



## MODELAGEM DO BALANÇO DE ENERGIA POR REGRESSÃO ROBUSTA COM VALIDAÇÃO CRUZADA E BOOTSTRAP

Renata G. Aguiar<sup>1</sup>, Carlo R. De Muisis<sup>2</sup>

1 Matemática, doutoranda do PPG em Física Ambiental da UFMT, Profa. Assistente, Departamento de Engenharia Ambiental, UNIR, Ji-Paraná – RO, Fone: (0xx69) 3421 3595, rgaguiar@gmail.com

2 Engenheiro Civil, professor orientador do PPG em Física Ambiental da UFMT, Profa. Permanente, UNIC, Cuiabá – MT, Fone: (0xx65) 3363 1264, carlo.ralph@kroton.com.br

Apresentado no XVIII Congresso Brasileiro de Agrometeorologia – 02 a 06 de Setembro de 2013 – Centro de Convenções e Eventos Benedito Silva Nunes, Universidade Federal do Pará, Belém, PA

**RESUMO:** O objetivo do presente estudo foi verificar a eficiência do fechamento do balanço de energia em ecossistema amazônico por meio do modelo de regressão linear simples e robusta com *bootstrap* e validação cruzada no intuito de criar subsídios para elucidar a contribuição da floresta no complexo sistema absorção-emissão de dióxido de carbono. As medidas de saldo de radiação, fluxo de calor latente, fluxo de calor sensível e fluxo de calor no solo foram realizadas de janeiro a dezembro de 2008 em uma área de floresta tropical úmida no sudoeste da Amazônia, em um sítio experimental pertencente à rede de torres do Programa LBA. Os resultados evidenciaram que a regressão linear simples não é um modelo adequado para analisar o fechamento do balanço de energia e que a aplicação do modelo de regressão linear robusta com *bootstrap* melhorou o ajuste, embora tenha aumentado o resíduo do fechamento.

**PALAVRAS-CHAVE:** regressão linear, *bootstrap*, validação cruzada.

### ENERGY BALANCE MODELING FOR ROBUST REGRESSION WITH CROSS-VALIDATION AND BOOTSTRAP

**ABSTRACT:** The aim of this study was to verify the efficiency of the the energy balance closure in the Amazon ecosystem through linear regression model and robust with bootstrap and cross-validation in order to create subsidies to elucidate the contribution of the forest complex system in absorption-emission of carbon dioxide. Measurements of net radiation, latent heat flux, sensible heat flux and soil heat flux were made from January to December 2008 in an area of tropical rain forest in southwest Amazonia, in an experimental site belongs to the network towers of the LBA Program. The results showed that the simple linear regression is not an appropriate model to analyze the energy balance closure and the application of linear robust regression model with bootstrap improved the fit, although it increased the residue of closing.

**KEYWORDS:** linear regression, bootstrap, cross-validation.





## INTRODUÇÃO

O aporte das florestas no processo de absorção do dióxido de carbono excedente na atmosfera devido às ações antrópicas continua sendo um enigma para a comunidade científica. Dentre os fatores que recomendam atenção ao dar ênfase às pesquisas que concluíram ser a floresta um sumidouro relevante desse gás, está o método de covariância de vórtices turbulentos (*eddy covariance*).

Esse método apresenta limitações, principalmente ao ser usado em áreas com vegetação alta, uma vez que foi desenvolvido para vegetações baixas e terrenos horizontalmente uniformes (BALDOCCHI et al. 1988; BALDOCCHI, 2003). No intento de verificar a eficiência do método exposto, diversos estudos utilizaram um modelo de regressão linear simples nas componentes do balanço de energia, uma vez que o fluxo de calor sensível e latente – medido pelo mesmo sensor que afere o fluxo de dióxido de carbono – e as medidas do saldo de radiação, do fluxo de calor no solo, da energia armazenada na biomassa e no dossel são realizadas por sensores diferentes, o fechamento do balanço indicaria concordância entre as medidas. No entanto, esses estudos não verificaram a adequação dos pressupostos que demandam uma análise de regressão, o que leva a conclusões errôneas quando o modelo não é satisfatório.

Portanto, o presente estudo teve a finalidade de verificar a eficiência do fechamento do balanço de energia em ecossistema amazônico por meio do modelo de regressão linear robusta com *bootstrap* no intuito de criar subsídios para elucidar a contribuição da floresta no complexo sistema absorção-emissão de dióxido de carbono.

## MATERIAL E MÉTODOS

No período de janeiro a dezembro de 2008 foram realizadas medidas de saldo de radiação, fluxo de calor latente, fluxo de calor sensível e fluxo de calor no solo em uma área de floresta tropical úmida no sudoeste da Amazônia, em um sítio experimental pertencente à rede de torres do Experimento de Grande Escala da Biosfera-Atmosfera na Amazônia – Programa LBA. A área de floresta tropical está localizada na Reserva Biológica do Jaru (REBIO Jaru), 10°11'11,4''S; 61°52'29,9''W, em Rondônia.

Medidas dos fluxos de calor latente e sensível foram determinadas pelo método de covariância de vórtices turbulentos com o emprego de um sistema de medidas de alta frequência dos fluxos de superfície, composto por um anemômetro sônico tridimensional (Solent 1012R2, Gill Instruments, UK), – que mede as três componentes da velocidade do vento - velocidade horizontal do vento a leste ( $u$ ), velocidade horizontal do vento a norte ( $v$ ) e a velocidade vertical do vento ( $w$ ) – e a temperatura do ar, e um analisador de gás por infravermelho de caminho aberto (IRGA, Li-7500, LICOR, USA), que mede as concentrações de vapor de água e dióxido de carbono.

Os sensores que mediram os dados de saldo de radiação e do fluxo de calor no solo foram acoplados a um sistema de baixa frequência de aquisição de dados (Datalogger CR10X, Campbell Scientific Instrument, Utah, USA) que armazenava os dados em um módulo de memória a cada 10 min e eram semanalmente transferidos para um disco rígido de





computador.

Primeiramente foi realizada uma análise de regressão linear simples entre  $R_n - G$  (eixo das abscissas) e  $H + LE$  (eixo das ordenadas) para verificar o quão próximos esses dados estariam de uma reta 1:1 por meio da análise do intercepto ( $a$ ) e do coeficiente de regressão ( $b$ ) da equação gerada pelo modelo e também do coeficiente de determinação ( $r^2$ ).

Então procedeu-se uma Análise de Variância (ANOVA) para testar a significância da regressão, com nível de significância ( $\alpha$ ) de 0,05. Para testar o pressuposto de normalidade dos erros foram utilizados, considerando o menor nível de significância calculado, os testes de Anderson-Darling, Kolmogorov-Smirnov e Ryan-Joiner que é similar ao Shapiro-Wilk. Para testar a homogeneidade de variâncias foi utilizado o teste de Levene e de Wald-Wolfowitz. O pressuposto de que os resíduos são independentes uns dos outros foi avaliado por meio do teste de Durbin-Watson ( $\alpha = 0,05$ ).

Antes de proceder à aplicação da regressão linear robusta, primeiramente foram testadas e validadas as oito funções que constam no Matlab (2012a) para encontrar a constante de ajuste mais adequada para calcular o peso dos valores extremos dos dados do balanço de energia, visto que ainda não há um consenso sobre qual função estima melhor (S-PLUS, 1998).

No teste e validação das funções foi utilizada a validação cruzada denominada  $k$ -fold. Os dados foram divididos em dez subconjuntos mutuamente exclusivos e de tamanhos iguais. A proporção utilizada para o teste foi de 90% e de 10% para a validação. Para diferenciar qual função apresentou a constante de ajuste mais adequada utilizou-se o desvio mediano, que consiste nas medianas dos módulos dos desvios entre valores observados e estimados dentro de cada késimo particionamento da validação cruzada. Conforme detalhado na Equação 1.

$$dm = md(md_k(|(H + LE)_{ik} - \widehat{(H + LE)}_{ik}|)) \quad (1)$$

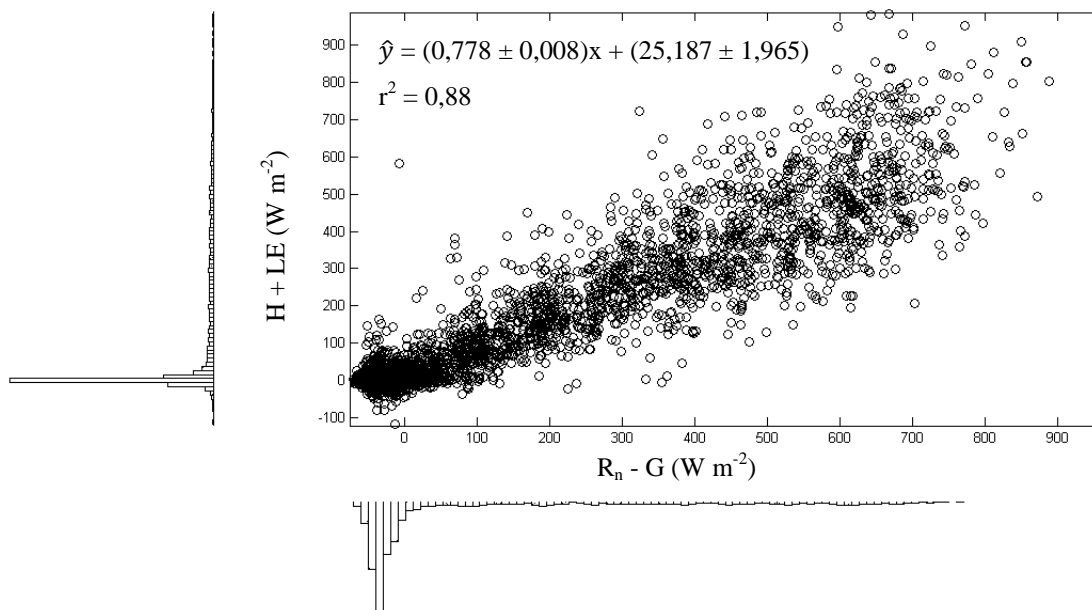
Em que:  $dm$  é o desvio mediano,  $md$  é a mediana,  $k$  é número de particionamentos da validação cruzada,  $H$  é o fluxo de calor sensível e  $LE$  é fluxo de calor latente.

Após definir qual a função e a constante de ajuste adequada o número de reamostragens foi replicado de 10 a 5000, variando de 10 em 10, com a finalidade de encontrar qual seria o número ideal de iterações para estimar os intervalos de confiança de 95% dos coeficientes do modelo de regressão linear robusta.

## RESULTADOS E DISCUSSÃO

A regressão linear simples entre a energia disponível no sistema ( $R_n - G$ ) e a soma dos fluxos turbulentos ( $H + LE$ ) gerou um coeficiente de regressão  $b = 0,778 \pm 0,008$  (todo valor após o sinal  $\pm$  corresponde a um intervalo de confiança de 95%), resultando em um resíduo de 22% (Figura 1).





**Figura 1** – Relação entre a energia disponível no sistema ( $R_n - G$ ) e a soma dos fluxos de calor sensível e calor latente ( $H + LE$ ).  $R_n$  é o saldo de radiação e  $G$  é o fluxo de calor no solo. Os pontos representam médias a cada 30 min ( $n = 5.320$ ).

Os coeficientes de regressão e de determinação foram condizentes com os valores encontrados em vários estudos que utilizaram o método *eddy covariance* (EC) em diversos tipos de superfícies. Entretanto, não basta que os coeficientes de regressão e de determinação estejam dentro de uma margem aceitável para avaliar que os dados de fluxo de energia obtidos pelo método EC apresentam qualidade satisfatória, é imprescindível verificar a adequação do modelo. Sem a qual tornaria os resultados gerados pelo modelo sem significância, portanto, comprometendo a confiança nas estimativas dos coeficientes.

Chatterjee & Hadi (2006) enfatizam que a validade dos procedimentos estatísticos depende da validade dos pressupostos. Desta forma, quando os pressupostos são violados os resultados não são confiáveis e conclusões com base neles podem levar a graves erros.

Comumente a primeira etapa para atestar a adequação do modelo de regressão linear simples é verificar os pressupostos. No entanto, optou-se, por uma questão de otimização do tempo, primeiro realizar o teste de hipóteses, pois caso a regressão não fosse validada não haveria motivo para utilizar o modelo de regressão linear simples, e conseqüentemente, de avaliar passo a passo todos os pressupostos.

A análise de variância, foi o teste estatístico de hipóteses utilizado para verificar a adequação da regressão linear simples, com  $\alpha = 0,05$ . Mediante exame da estimativa do p-valor (nível de significância observado), verificou-se que existe regressão dos fluxos de calor sensível e latente em relação à energia disponível.

Ao verificar os pressupostos, constatou-se violação de dois importantes pressupostos, o de normalidade e o de homocedasticidade, desta forma, o modelo de regressão linear simples não deve ser aplicado, pois implicaria em inferências equivocadas sobre as estimativas dos coeficientes da regressão das variáveis que integram o fechamento do balanço de energia.



Contudo, a Estatística dispõe de técnicas de ação corretiva que possibilitam transformar os dados, e em alguns casos obter normalidade e ou homocedasticidade. Foram aplicadas as transformações logarítmica, inversa, raiz quadrada e Box-Cox, no entanto, os resultados evidenciaram que as transformações não aproximaram a distribuição dos resíduos de uma normal e também não estabilizaram a variância.

Na busca por um método que apresentasse melhor ajuste aos dados do balanço de energia foi utilizada a regressão linear robusta com *bootstrap* e validação cruzada, pois quando os erros de um modelo de regressão não estão normalmente distribuídos, o cálculo dos intervalos de confiança por meio da estatística paramétrica não apresenta resultados satisfatórios, sendo possível contornar o problema utilizando *bootstrap* (CHERNICK; FRIIS, 2003).

Para encontrar a função de ponderação mais adequada às componentes do balanço de energia foram testadas e validadas por meio da validação cruzada *k-fold* oito funções de ponderação de modo a pesquisar, para cada uma delas, qual o valor da constante de ajuste (*tune*) mais adequada para ponderar os pesos da regressão robusta aplicada aos dados do balanço de energia. Os resultados estão dispostos na Tabela 1.

Para escolher a função de ponderação adequada observou-se qual apresentou o menor desvio mediano, que no caso em questão foi a função Bisquare. A constante de ajuste referente ao menor resíduo foi 1,614.

**Tabela 1** – Constantes de ajuste otimizada e respectivo desvio mediano das diferentes funções de ponderação

Função	Constante de Ajuste	<i>dm</i>
Andrews	0,450	13,590
Bisquare	1,614	13,585
Cauchy	0,560	14,641
Fair	-*	-*
Huber	-*	-*
Logistic	-*	-*
Talwar	0,990	13,664
Welsch	1,130	13,938

\* A constante de ajuste e desvio mediano não convergiu para um mínimo.

Nota: *dm* - mediana das medianas dos módulos dos desvios entre valores observados e estimados dentro de cada késimo particionamento da validação cruzada.

Antes de aplicar o *bootstrap* para estimar o intervalo de confiança de 95% dos coeficientes do modelo de regressão linear simples, o número de reamostragens foi replicado e diferentemente das 1000 replicações comumente indicada, observou-se que o coeficiente de regressão e o intercepto estabilizaram a partir de 2500 iterações *bootstrap*.

Os intervalos de confiança dos coeficientes da regressão robusta foram calculados utilizando 2500 replicações *bootstrap* e estão dispostos na Equação 2. O ajuste do modelo, medido pelo coeficiente de determinação passou a ser de 0,99. Evidenciando um ajuste muito melhor para a regressão linear robusta ao comparar com os resultados da regressão linear simples.

$$\hat{y} = 20,362 [18,251; 21,004] + 0,604x [0,546; 0,616] \quad (2)$$

$$r^2 = 0,99 [0,988; 0,992]$$

O resíduo do fechamento do balanço de energia aumentou consideravelmente de 22,2%,





considerando a regressão linear simples, para 39,6% utilizando a regressão linear robusta. Desta forma, a aplicação da regressão linear robusta com *bootstrap* e validação cruzada evidenciou que a utilização de uma técnica adequada aos dados do balanço de energia gerou uma discrepância no fechamento do balanço de energia mais expressiva. O desequilíbrio energético conforme ressaltam Wohlfahrt & Widmoser (2013) continua sendo um problema pendente em micrometeorologia e esforços contínuos por parte da comunidade científica são necessários a fim de fazer avançar esse assunto.

## CONCLUSÕES

A regressão linear simples não é um modelo adequado para analisar o fechamento do balanço de energia, pois não foi possível validar os pressupostos que o modelo demanda. A regressão linear robusta melhorou o ajuste. Entretanto, houve um aumento do resíduo do fechamento do balanço de energia, evidenciando que ainda são necessárias melhorias na técnica de EC para determinação dos fluxos turbulentos e que os resultados dessa técnica devem ser avaliados com parcimônia.

## AGRADECIMENTOS

Ao Experimento de Grande Escala da Biosfera-Atmosfera na Amazônia – Programa LBA e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.

## REFERÊNCIAS

- BALDOCCHI, D. D. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. **Global Change Biology**, v. 9, p. 479-92, 2003.
- BALDOCCHI, D. D.; HINCKS, B. B.; MEYERS, T. P. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. **Ecology**. v. 69, p. 1331-40, 1988.
- CHATTERJEE, S.; HADI, A. S. **Regression analysis by example**. 4. ed. New Jersey: John Wiley & Sons, 2006. 375p.
- CHERNICK, M. R.; FRIIS, R. H. **Introductory biostatistics for the health sciences: modern applications including bootstrap**. New Jersey: John Wiley & Sons, 2003. 406p.
- S-PLUS. **S-Plus 4 guide to Statistics**. Seattle: MathSoft. 1998. 881 p. Disponível em: <<http://www.mathsoft.com>> Acesso em: 08 ago. 2012.





XVIII Congresso Brasileiro de Agrometeorologia – XVIII CBA  
2013 e VII Reunião Latino Americana de Agrometeorologia  
Belém - PA, Brasil, 02 a 06 de Setembro 2013  
*Cenários de Mudanças Climáticas e a Sustentabilidade  
Socioambiental e do Agronegócio na Amazônia*



WOHLFAHRT, G.; WIDMOSER, P. Can an energy balance model provide additional constraints on how to close the energy imbalance? **Agricultural and Forest Meteorology**, v. 169, p. 85-91, 2013.



Secretaria do XVIII Congresso Brasileiro e VII Reunião Latino Americana de Agrometeorologia – 2013  
Rua Augusto Corrêa, 01. Cidade Universitária Prof. José da Silveira Netto  
CEP 66075-900 Guamá. Belém - PA - Brasil  
<http://www.sbagro.org.br>

