

## Seleccção de modelos em amostragem por distâncias

Fátima Morgado

*Departamento de Matemática, Universidade de Évora - romaofatima@sapo.pt*

Russell Alpizar-Jara

*Departamento de Matemática, Universidade de Évora - alpizar@uevora.pt*

José Saias

*Departamento de Informática, Universidade de Évora - jsaias@di.uevora.pt*

**Resumo:** A selecção de um modelo que se ajuste adequadamente a um conjunto de dados é uma parte fulcral no processo da inferência estatística. O critério mais comum para seleccionar uma função de detectabilidade, que se ajuste a um conjunto de dados de distâncias, denomina-se critério de informação de Akaike (AIC). Para avaliar a performance da rotina da selecção de modelos implementada no programa DISTANCE (Thomas *et al.*, 2006) realizaram-se simulações Monte Carlo para gerar dados de duas funções densidade de probabilidade frequentemente usadas. Aplicando esta metodologia, observou-se um enviesamento substancial na selecção do modelo, sendo os estimadores dos parâmetros baseados no modelo seleccionado frequentemente enviesados. Utilizaram-se duas abordagens na estimação dos parâmetros: selecção de modelos tradicional (ST) e ponderação de modelos (PM). Para ilustrar a metodologia da selecção de modelos usaram-se dados provenientes de um estudo controlado.

**Palavras-chave:** Amostragem por distâncias, critério de informação de Akaike, ponderação de modelos, selecção de modelos.

**Abstract:** The selection of a model that adequately fits a data set is an important part in the statistical inference process. The most commonly used criterion to select a detectability function that fits a distance sampling data set it is known as the Akaike information criterion (AIC). In order to evaluate the performance of the model selection routine implemented in software DISTANCE (Thomas *et al.*, 2006), we carried out Monte Carlo simulations generating data from two frequently used probability density functions. We found substantial model selection biases when using this methodology. We further observed that parameter estimators based on the selected model were often biased. Parameter estimation was conducted under two approaches: traditional model selection (ST) and model averaging (PM). We used data from a controlled field experiment to illustrate the model selection methodology.

**Keywords:** Akaike information criterion, distance sampling, model averaging, model selection.