Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Mensuração da Biomassa e Seleção de Modelos para a Construção de Equações de Biomassa.

Edgar de Souza Vismara

Dissertação apresentada para obtenção do título de Mestre em Recursos Florestais. Opção em Conservação de Ecossistemas Florestais

Piracicaba
2009

2 SELEÇÃO DE MODELOS EMPÍRICOS ATRAVÉS DO CRITÉRIO DE INFORMAÇÃO DE AKAIKE

Resumo

A questão dos métodos de seleção de modelos usados na área florestal é discutida, no que diz respeito a suas limitações, e o AIC é apresentado como abordagem alternativa para seleção de modelos empíricos no meio florestal. Na década de 70 Akaike propôs seu critério de seleção de modelos baseado no principio de máxima entropia entre duas distribuições e no de máxima verossimilhança na estimativa de parâmetros. Este capítulo apresenta sucintamente as bases teóricas desse critério, bem como discute sua importância no contexto da modelagem no meio florestal.

Palavras-chave: seleção de Modelos, critério de Informação de Akaike, AIC

Abstract

The question of model selection metods used in forestry is discussed, with regard to its limitations, and the AIC is presented as alternative approach for the selection of empirical models in forest environment. In the 70's Akaike proposed the criterion of model selection based on the principle of maximum entropy between two distributions and of the maximum likelihood estimation of parameters. This paper presents briefly the theoretical basis of this criterion and discusses its importance in the context of modeling in the forestry.

Keywords: Model Selection, Akaike Information Criterion, AIC.

2.1 Introdução

O processo de análise estatística consiste basicamente em encontrar um modelo apropriado, estimar os parâmetros e, por fim, determinar a ordem ou tamanho deste modelo (BOSDOGAN, 1987).

Nesse contexto, assume-se que exista um único modelo correto (ou até mesmo verdadeiro) ou pelo menos um melhor modelo que seria suficiente como base para inferir a partir dos dados (BURHAM; ANDERSON, 2004). No entanto, de acordo com Parzen (1982), a modelagem estatística de dados é um campo da estatística que busca ajustar um modelo aos dados sem conhecimento de como o modelo verdadeiro é ou deve ser.

De fato, o que ocorre normalmente nos diversos campos de aplicação da estatística, incluindo o meio florestal, é que uma vez selecionado o modelo o problema da inferência passa a ser de

estimação e determinação da ordem do modelo, definida pelo número de variáveis preditoras. Desta maneira, no processo de inferência, a incerteza na escolha do modelo não é considerada, tornando-se apenas mero componente da variância (BURHAM; ANDERSON, 2004).

Em trabalhos recentes da literatura (SAKAMOTO; ISHIGURO; KITAGAWA, 1986; BOSDOGAN, 1987; CAMERON; WINDMEIJER, 1996; BUCKLAND; BURHAM; AUGUSTIN, 1997; BOSDOGAN, 2000; BURHAM, ANDERSON, 2002, 2004) o problema da seleção de modelos vem sendo tratado de forma distinta. Tem-se buscado um critério que avalie o melhor modelo de aproximação, entre uma série de modelos candidatos com diferentes relações funcionais e com diferentes número de parâmetros, para descrever os dados.

Esse critérios, além de incluir a incerteza na seleção dos modelos, que permite a comparação de modelos não hierárquicos, ainda levam em consideração a complexidade do modelo de acordo com o princípio da parcimônia, ou seja, dentre uma série de modelos concorrentes de mesmo desempenho, os modelos mais complexos ou com maior numero de parâmetros devem ser penalizados.

Essa abordagem iniciada por Akaike (1974, 1981) estabelece uma simples relação entre a distância de Kulbach-Leibler e a função de máxima log-verossimilhança de Fisher levando a uma simples e efetiva metodologia para seleção de modelos parcimoniosos para análise de dados empíricos, denominado de critério de informação de Akaike (AIC).

Este critério estima a distância relativa de Kulbach-Leibler entre dois modelos, possuindo a vantagem de permitir a comparação de modelos não-hierárquicos, considerando-os apenas concorrentes (SAKAMOTO; ISHIGURO; KITAGAWA, 1986), sendo portanto, uma ferramenta valiosa na seleção de modelos empíricos de predição no meio florestal.

Desta maneira, este capítulo se propõem discutir as limitações do uso dos critérios de seleção de modelos empíricos comumente usados no meio florestal e apresentar como o AIC poderá vir a suprir essas limitações, tornado-se ferramenta importante no processo de seleção e avaliação desses modelos.

2.2 Abordagem tradicional

Em trabalhos realizados no meio florestal (SOARES; OLIVEIRA, 2002; MAESTRI et. al., 2004; SOARES; LEITE; GÖRGENS, 2005) modelos empíricos de predição de volume ou biomassa arbórea são obtidos através de métodos estatísticos de regressão linear e não linear. Nesses trabalhos, a verificação da qualidade do ajuste e a consequente escolha do melhor modelo ocorrem através da abordagem clássica dos testes F do modelo e parcial, bem como através do coeficiente de determinação (R^2).

Nesses trabalhos, o teste F do modelo e o coeficiente de determinação são tratados como critérios de seleção absolutos, já que são usados para verificar se o modelo é ou não adequado para aplicação num determinado conjunto de dados. Trata-se de um equivoco pois, na verdade,

esses dois critérios comparam a qualidade do ajuste de um modelo em relação a um outro modelo mais simples que é a média amostral .

Esse equivoco fica claro a partir da observação da natureza desses critérios. O teste F do modelo, segundo (CHATERJEE; PRICE, 1991) é construído através da investigação das seguintes hipóteses:

H_0: Todos os coeficientes da regressão são iguais à zero.

$vs.$

H_a: Pelo menos um coeficiente da regressão difere de zero.

Se a hipótese nula não for rejeitada, conclui-se que não há relação entre a variável resposta e as p variáveis preditoras, ou seja, o modelo testado não difere da média amostral. No caso de se rejeitar a hipótese nula conclui-se que o modelo de regressão explica melhor a variabilidade dos dados que a média amostral.

No entanto, se forem considerados dois modelos concorrentes, em que a hipótese nula tenha sido rejeitada para ambos, o teste F do modelo é de pouca utilidade como critério de decisão de qual modelo é mais adequado (DRAPER; SMITH, 1998).

O coeficiente de determinação (R^2), por sua vez, quantifica a proporção da variabilidade da variável resposta que é explicada por um modelo de aproximação qualquer (RAO, 1973). Este é definido por

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}, \tag{2.1}$$

onde $\sum(y_i - \hat{y}_i)^2$ é a soma dos desvios em relação aos valores ajustados e $\sum(y_i - \bar{y}_i)^2$ é a soma dos desvios em relação à média.

O R^2 pode assumir valores no intervalo [0,1], sendo que valores próximos de 1 (um) denotam uma boa relação entre a variável resposta e as p variáveis preditoras, indicando um bom ajuste. Por outro lado, valores próximos ou iguais a 0 (zero) denotam pouca ou nenhuma relação entre a variável resposta e as p variáveis preditoras, indicando que o modelo não é superior a média amostral (DRAPER; SMITH, 1981).

Apesar do caráter relativo do coeficiente de determinação, parece lógico usa-lo como critério de seleção de modelos, já que, ao contrário do teste F do modelo, fornece uma "medida" do quanto o modelo é superior à média amostral. Sendo assim, se forem considerados dois modelos concorrentes cujos teste F do modelo rejeitassem a hipótese nula optaria-se pelo R^2 de maior valor, ou com melhor qualidade de ajuste.

No entanto, este caráter de medida de discrência em relação à média limita muito a utilização do R^2 na seleção de modelos quando este é aplicado isoladamente (CHATERJEE; PRICE, 1991). Para corroborar com essa afirmação os autores citam o trabalho de Anscombe (1973), onde quatro diferentes conjuntos de dados, com diferentes padrões, foram gerados e ajustados a um modelo de regressão linear simples.

Os coeficientes do modelo, os coeficientes de determinação, bem como os resultados dos testes de hipóteses foram os mesmos para os exemplos (a), (b), (c) e (d) da Figura 2.1, demonstrando o problema do uso dessas estatísticas como evidência de qualidade de ajuste quando a relação entre as variaveis escapa à linearidade.

Notadamente, (a) é o unico conjunto de dados em que a relação linear se aplica. Por ser o mesmo para todos os casos, se o R^2 fosse aplicado como critério para seleção entre os modelos (a), (b), (c) e (d), este não conseguiria captar a discrepância entre o modelo ajustado e o verdadeiro padrão sugerido pelo conjunto de dados.

Desta forma, a aplicação do R^2 como critério de seleção na análise de regressão não linear também é complicada. Bates e Watts (1988) afirmam que as propriedades dos modelos lineares, não

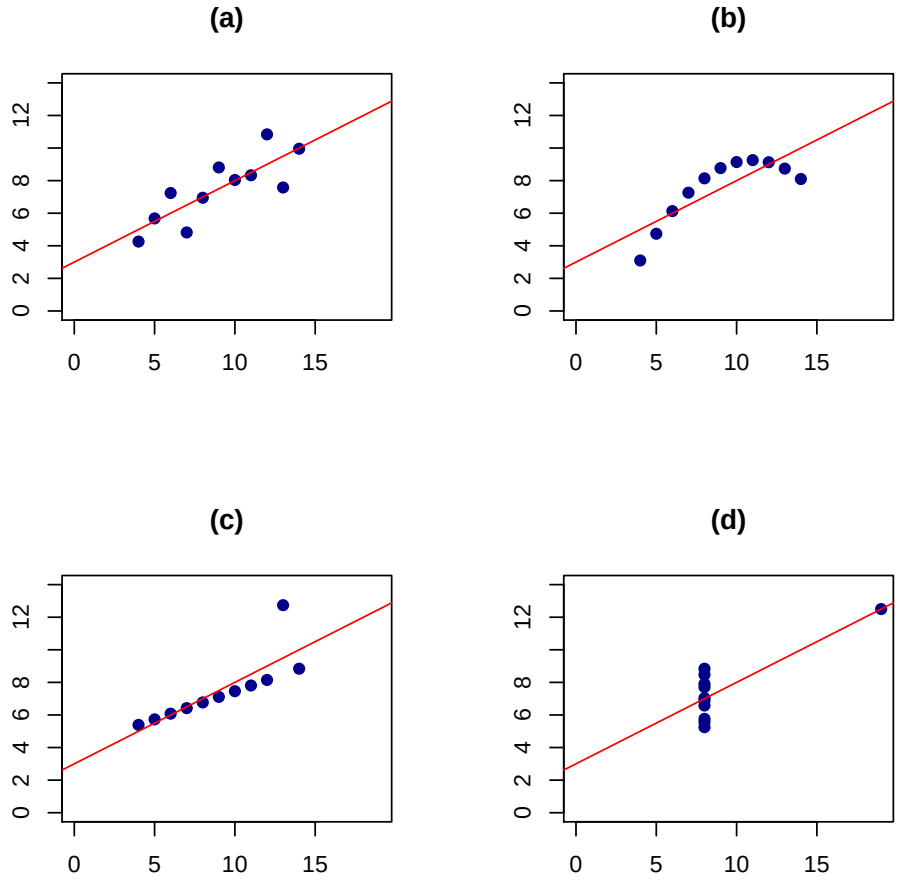

Figura 2.1 – Ajuste de quatro conjunto de dados (a, b, c e d) com padrões distintos, via regressão linear, onde foram obtidos o mesmo modelo $(y = 3 + 0,5x + e)$ e o mesmo $R^2(0,666)$ (retirado de Ascombe, 1977).

são válidas para os modelos não lineares. Por exemplo, a soma dos resíduos não necessariamente é

igual à zero e a soma de quadrados do erro mais a soma de quadrados da regressão, não necessariamente é igual à soma de quadrados total. Consequentemente, o coeficiente de determinação, pode assumir valores diferentes do intervalo [0,1], não sendo uma estatística descritiva importante para os modelos não lineares (CAMERON; WINDMEIJER, 1996).

Outra limitação ao uso do R^2 na seleção de modelos se refere à análise de regressão multipla. Neste tipo de análise nem todas variáveis preditoras são necessariamente efetivas para predição da variável resposta, mas a inclusão de novas variáveis geralmente reduz a soma de quadrados dos resíduos, consequentemente aumentando o valor do R^2 (SAKAMOTO; ISHIGURO; KITAGAWA, 1986).

O coeficiente de determinação ajustado (eq. 2.2), é uma tentativa de tentar corrigir esse problema ajustando o numerador e o denominador da eq. (2.1), através dos respectivos graus de liberdade (DRAPER; SMITH, 1998):

$$R_a^2 = \left(\frac{n-1}{n-k-1}\right) 1 - R^2 \tag{2.2}$$

onde R^2 é um argumento já definido, n é o tamanho da amostra e k o numero de parâmetros do modelo.

Contrariamente ao coeficiente de determinação, o coeficiente de determinação ajustado pode diminuir em valor se a contribuição da variável adicional na explicação da variação total, for inferior ao impacto que essa adição acarreta nos graus de liberdade (DRAPER; SMITH, 1998).

O coeficiente de determinação ajustado não possui a mesma interpretação do coeficiente de determinação, no que diz respeito a proporção da variabilidade da variável resposta explicada pelo modelo (CHATERJEE; PRICE, 1991). No entanto, o R_a^2 é também geralmente usado para julgar a qualidade do ajuste de modelos de regressão multipla possuindo, porém, as mesmas limitações do R^2, no que se refere a sua aplicação como critério unico de seleção de modelos lineares e não linerares.

Por fim, outro critério de seleção de modelos usado no meio florestal, aplicado a regressão multipla, é o teste F parcial. Este teste é, segundo Wada e Kashiwagi (1990), usado para comparar dois modelos concorrentes investigando as seguintes hipóteses:

H_0: Um subconjunto dos coeficientes é igual à zero.

$vs.$

H_a: Nenhum dos coeficientes da regressão é igual a zero.

Para Draper e Smith (1991), o teste F parcial, diferentemente do teste F do modelo, compara modelos hierárquicos, ou seja, o modelo completo com p coeficientes contra um modelo reduzido composto por um subconjunto desses coeficientes. Trata-se de um critério relativo, pois procura investigar se o modelo reduzido é tão adequado quanto o modelo completo.

Apesar de mais adequado como critério de seleção de modelos, o teste F parcial, compara somente modelos hierárquicos não possibilitando a comparação de modelos com diferentes relações funcionais (DRAPER; SMITH, 1998). Isto ocorre, pois, quando o tamanho da amostra é suficientemente grande, o que geralmente ocorre em problemas de modelagem florestal, o teste F parcial tende a escolher modelos mais complexos e rejeitar modelos parcimoniosos por apresentarem significativa falta de ajuste (KUHA, 2004). Além disso, Sakamoto, Ishiguro e Kitagawa (1986) afirmam que um modelo com um número desnecessáriamente grande de variáveis preditoras pode ser instável e gerar uma idéia falsa de super ajuste.

A partir das limitações da abordagem clássica torna-se interessante considerar abordagens distintas de seleção de modelos para o meio florestal. Uma possível alternativa é o da abordagem das discrepâncias entre dois modelos. Esta abordagem, além da utilidade e flexibilidade, permite uma nova visão no contexto da modelagem (FORSTER, 2000). Ela é particularmente útil por fornecer critérios de seleção relativos, parcimoniosos e aplicáveis a diferentes relações funcionais, ampliando assim, a liberdade do pesquisador e evitando os problemas de super ajuste.

Para se entender essa abordagem primeiro faz-se necessário definir alguns conceitos que auxiliam na compreensão do processo de modelagem de dados empíricos. O conceito de discrepância entre modelos como medida de falta de ajuste, bem como os elementos que compõe essa discrepância serão apresentados a seguir.

2.3 Abordagem das discrepâncias

2.3.1 Conceitos de discrepância

Em modelagem de dados estatísticos busca-se, a partir de um conjunto de dados tomados de uma população, descrever e ou inferir sobre um determinado fenômeno. A descrição completa desse fenômeno é muito difícil, já que as inferências são realizadas a partir de alguns parâmetros da população, arbitrariamente selecionados e estimados segundo um determinado esforço amostral (FORSTER, 2000).

Linhart e Zuchini (1986) denominam de modelo operacional $f(x)$, o modelo mais próximo do fenômeno a ser descrito pelo pesquisador, sendo que apenas em casos excepcionais se tem informação disponível para especificar completamente o modelo operacional. Para os mesmos autores, como regra, em modelagem pode-se apenas obter um modelo relativamente próximo a $f(x)$ a partir do conjunto de dados.

Zuchinni (2000) afirma que, é necessário primeiramente, especificar uma família de modelos de aproximação $g(x|\theta)$ com $\theta \in \Theta$, cujos membros individuais são indentificados pelo vetor de parâmetros $\theta = (\theta_1, \theta_2, ..., \theta_p)^T$. O modelo ajustado $g(x|\hat{\theta})$, um membro da familia de modelos $g(x|\theta)$ com $\theta \in \Theta$, é então obtido através de algum método de estimação dos parâmetros (mínimos

quadrados, máxima verossimilhança, por exemplo).

O próximo passo é selecionar o modelo ajustado $g(x|\hat{\theta})$ mais próximo do modelo operacional $f(x)$ e independente da estratégia utilizada para esta seleção existe como regra um número de aspectos em que o modelo ajustado e o modelo operacional diferem. Cada um desses aspectos de falta de ajuste pode ser visto como alguma discrepância Δ e esta, para que se realize a escolha do melhor modelo, deve ser minimizada (BUCKLAND; BURHAM; AUGUSTIN, 1997). Os mesmos autores dividem a dicrepância total em dois componentes: discrepância na aproximção e discrepância na estimação.

A discrepância entre o modelo operacional $f(x)$ e a família de modelos $g(x|\theta)$ é definida por:

$$\Delta(\theta) = \Delta[f(x), g(x|\theta)]. \tag{2.3}$$

Forster (2000) define que a discrepância na aproximação entre a família de modelos $g(x|\theta)$ e o modelo operacional $f(x)$ é dada por $\Delta(\theta_0)$, onde θ_0 é o valor que minimiza a eq. (2.3), tornando $g(x|\theta_0)$ o melhor modelo de aproximação para a familia $g(x|\theta)$. Assume-se que θ_0 exista e que é único e isso claramente ocorre se $f(x) \in g(x|\theta)$, com $\theta \in \Theta$.

Desta forma, a discrepância na aproximação não depende de forma alguma do tamanho da amostra e do método de estimação utilizado, mas somente do espaço paramétrico considerado (LINHART; ZUCHINNI, 1986).

Em contrapartida, Linhart e Zuchini (1986) afirmam que a discrepância devido à estimação se refere à discrepância entre o modelo ajustado $g(x|\hat{\theta})$ e o melhor modelo de aproximação $g(x|\theta_0)$ entre os membros a familia de modelos de aproximação $g(x|\theta)$, como segue:

$$\Delta[g(x|\theta_0), g(x|\hat{\theta})]. \tag{2.4}$$

A discrepância na estimação, ao contrário da discrepância na aproximação, depende fortemente do tamanho da amostra e dos métodos de estimação utilizados (FORSTER, 2000).

No processo de seleção de modelos, quanto maior for o espaço paramétrico considerado de um modelo, menor será sua discrepância na aproximação em relação ao modelo operacional. Em outras palavras, se apenas a discrepancia na aproximação for utilizada como quesito para escolha de modelos concorrentes esta privilegiará sempre os modelos de maior complexidade (FORSTER, 2000). Por outro lado, segundo o mesmo autor, modelos mais complexos tendem à apresentar maior incerteza na estimativa dos seus parâmetros, ou seja, tendem à apresentar maior discrepância na estimação.

O problema de seleção de modelos é encontrar, dentro deste contexto, um caminho para conciliar essas duas propriedades opostas e encontrar o modelo de melhor performance geral (ZUCHINNI, 2000).

A discrepância geral $\Delta(\theta_0)$ ou a discrepância entre o modelo operacional $f(x)$ e o modelo ajustado $g(x|\hat{\theta})$ é definida em (ZUCHINNI, 2000) por:

$$\Delta(\theta_0) = \Delta[f(x), g(x|\hat{\theta})]. \tag{2.5}$$

Burham e Anderson (2002) afirmam que, na prática, não é possivel calcular nenhuma das dicrepâncias acima, pois não se conhece completamente o modelo operacional $f(x)$. Ou seja, para uma determinada amostra não é possivel calcular diretamente a dicrepância geral e nem mesmo seus dois componentes.

No entanto, segundo Zuchinni (2000), pode-se obter alguma ajuda na seleção de modelos através do calculo do seu valor esperado para uma determinada amostra. Esta é chamada de função de discrepância geral esperada, definida por:

$$E\Delta(\theta_0) = \Delta[f(x)), g(x|\hat{\theta})]. \tag{2.6}$$

Porém, também não é possível calcular a discrepância geral estimada, representada pela eq. 2.6, sem o conhecimento do modelo operacional $f(x)$. Uma saída para este problema é estimar essa discrêpancia a partir de evidencias fornecidas pelo conjunto de dados, obtendo-se um estimador da dicrepância geral esperada. Este estimador é, de acordo com Zuchinni (2000), um critério de seleção de modelos baseado na perspectiva das discrepâncias.

Uma função de discrepância geral esperada largamente utlizada é a chamada distancia entre dois modelos de Kulbach-Leibler. Esta medida de discrepância, baseada no principio de entropia máxima de Boltzman (veja Burham e Anderson (2002) para mais detalhes) forneceu uma importante base teórica para criação de diversos critérios de seleção de modelos nas ultimas décadas, incluindo o Critério de Informação de Akaike.

2.3.2 Distância de Kulbach-Leibler e Entropia

A interpretação estatística de entropia, uma medida do aumento de energia de qualquer sistema isolado termodinamicamente, foi desenvolvida por Boltzmann no final do século XIX (AKAIKE, 1985). Essa foi definida em Boltzman (1877 apud AKAIKE, 1985) como uma medida da entropia de uma distribuição $g(x)$ em relação a outra $f(x)$:

$$B(f, g) = -\int f(x)\, log\left[\frac{f(x)}{g(x)}\right]\, dx. \tag{2.7}$$

A entropia é critério natural do ajuste de $g(x)$ em relação a $f(x)$ e quanto maior a entropia melhor é aproximação entre os modelos, sendo este chamado de principio da máxima entropia (AKAIKE, 1981).

Desta forma, dois modelos estão mais próximos quanto maior a entropia entre eles ou quanto menor a entropia negativa ou "neg-entropia" (BOSDOGAN, 2000).

$$I(f,g) = -B(f,g) \tag{2.8}$$

A entropia negativa é também chamada de distância de Kulbach-Leibler $(I(f,g))$ e por consequência a minimização desta distância é equivalente a maximização da entropia de Boltzman (BOSDOGAN, 2007).

No processo de modelagem precisamos escolher o melhor modelo de aproximação $g(x|\theta_0)$ em relação ao modelo operacional $f(x)$, que gerou os dados. Sendo assim e considerando que se conhece o modelo operacional, pode-se medir a distância entre este e diversas familias de aproximação propostas $g(x|\theta)$, através do cálculo da distância de Kulbach-Leibler (LiNHART; ZUCHINNI, 1986).

Neste caso, considerando-se o modelo operacional $f(x)$ como dado (fixo) e somente $g(x|\theta)$ variando num espaço de modelos θ, procura-se uma familia de aproximação que minimize $I(f,g)$. Isto pode ser feito aplicando a eq. 2.9 às diferentes familias de aproximação $g(x|\theta)$ e ao modelo operacional $f(x)$

$$I(f,g) = \int f(x)\ log\left[\frac{f(x)}{g(x|\theta)}\right]\ dx. \tag{2.9}$$

A distância de Kulbach-Leibler, segundo Burham e Anderson (2004) pode ser vista como medida de discrepância da família $g(x|\theta)$ ao se aproximar de $f(x)$, sendo, no entanto, necessário se conhecer completamente $f(x)$ e ainda conhecer os parâmetros de $g(x|\theta)$ para ser calculada.

Sendo assim, a distancia de Kulbach-Leibler pode ser conceitualizada como uma distância direta entre dois modelos e, segundo Wada e Kashiwagi (1990), é a mais fundamental de todas as medidas de informação pela simplicidade e propriedades aditivas, tornando-a, em conjunto com a teoria de máxima verossimilhança, uma base racional para seleção de modelos.

Como dito até agora, para se calcular a distância de Kulbach-Leibler entre $f(x)$ e $g(x)$ faz-se necessário conhecer ambos os modelos bem como seus parâmetros. Porém, usando uma medida de distancia relativa, pode-se comparar vários modelos de aproximação $g(x|\theta)$ sem a necessidade de conhecer o modelo operacional. Isto ocorre, pois, a eq. 2.9 pode ser também apresentada da seguinte forma:

$$I(f,g) = \int f(x)\ log[f(x)]\ dx - \int f(x)\ log[g(x|\theta)]\ dx. \tag{2.10}$$

Cada um dos dois termos da eq. (2.10), segundo Burham e Anderson (2000), se refere ao valor esperado em relação ao modelo operacional $f(x)$, podendo ser reescrita como a diferença de duas esperanças em relação a $f(x)$, como segue:

$$I(f,g) = E_f[log(f(x))] - E_f[log(g(x|\theta))]. \tag{2.11}$$

O primeiro termo é uma constante que depende do modelo operacional $f(x)$. Chamando esse termo constante de C, tem-se:

$$I(f,g) = C - E_f[log(g(x|\theta))] \tag{2.12}$$

ou ainda

$$I(f,g) - C = -E_f[log(g(x|\theta))]. \tag{2.13}$$

Sendo assim, tratando o termo desconhecido como uma constante, é possível se calcular a distância relativa entre várias familias de aproximação candidatas $g(x|\theta)$ e o modelo operacional $f(x)$ (BOSDOGAN, 1987).

O termo $I(f,g)-C$ é a distância relativa entre $f(x)$ e $g(x|\theta)$, tornando $-E_f[log(g(x|\theta)]$ a quantidade de interesse para selecionar o melhor modelo de aproximação $g(x|\theta_0)$ entre várias familias de aproximação $g(x|\theta)$ candidatas. Desta forma pode-se postular várias familias de aproximação $g(x|\theta)$ e selecionar a melhor entre elas.

Calcular, no entanto, a distância relativa entre o modelo operacional $f(x)$ e as várias candidatas $g(x|\theta)$, é calcular somente a discrepância na aproximação. Mas o que se busca é um modelo que minimize a discrepância geral, já que na prática não se tem informação sobre o parâmetro θ da familia de aproximação, sendo necessário estima-lo a partir do conjunto de dados. Neste caso, precisa-se estimar a distância entre $g(x|\hat{\theta})$ e $f(x)$ para que se escolha o modelo com a menor distância relativa estimada, ou modelo com menor discrepância geral esperada estimada. Akaike (1985) encontrou uma maneira de obter a estimativa da discrepância geral esperada de Kulbach-Leibler, baseada na função de log-verossimilhança no seu ponto máximo (BURHAM; ANDERSON, 2002).

2.3.3 Critério de informação de Akaike

Akaike (1985, 1994) afirma que aplicar a função de dicrepância geral esperada de Kulbach-Leibler como critério de seleção de modelos é aplicar essa dupla esperança:

$$E_y E_x[log(g(x|\hat{\theta}(y)))] \tag{2.14}$$

onde o termo interno é equivalente a $E_f[log(g(x|\theta))]$ (distância de Kulbach-Leibler), com o parâmetro θ sendo substituido pela estimativa de máxima verossimilhança de θ, baseada na familia de distribuição $g(x|\theta)$ e no conjunto de dados y.

Segundo Burham e Anderson (2004), apesar de y denotar o conjunto de dados, x e y devem ser conceitualizadas como sendo amostras aleatórias tomadas do mesmo modelo operacional $f(x)$ e, portanto, com as duas esperanças tomadas em relação a este mesmo modelo.

Akaike (1973, 1974) relacionou, então a distância de Kulbach- Leibler e teoria da verossimilhança, demonstrando que a estimativa de máxima log-verossimlhança era um estimador enviesado de $E_y E_x[log(g(x|\hat{\theta}(y)))]$.

Bosdogan (2000)e Sakamoto (1986) afirmam que, apesar de $log(L(\hat{\theta}|dados))$ ser um estimador enviesado de $E_y E_x[log(g(x|\hat{\theta}(y)))]$, este viés é aproximadamente igual ao número de parâmetros p da familia de aproximação $g(x|\theta)$. Sendo esse um resultado assintótico de fundamental importância (BURHAM; ANDERSON, 2004).

Sendo assim, um estimador não enviesado de $E_y E_x[log(g(x|\hat{\theta}(y)))]$ para grandes amostras, como se tem geralmente na meio florestal, torna-se:

$$log(L(\hat{\theta}|dados)) - p. \tag{2.15}$$

Sendo esse resultado, então equivalente à:

$$log(L(\hat{\theta}|dados)) - p = C - \hat{E}_{\hat{\theta}}[I(f, \hat{g})], \tag{2.16}$$

onde $\hat{g} = g(x|\hat{\theta})$.

Este estimador da distância relativa de Kulbach-Leibler, segundo Bosdogan (2000), torna possivel combinar estimação e seleção de modelos numa estrutura de otimização unificada. Em outras palavras, pode-se através de um unico critério estimar a discrepancia geral esperada de uma familia de aproximação a partir de um conjunto de dados.

Desta maneira, Akaike (1973 apud SAKAMOTO; ISHIGURO; KITAGAWA, 1986) encontrou um estimador da distância relativa esperada de Kulbach-Leibler baseada na função de máxima log-verossimilhança, corrigida pelo viés assintótico p

$$log(L(\hat{\theta}|dados)) - p = \hat{E}[I(f, g)]\ relativa. \tag{2.17}$$

O termo p é o termo de correção do viés da estimativa e é importante ressaltar que esse termo não é de forma alguma arbitrário e confere ao AIC propriedades assintóticas desejadas (veja Burham e Anderson (2004) para mais detalhes)

Sendo assim, Akaike (1973) (1973 apud SAKAMOTO; ISHIGURO; KITAGAWA, 1986) definiu "an information criterion" (AIC) multiplicando os dois termos desse resultado por -2, ficando este conhecido como Akaike Information Criterion, ou simplesmente AIC.

$$AIC = -2log(L(\hat{\theta}|dados)) + 2p \tag{2.18}$$

O motivo da multiplicação desses termos por -2 é importante, mas completamente técnico e não será alvo desta revisão, sendo que para maiores detalhes veja Bosdogan (1987) e Akaike (1985).

Sakamoto; Ishiguro e kitagawa (1986) afirmam que para o modelo homocedástico e com erros gaussianos, comumente usados no meio florestal, o AIC torna-se:

$$AIC = -2log(L(\hat{(\sigma^2)}) + 2p \tag{2.19}$$

onde n é o tamanho da amostra, p é o número de parâmetros (incluindo os coeficientes do modelo e $\hat{\sigma^2}$) e $\hat{\sigma^2} = \frac{\sum(\hat{\epsilon^2})}{n}$.

Desta forma, o valor de AIC é simples de ser obtido para os casos de estimação por mínimos quadrados, como é o caso do ajuste de modelos de regressão, e para os casos de análises baseadas na estimativas de verossimilhança de uma forma geral (SEVERINI, 2000).

Assumindo uma série de familias de aproximação $a\ priori$ selecionadas e fundamentadas em alguma teoria científica , o AIC pode ser obtido para cada familia, classificando-as da melhor para a pior com base no conjunto de dados disponível. As melhores familias de aproximação serão aquelas com menores valores de AIC, ou com equivalentes menores estimativas de discrepância geral esperada, em relação ao modelo operacional (BOSDOGAN, 1987).

Se parâmetros conhecidos forem adicionados a uma família de aproximação candidata, o valor de AIC diminuirá. Mas como não se conhece em geral os parâmetros de uma determinada familia, sendo estes apenas estimados a partir dos dados, o acréscimo de parâmetros irá aumentar a incerteza na estimação, e por conseqüência, aumentar o valor de AIC (KUHA, 2004).

Neste sentido o AIC é um critério de seleção que privilegiará famílias de aproximação parcimoniosas mais próximas ao modelo operacional. No entanto é preciso ressaltar que este critério irá apenas classificar as familias propostas, escolhendo uma entre elas, mesmo que todas estejam bem distantes do modelo operacional (BUCKLAND; BURHAM; AUGUSTIN, 1997).

2.4 Considerações Finais

Alguns autores (FORSTER, 2000; KUHA, 2004), interpretam o AIC também como um otimizador do balanço entre entrada e saida de parâmetros no processo de escolha de modelos parcimoniosos. No entanto, apesar dessa interpretação não ser incorreta, o presente trabalho prefere a interpretação do AIC como estimador da discrepância geral esperada, de acordo com a perspectiva de Burham e Anderson (2004) e Linhart e Zuchinni (1986).

Na segunda perspectiva fica claro que o AIC, apesar da sua simplicidade quanto a forma e quanto a aplicação, é baseado em conceitos profundos e extremamente consolidados da teoria da informação (Distância de Kulbach-Leibler) e da teoria estatística (paradigma da verossimilhança).

Outro conceito importante no processo de modelagem, inserido nessa perspectiva, é o da intangibilidade do modelo operacional. Em outras palavras, a realidade completa nunca pode ser descrita a partir de um modelo, sendo esta apenas circunscrita através de uma série de modelos de aproximação concorrentes.

Isto implica que o processo de modelagem não é simplesmente um processo de estimação, mas também um processo de aproximação de uma realidade que não pode ser descrita totalmente pelo modelo.

No meio florestal, como ja visto, se usam critérios de seleção de modelos que não permitem

comparar modelos de diferentes relações funcionais, simplificando o processo de modelagem à estimação e à definição da ordem (número de parâmetros) do modelo. Além disso, não existe nenhum critério que aplicado isoladamente classifique os modelos, quanto a sua falta de ajuste, em relação ao modelo operacional.

A hipótese deste trabalho é que o AIC pode, se interpretado e usado corretamente sob o ponto de vista das discrepâncias, vir a ser uma excelente ferramenta de seleção de modelos empíricos de predição no meio florestal.

Referências

AKAIKE, H. A New Look at Statistical Model Identification, **IEEE Transactions on Automatic Control**, Tokio, v. 19, n. 6, p. 717-723, Dec. 1974.

AKAIKE, H. Likelihood of a Model and Information Criteria, **Journal of Econometrics**, Amsterdan, v. 16, n.1, p. 3-14, May. 1981.

ANSCOMBE, F. J. Graphs in Statistical Analysis, **The American Statistician**, v. 27, n. 1, p. 17-21, Alexandria, Feb. 1973.

BOSDOGAN, H. Model Selection and Akaike Information Criterion (AIC): The General Theory and Its Analitical Extensions, **Psychometrika**, New York, v. 52, n. 3, p. 345-370, Sept. 1987.

BOSDOGAN, H. Akaike Information Criterion and Recent Developments in Information Complexity, **Journal of Mathematical Psychology**, Oxford, v. 44, n. 1, p. 62-91, Mar. 2000.

BUCKLAND, S. T.; BURNHAM, K. P. e AUGUSTIN, N. H. Model Selection: A Integral Part of Inference, **Biometrics**, London, v. 53, n. 2, p. 603-618, Jun. 1997.

BURNHAM, K.P. e ANDERSON, D.R. Model Selection and Multimodel Inference: Practical Information Theoretic Approach. 2.ed. New York: Sprinter, 2002. 488 p.

BURNHAM, K.P. e ANDERSON, D.R. Multimodel Inference: Understanding AIC e BIC in Model Selection, **Sociological Methods Reseach**, London, v. 33, n. 2, p. 261-304, Nov. 2004.

CAMERON, A. C. e WINDMEIJER, F.A.G. An R-squared Measure of Goodness of Fit for Some Commom Nonlinear Regression Models, **Journal of Econometrics**, Amsterdan, v. 77, n. 2, p. 329-342, Apr. 1997.

CAMPOS, J. C. C., CAMPOS, A. L. A. S. e LEITE, H. G. Decisão Silvicultural Empregando um Sistema de Predição do Crescimento e da Produção, **Revista Árvore**, Viçosa, v. 12, n. 2, p. 100-110, mai. 1988.

CAMPOS, J. C. C., SOARES, C. P. B., LEITE, H. G. e CAMPOS, M. B. Estimação de Diferentes Volumes Comerciais Utilizando um Modelo do tipo Povoamento Total, **Revista Árvore**, Viçosa, v. 25, n. 2, p. 223-230, set. 2001.

CHATTERJEE, S., PRICE, B. Regression analysis by example. 2.ed. New York : Wiley, 1977, 228 p.

DRAPER, N. R. e SMITH, H. Applied Regression Analysis. 3.ed. United States of America: A Wiley-Interscience Publication, 1998, 706 p.

FORSTER, M. R. Key Concepts in Model Selection: Performance and Generalizability, **Journal of Mathematical Psychology**, Oxford, v. 44, n. 1, p. 205-231, Mar. 2000.

KUHA, J. AIC e BIC: Comparisions of Assumptions and Performance, **Sociological Methods Research**, london, v. 33, n. 2,p. 188-228, Nov. 2004.

LEITE, H. G., NOGUEIRA, G. S., MOREIRA, A. M. e LIMA, J. E. Um Modelo de Crescimento e Produção para *Pinus taeda L.* na Região Sul do Brasil, **Revista Árvore**, Viçosa, v. 25, n. 1, p. 105-112, mai. 2001.

MAESTRI, R., SANQUETTA, C. R., MACHADO, S. A., SCOLFORO, J. R. S. e CORTE, A. P. D. Viabilidade de um projeto florestal de *Eucalyptus grandis* considerando o seqüestro de carbono, **Floresta**, Curitiba, v. 34, n. 3, p. 347-360, dez. 2004.

SAKAMOTO, Y., ISHIGURO, M., KITAGAWA, G. Akaike Information Criterion Statistics. 1.ed. Tokio: KTK Scientific Publisher/D. Riedel, 1986, 290 p.

SOARES, C. P. B. e OLIVEIRA, M. L. R. Equações para estimar o carbono na parte aérea de árvores de eucalipto em Viçosa, Minas Gerais, **Revista Árvore**, Viçosa, v. 26, n. 5, p. 533-539, set. 2002.

SOARES, C. P. B., LEITE, H. G. e GÖRGENS, E. B. Equações para estimar o estoque de carbono no fuste de árvores individuais e em plantios comerciais de eucalipto, **Revista Árvore**, Viçosa, v. 29, n. 5, p. 711-718, set. 2005.

WADA, Y. e KASHIWAGI, N. Selecting Statistical Models with Information Statistics, **Journal of Dairy Science**, Palo Alto, v.73, n. 2, p. 3575-3582, Jun. 1990.

RAO, C. Linear statistical inference and its applications. 2.ed. New York: Wiley, 1973, 625 p

LINHART, H., ZUCHINNI, W. Model selection. 2.ed. New York: Wiley, 1986, 301 p.

ZUCHINNI, W. An introduction to model selection, **Journal of Mathematical Psychology**, Oxford, v. 44, n. 1, p. 41-61, Mar. 2000.