modelo de regressão do DAP, verificou-se que para o caso 17, a distância de Cook ( $D_i$ ) correspondeu a um percentil de 7,63% de uma distribuição  $F$ , para 3 parâmetros, e 29 casos, não sendo considerado um *outlier* (limite de 20%). A análise de *outliers* pelo critério DFFITS para o caso 17, indicou não se tratar de um *outlier*, pois o valor obtido de  $DFFITS_{17}$  foi de 0,249 sendo muito menor que a referência (1,0). Analisando os valores obtidos do critério de DFBETAS para os diferentes  $b_k$  ( $b_0$ : 0,043;  $b_1$ : 0,042;  $b_2$ : -0,039), com a inserção e eliminação do caso 17, os mesmos não ultrapassaram o limite de referência 1,0, indicando desta forma que o caso 17 não se tratava de um *outlier*. O teste de Teste de Bonferroni, para verificação de *outlying* em  $Y$ , apresentou valores muito inferiores ao teste  $t$  (2,90 para:  $\alpha = 0,1$ ;  $n = 31$  e  $p = 6$ ), conclui-se por este teste que o conjunto de dados não contem *outliers*.

Para o modelo de regressão do LogBio, verificou-se que para o caso 16, a distância de Cook ( $D_i$ ) correspondeu a um percentil de 17,22% de uma distribuição  $F$ , para 2 parâmetros, e 29 casos, indicou não se tratar de um *outlier* (limite de 20%). A análise de *outliers* pelo critério DFFITS para o caso 16, indicou se tratar de um *outlier*, pois, o valor de  $DFFITS_{16}$  obtido foi de 1,28 sendo maior que a referência (1,0). Analisando os valores obtidos do critério de DFBETAS para os diferentes  $b_k$  ( $b_0$ : -0,07673;  $b_1$ : 0,07843), com a inserção e eliminação do caso 16, os mesmos não ultrapassaram o limite de referência 1,0, indicando desta forma que o caso 16 não se tratava de um *outlier*. O teste de Teste de Bonferroni apresentou valores muito inferiores ao teste  $t$  (3,21 para:  $\alpha = 0,1$ ;  $n = 31$  e  $p = 2$ ), conclui-se por este teste que o conjunto de dados não contem *outliers*.

Para o modelo de regressão do Volume, verificou-se que para o caso 17, a distância de Cook ( $D_i$ ) correspondeu a um percentil de 17,22% de uma distribuição  $F$ , para 2 parâmetros, e 29 casos, indicou não se tratar de um *outlier* (limite de 20%). Analisando os valores obtidos do critério de DFBETAS para os diferentes  $b_k$  ( $b_0$ : - 0,4840;  $b_1$ : - 0,5618), com a inserção e eliminação do caso 17, os mesmos não ultrapassaram o limite de referência 1,0, indicando desta forma que o caso 17 não se tratava de um *outlier*. O teste de Teste de Bonferroni apresentou valores muito inferiores ao teste  $t$  (3,21 para:  $\alpha = 0,1$ ;  $n = 31$  e  $p = 2$ ), conclui-se por este teste que o conjunto de dados não contem *outliers*.

#### 4.4. REFINAMENTO DO MODELO

Para o caso particular do estudo da biomassa, a matriz de pesos foi calculada utilizando a planilha Excel e gerou-se a regressão por pesos no software Minitab-13. O processo levou a uma melhoria no desvio padrão um decréscimo nos coeficientes de determinação na segunda interação do método, cujo modelo de regressão obtido segue abaixo:

$$\text{Biomassa} = 195 + 18,8 \text{ HH} \quad (R^2_a = 20,2\%) \quad (\text{Desvio Padrão} = 6,43)$$

Aplicando o Teste de Levene para a nova regressão obtida, verificou-se uma melhoria no valor de  $t_L^*$  (2,44), mas ainda superior ao valor de distribuição  $t_{0,975;29}$  (2,04), indicando que o refinamento não conseguiu tornar os resíduos heterocedasticidade.

## 5. CONCLUSÕES E RECOMENDAÇÕES

Analisando os quatro modelos obtidos, verificou-se que o modelo da Biomassa foi o que apresentou maior variância nos seus resíduos, tornando necessária a utilização de regressão por pesos para diminuir este efeito. Isto permitiu a melhoria do desempenho do modelo de biomassa em relação a heterocedasticidade, mas resultando em um ajustamento mais fraco do modelo devido ao decaimento no índice  $R_a^2$ . Isto ocorreu devido à regressão dos resíduos pelas variáveis X, na técnica de geração dos pesos ter apresentado um forte espalhamento.

Por sua vez, o modelo realizado com a variável DAP apresentou os melhores resultados com relação aos resíduos, por serem simétricos e uniformes, em que os testes de *outlier* não acusaram presença de casos significativos, resultando em um modelo com melhor ajustamento que a Biomassa, observando o  $R_a^2$  e a variância dos resíduos, não necessitando empregar a regressão por pesos. O modelo do Volume não apresentou heterocedasticidade, mas por sua vez apresentou o índice  $R_a^2$  superior ao modelo da Biomassa, não apresentando casos de *outliers*.

O modelo do LogBio apresentou a suspeita de presença de *outlier* pelo teste DFFITS, que não foi confirmado pelos demais testes. Este modelo não apresentou heterocedasticidade, mas o seu índice  $R_a^2$  foi superior ao do modelo da Biomassa e do Volume, mas não superando o índice do modelo do DAP. Os modelos de regressão do LogBio e Volume são sofreram o refinamento da regressão devido à variância dos resíduos não terem sido tão grandes quanto ao da biomassa.

Como continuidade desta linha de pesquisa será realizada uma nova missão de mapeamento empregando radar aerotransportado em áreas de reflorestamento de *Eucalyptus*, o que permitirá obter um melhor entendimento da interação do sinal de microondas com a vegetação, estabelecendo com maior facilidade os parâmetros estruturais da floresta que possuam maior correlação com a resposta do radar, com o objetivo de buscar um modelo de relacionamento entre as características biométricas de áreas reflorestadas com a resposta do sensor radar.

## Referências

- Beaudoin, A. *et al.* Retrieval of Forest Biomass from SAR Data. **International Journal of Remote Sensing**, v.15, n.14, p.2777-2796, 1994.
- Gama F.F., Santos J.R., Mura J.C., Freitas C.C. Estimativa de parâmetros biofísicos da cobertura florestal através de dados polarimétricos de radar (banda P) aerotransportado. In: Disperati, A.A., Santos, J.R. (ed.) Aplicações de geotecnologias na Engenharia Florestal, Curitiba: Copiadora Gabardo Ltda. 2004 **Anais...** Curitiba. 2004. p. 270-272. ISBN 85-904724-1-8.
- Neter, J., Kutner M.H., Nachtsheim, C.J., Wasserman, W. **Applied Linear Statistical Models**, Boston, MA, McGraw-Hill, 1996. 1408 p.
- Pope K.O., Rey-Benayas, J.M., Paris, J.F., Radar Remote Sensing of forest and wetland ecosystems in the Central American Tropics. **Remote Sensing of Environment**, v.2, n. 48, p. 205-219, 1994.
- Projeto RADAMBRASIL (1976). Folha SA. 21 Santarém; geologia, geomorfologia, solos, vegetação e uso potencial da terra. Rio de Janeiro, 522pp. (Levantamento de recursos naturais, 10)
- Santos J.R., Freitas C.C., Araujo L.S., Dutra L.V., Mura J.C., Gama F.F., Soler L.S., Sant'Anna S.J.S. Airborne P-band SAR applied to the above ground biomass studies in the Brazilian tropical rainforest. **Remote Sensing of Environment**. v.87, n.4, p.482-493, 2003.