Psychonomic Bulletin & Review
2004, 11 (1), 192-196

Notes and Comment

AIC model selection using Akaike weights

ERIC-JAN WAGENMAKERS and SIMON FARRELL
Northwestern University, Evanston, Illinois

The Akaike information criterion (AIC; Akaike, 1973) is a popular method for comparing the adequacy of multiple, possibly nonnested models. Current practice in cognitive psychology is to accept a single model on the basis of only the "raw" AIC values, making it difficult to unambiguously interpret the observed AIC differences in terms of a continuous measure such as probability. Here we demonstrate that AIC values can be easily transformed to so-called Akaike weights (e.g., Akaike, 1978, 1979; Bozdogan, 1987; Burnham & Anderson, 2002), which can be directly interpreted as conditional probabilities for each model. We show by example how these Akaike weights can greatly facilitate the interpretation of the results of AIC model comparison procedures.

The evaluation of competing hypotheses is central to the process of scientific inquiry. When the competing hypotheses are stated in the form of predictions from quantitative models, their adequacy with respect to observed data can be rigorously assessed. Given K plausible candidate models of the underlying process that has generated the observed data, we should like to know which hypothesis or model approximates the "true" process best. More generally, we should like to know how much statistical evidence the data provide for each of the K models, preferably in terms of likelihood (Royall, 1997) or the probability of each of the models' being correct (or the most correct, because the generating model may never be known for certain). The process of evaluating candidate models is termed *model selection* or *model evaluation.*

A straightforward solution to the problem of evaluating several candidate models is to select the model that gives the most accurate description of the data. However, the process of model evaluation is complicated by the fact that a model with many free parameters is more flexible than a model with only a few parameters. It is clearly not desirable to always deem the most complex model the best, and it is generally accepted that the best model is the one that provides an adequate account of the data while using a minimum number of parameters (e.g., Myung, Forster, & Browne, 2000; Myung & Pitt, 1997). Any criterion for model selection needs to address this tradeoff between descriptive accuracy and parsimony.

One of the more popular methods of comparing multiple models, taking both descriptive accuracy and parsimony into account, is the Akaike information criterion (AIC; see, e.g., Akaike, 1973, 1974, 1978, 1979, 1983, 1987; Bozdogan, 1987; Burnham & Anderson, 2002; Parzen, Tanabe, & Kitagawa, 1998; Sakamoto, Ishiguro, & Kitagawa, 1986; Takane & Bozdogan, 1987).[1] The AIC is a very general method that is applied in a wide range of situations relevant to cognitive psychology. For instance, the AIC often used as a measure of model adequacy in both structural equation modeling (Jöreskog & Sörbom, 1996) and time series analysis (e.g., McQuarrie & Tsai, 1998). The AIC is also applied in factor analysis (e.g., Akaike, 1987), regression (e.g., Burnham & Anderson, 2002; see also Efron, 1986), and latent class analysis (e.g., Eid & Langeheine, 1999). A sample of other recent contexts in which the AIC has been employed includes competitive testing of models of categorization (e.g., Maddox & Bohil, 2001; Nosofsky, 1998; Thomas, 2001), modeling of mixture distributions (Raijmakers, Dolan, & Molenaar, 2001), modeling of luminance detection (Smith, 1998), and modeling of perception with the use of stochastic catastrophe models (Ploeger, van der Maas, & Hartelman, 2002).

After a brief description of the AIC, we will discuss how the usual manner in which results from an AIC analysis are reported may often be imprecise or confusing. We will then show how the raw AIC values can be transformed to conditional probabilities for each model. This transformation has been recommended by Akaike (1978, 1979), Bozdogan (1987), and Burnham and Anderson (2002), but hitherto it has not been applied in cognitive psychology. By means of a hypothetical example, we will illustrate how the proposed transformation makes the results of AIC analyses easier to interpret and also increases the accessibility of the results for further analysis. Finally, we will discuss some similarities and differences between the AIC and another popular model selection criterion, the Bayesian information criterion (BIC).

AIC Model Selection

The objective of AIC model selection is to estimate the information loss when the probability distribution f associated with the true (generating) model is approximated by probability distribution g, associated with the model that is to be evaluated. A measure for the dis-

We thank Han van der Maas and In Jae Myung for helpful comments on an earlier draft of this paper. Correspondence concerning this article can be addressed to E.-J. Wagenmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands (e-mail: ewagenmakers@fmg.uva.nl), or to S. Farrell, Department of Experimental Psychology, University of Bristol, 8 Woodland Road, Clifton, Bristol BS8 1TN, England (e-mail: simon.farrell@bristol. ac.uk).

crepancy between the true model and the approximating model is given by the Kullback–Leibler (1951) information quantity $I(f, g)$, which is equal to the negative of Boltzmann's (1877) generalized entropy (for details, see Bozdogan, 1987; Burnham & Anderson, 2002; Golan, 2002; and Sakamoto et al., 1986).

Akaike (1973; Bozdogan, 1987) has shown that choosing the model with the lowest expected information loss (i.e., the model that minimizes the expected Kullback–Leibler discrepancy) is asymptotically equivalent to choosing a model M_i, $i = 1, 2, \ldots, K$ that has the lowest AIC value. The AIC is defined as

$$\mathrm{AIC}_i = -2\log L_i + 2V_i, \tag{1}$$

where L_i, the maximum likelihood for the candidate model i, is determined by adjusting the V_i free parameters in such a way as to maximize the probability that the candidate model has generated the observed data. Equation 1 shows that the AIC rewards descriptive accuracy via the maximum likelihood, and penalizes lack of parsimony according to the number of free parameters (note that models with smaller AIC values are to be preferred). Equation 1 is based on asymptotic approximations and is valid only for sufficiently large data sets. The finite sample correction

$$\mathrm{AIC}_c = -2\log L + 2V + \frac{2V(V+1)}{(n-V-1)}$$

(e.g., Hurvich & Tsai, 1995; Sugiura, 1978) is generally recommended when $n/V < 40$ (Burnham & Anderson, 2002, p. 445).

Table 1 shows hypothetical results which might be obtained from fitting five different models (i.e., A1, A2, B1, B2, and C) to an experiment, in this case a categorization task. The groups of Models A, B, and C can be considered different models (e.g., an exemplar model, a boundary model, and a prototype model). Models A2 and B2 differ from A1 and B1, respectively, by the inclusion of an extra parameter that might correspond, for example, to the level of attention of the participants (see Maddox & Ashby, 1993, for an example of such a set of models).

The first column of Table 1 shows the number of free parameters, and the second column shows the log likelihood for each model obtained by using maximum likelihood parameter estimation (MLE; e.g., Eliason, 1993; Sakamoto et al., 1986). Given the number of free parameters and the log likelihood, Equation 1 can be applied to yield AIC values; these are shown in the third column of Table 1. From an inspection of the AIC values it is apparent that Model A2 is the preferred model (i.e., it has the lowest AIC value of the five candidate models). However, it is difficult to intuit how much statistical importance we should attach to a difference in the AIC values between the best model (A2, AIC = 202) and the next best model (A1, AIC = 204). It is especially important to assess the weight of evidence in favor of the best model when a binary decision is made and the other candidate models (with higher AIC values) are simply discarded. When the AIC differences are very small, the acceptance of a single model may lead to a false sense of confidence. In addition, the raw AIC values cannot tell us what the weight of evidence is in favor of Models A1 and A2 over Models B1 and B2—that is, the extent to which the data support Model A over Model B. Such considerations are important in situations where a specific Model A2 may have the lowest AIC, but Model B may generally be the overall better model.

Akaike Weights

We now show how these questions can be easily and satisfactorily addressed by considering a simple transformation of the raw AIC values. First, we compute, for each model, the differences in AIC with respect to the AIC of the best candidate model (e.g., Akaike, 1978; Burnham & Anderson, 2002); that is,

$$\Delta_i(\mathrm{AIC}) = \mathrm{AIC}_i - \min \mathrm{AIC}. \tag{2}$$

Continuing our example, column 4 of Table 1 gives $\Delta_i(\mathrm{AIC})$ for each of the five models. This procedure reflects our interest in the relative performance of the models, not their absolute AIC values. For the next step, we note that the AIC is an unbiased estimator of minus twice the expected log likelihood of the model (Akaike,

Table 1
Results of AIC and BIC Analysis for Five Competing Models (Hypothetical Data)

Model	No. Par_i	$\log(L_i)$	AIC_i	$\Delta_i(\mathrm{AIC})$	$w_i(\mathrm{AIC})$	BIC_i	$\Delta_i(\mathrm{BIC})$	$w_i(\mathrm{BIC})$
A1	2	−100	204	2	.2242	211.0	0	.6439
A2	3	−98	202	0	.6094	212.4	1.48	.3071
B1	3	−100	206	4	.0824	216.4	5.48	.0416
B2	4	−99	206	4	.0824	219.9	8.96	.0073
C	4	−103	214	12	.0015	227.9	16.96	.0001

Note—No. Par_i = number of estimated parameters for model i; $\log(\mathrm{L}_i)$ = natural logarithm of the maximum likelihood for model i; $\Delta_i(\mathrm{AIC}) = [\mathrm{AIC}_i - \min(\mathrm{AIC})]$; $w_i(\mathrm{AIC})$ = the rounded Akaike weights; $\Delta_i(\mathrm{BIC}) = [\mathrm{BIC}_i - \min(\mathrm{BIC})]$; $w_i(\mathrm{BIC})$ = the rounded Schwarz weights. The number of observations that enters into the calculation of the maximum likelihood is 240. See text for details.

1978, pp. 218–220, and 1979, p. 239; Bozdogan, 1987, pp. 353–356). From the differences in AIC, we can then obtain an estimate of the relative likelihood L of model i by the simple transform:

$$L(M_i \mid \text{data}) \propto \exp\left\{-\frac{1}{2}\Delta_i(\text{AIC})\right\}, \tag{3}$$

where $\propto$ stands for "is proportional to." In the last step, the relative model likelihoods are normalized (i.e., divided by the sum of the likelihoods of all models) to obtain Akaike weights w_i(AIC) (e.g., Burnham & Anderson, 2002), where

$$w_i(\text{AIC}) = \frac{\exp\left\{-\frac{1}{2}\Delta_i(\text{AIC})\right\}}{\sum_{k=1}^{K}\exp\left\{-\frac{1}{2}\Delta_k(\text{AIC})\right\}}, \tag{4}$$

so that $\sum w_i(\text{AIC}) = 1$. Table 1, column 5, shows the set of Akaike weights for the illustrative data.[2] Weight w_i(AIC) can be interpreted as the probability that M_i is the best model (in the AIC sense, that it minimizes the Kullback–Leibler discrepancy), given the data and the set of candidate models (e.g., Burnham & Anderson, 2001). Thus, the strength of evidence in favor of one model over the other is obtained by dividing their Akaike weights. Note that the Akaike weights are subject to sampling variability, and that a different sample will most likely generate a different set of weights for the models in the candidate set.

Recall that the conclusion from the raw AIC values was that Model A2 is the preferred model. From an inspection of the Akaike weights in Table 1, it can be easily inferred that the best-fitting Model A2 is

$$\frac{w_{\text{A2}}(\text{AIC})}{w_{\text{A1}}(\text{AIC})} = \frac{.6094}{.2242} \approx 2.7$$

times more likely to be the best model in terms of Kullback–Leibler discrepancy than is the next-best Model A1. Note that the evidence ratio of Akaike weights for model i over model j can also be directly calculated by

$$\begin{aligned}\frac{w_i(\text{AIC})}{w_j(\text{AIC})} &= \exp\left(\log L_i - \log L_j + V_j - V_i\right)\\ &= \frac{L_i}{L_j}\exp\left(V_j - V_i\right).\end{aligned}$$

To get an intuitive feeling for how much support this evidence ratio provides in favor of A2 over A1, we can also express the evidence ratio as the normalized probability that Model A2 is to be preferred over Model A1:

$$\frac{w_{\text{A2}}(\text{AIC})}{w_{\text{A1}}(\text{AIC}) + w_{\text{A2}}(\text{AIC})} = \frac{.61}{.61 + .22} \approx .73.$$

Thus, using Akaike weights, we will again arrive at the conclusion that Model A2 is to be preferred over its competitors, but in addition the Akaike weights provide a continuous measure of strength of evidence. Finally, using the evidence ratio of the weights, we can also compute that Model A (i.e., both A1 and A2) is

$$\frac{w_{\text{A1}}(\text{AIC}) + w_{\text{A2}}(\text{AIC})}{w_{\text{B1}}(\text{AIC}) + w_{\text{B2}}(\text{AIC})} \approx 5.1$$

more likely than Model B (normalized probability .83). Similarly, the evidence ratio for models including an attentional parameter (i.e., A2 and B2) versus models without such a parameter (i.e., A1 and B1) is

$$\frac{w_{\text{A2}}(\text{AIC}) + w_{\text{B2}}(\text{AIC})}{w_{\text{A1}}(\text{AIC}) + w_{\text{B1}}(\text{AIC})} \approx 2.3$$

(giving a normalized probability of .69).

Comparison Between AIC and BIC

Despite the widespread use of the AIC, some believe that it is too liberal and tends to select overly complex models (e.g., Kass & Raftery, 1995). It has been pointed out that the AIC neglects the sampling variability of the estimated parameters. When the likelihood values for these parameters are not highly concentrated around their maximum value, this can lead to overly optimistic assessments (for an illustration, see Aitchison & Dunsmore, 1975, pp. 227–234). Furthermore, the AIC is not *consistent*. That is, as the number of observations n grows very large, the probability that the AIC recovers a true low-dimensional model does not approach unity (e.g., Bozdogan, 1987, p. 357). A popular alternative model selection criterion is the Bayesian information criterion or BIC (e.g., Burnham & Anderson, 2002; Hastie, Tibshirani, & Friedman, 2001; Kass & Raftery, 1995; Schwarz, 1978; Wasserman, 2000). The BIC (Schwarz, 1978) for model i is defined as

$$\text{BIC}_i = -2\log L_i + V_i \log n, \tag{5}$$

where n is the number of observations that enter into the likelihood calculation. The BIC is an asymptotic approximation to a Bayesian model selection (BMS) analysis, in which one integrates over the parameter space. Specifically, BMS requires the computation of the probability of the data given the model, $P(D \mid M_i)$, by incorporating the variability in the parameter vector θ_i: $P(D \mid M_i) = \int P(D \mid \theta_i, M_i)\,\pi(\theta_i \mid M_i)d\theta_i$, where $\pi(\theta_i \mid M_i)$ is the prior density. The choice of an adequate prior density is often a delicate matter that is left to the better judgment of the researcher. The BIC is much easier to compute than BMS, and it does not require the researcher to determine prior densities for the parameters. In contrast to the AIC, the BIC is consistent as $n \to \infty$ and does take parameter uncertainty into account.

A comparison of BIC (Equation 5) and AIC (Equation 1) shows that the BIC penalty term is larger than the AIC penalty term when $n > e^2$. Although the equations of AIC and BIC look very similar, they originate from quite different frameworks. The BIC assumes that the true generation model is in the set of candidate models, and it measures the degree of belief that a certain model is the true data-generating model. The AIC does not assume that any of the candidate models is necessarily true, but rather calculates for each model the Kullback–Leibler discrepancy, which is a measure of distance between the probability density generated by the model and reality. The merits and debits of AIC and BIC have been discussed elsewhere (cf. Burnham & Anderson, 2002, pp. 293–305, for a pro-AIC account, and Kass & Raftery, 1995, for a pro-BIC account). A formal comparison in terms of performance between AIC and BIC is very difficult, particularly because AIC and BIC address different questions. Most simulations that show BIC to perform better than AIC assume that the true model is in the candidate set and that it is relatively low dimensional. In contrast, most simulations that favor AIC over BIC assume that reality is infinitely dimensional, and hence the true model is not in the candidate set.

Just as raw AIC values may be converted to Akaike weights, a well-known procedure exists for transforming raw BIC values to BIC model weights (or "Schwarz weights"). Schwarz weights can be obtained by replacing the AIC values in Equation 4 by BIC values (cf. Buckland, Burnham, & Augustin, 1997; Burnham & Anderson, 2002, p. 297; Hastie et al., 2001, p. 207). The evidence ratio for model i over model j can also be calculated directly by

$$\frac{w_i(\mathrm{BIC})}{w_j(\mathrm{BIC})} = \frac{L_i}{L_j} n^{\frac{1}{2}(V_j - V_i)}.$$

For our hypothetical data set, we set the sample size n to a reasonable value of 240, and we calculated the Schwarz weight for each model (Table 1, last column). As can be seen from a comparison of the Schwarz weights and the Akaike weights, the BIC favors simple models (i.e., those with fewer parameters) to a greater extent than does the AIC. Note that in contrast to the AIC, the BIC prefers Model A1 over the more complex Model A2. Both the Akaike weights and the Schwarz weights, however, prefer Model A over Model B. According to the Schwarz weights, Model A (i.e., both A1 and A2) is

$$\frac{w_{\mathrm{A1}}(\mathrm{BIC}) + w_{\mathrm{A2}}(\mathrm{BIC})}{w_{\mathrm{B1}}(\mathrm{BIC}) + w_{\mathrm{B2}}(\mathrm{BIC})} \approx 19.45$$

more likely than Model B (normalized probability .95).[3]

Conclusion

Akaike weights are easy to compute from the raw AIC values and provide a straightforward interpretation as the probabilities of each model's being the best model in an AIC sense (i.e., the model that has the smallest Kullback–Leibler distance, given the data and the set of candidate models). The use of Akaike weights gives the reader greater insight into the relative merits of the competing models. In addition, Akaike weights quantify conclusions based on AIC analyses by specifying the amount of statistical confidence for the model with the lowest AIC value.[4] Given these considerable advantages, we believe that it is in many circumstances very useful to supplement the standard results of AIC model comparison analysis with presentation of Akaike weights.

REFERENCES

Aitchison, J., & Dunsmore, I. R. (1975). *Statistical prediction analysis*. Cambridge: Cambridge University Press.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267-281). Budapest: Akademiai Kiado.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

Akaike, H. (1978). On the likelihood of a time series model. *The Statistician*, **27**, 217-235.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, **66**, 237-242.

Akaike, H. (1983). Information measures and model selection. *Proceedings of the 44th Session of the International Statistical Institute*, **1**, 277-291.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317-332.

Boltzmann, L. (1877). Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht. *Wiener Berichte*, **76**, 373-435.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, **52**, 345-370.

Buckland, S. T., Burnham, K. P., & Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics*, **53**, 603-618.

Burnham, K. P., & Anderson, D. R. (2001). Kullback–Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, **28**, 111-119.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York: Springer-Verlag.

Efron, B. (1986). How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, **81**, 416-470.

Eid, M., & Langeheine, R. (1999). Latent class models. *Psychological Methods*, **4**, 100-116.

Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice*. Newbury Park, CA: Sage.

Golan, A. (Ed.) (2002). Information and entropy econometrics [Special issue]. *Journal of Econometrics*, **107** (1-2).

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer-Verlag.

Hurvich, C. M., & Tsai, C.-L. (1995). Model selection for extended quasi-likelihood models in small samples. *Biometrics*, **51**, 1077-1084.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Hillsdale, NJ: Erlbaum.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, **22**, 79-86.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, **53**, 49-70.

Maddox, W. T., & Bohil, C. J. (2001). Feedback effects on cost–benefit learning in perceptual categorization. *Memory & Cognition*, **29**, 598-615.

McQuarrie, A. D. R., & Tsai, C.-L. (1998). *Regression and time series model selection*. Singapore: World Scientific.

Myung, I. J., Forster, M. R., & Browne, M. W. (Eds.) (2000). Model selection [Special issue]. *Journal of Mathematical Psychology*, **44**(1).

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, 4, 79-95.

Nosofsky, R. M. (1998). Selective attention and the formation of linear decision boundaries: Reply to Maddox and Ashby (1998). *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 322-339.

Parzen, E., Tanabe, K., & Kitagawa, G. (Eds.) (1998). *Selected papers of Hirotugu Akaike*. New York: Springer-Verlag.

Ploeger, A., van der Maas, H. L. J., & Hartelman, P. A. I. (2002). Stochastic catastrophe analysis of switches in the perception of apparent motion. *Psychonomic Bulletin & Review*, **9**, 26-42.

Raijmakers, M. E. J., Dolan, C. V., & Molenaar, P. C. M. (2001). Finite mixture distribution models of simple discrimination learning. *Memory & Cognition*, **29**, 659-677.

Royall, R. M. (1997). *Statistical evidence: A likelihood paradigm*. London: Chapman & Hall.

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). *Akaike information criterion statistics*. Dordrecht: Reidel.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

Smith, P. L. (1998). Attention and luminance detection: A quantitative analysis. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 105-133.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B*, **64**, 1-34.

Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics, Theory & Methods*, **A7**, 13-26.

Takane, Y., & Bozdogan, H. (Eds.) (1987). AIC model selection [Special issue]. *Psychometrika*, **52**(3).

Thomas, R. D. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Perception & Psychophysics*, **63**, 625-650.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, **44**, 92-107.

NOTES

1. For recent discussions on the AIC and alternative model selection methods (which may be preferable to the AIC under certain circumstances, see Myung & Pitt, 1997), the interested reader is referred to the special issue of the *Journal of Mathematical Psychology* (Myung et al., 2000) on Model Selection.

2. The same Akaike weights are obtained from Equation 4 regardless of whether the raw AIC values or the AIC differences are used. The use of differences is encouraged because these are easier to understand in a table and because use of raw AIC values can result in extreme values when the exponential scaling is applied.

3. We should like to point out that a recently developed Bayesian model selection method, the deviance information criterion or DIC (Spiegelhalter, Best, Carlin, & van der Linde, 2002) is approximately equivalent to the AIC when the impact of prior information is negligible.

4. Akaike weights find further application in model-based parameter averaging (e.g., Buckland et al., 1997) and a Bayesian extension of the AIC procedure using priors (e.g., Akaike, 1979, 1983). A discussion of these issues is beyond the scope of this note.

(Manuscript received May 20, 2002;
revision accepted for publication November 25, 2002.)