

# An Introduction to Bayesian Statistics Without Using Equations

Tomoharu Eguchi

Fish Marine Turtle Research Program, Protected Resources Division  
National Marine Fisheries Service, La Jolla, CA 92037 USA (E-mail: Tomo.Eguchi@noaa.gov)

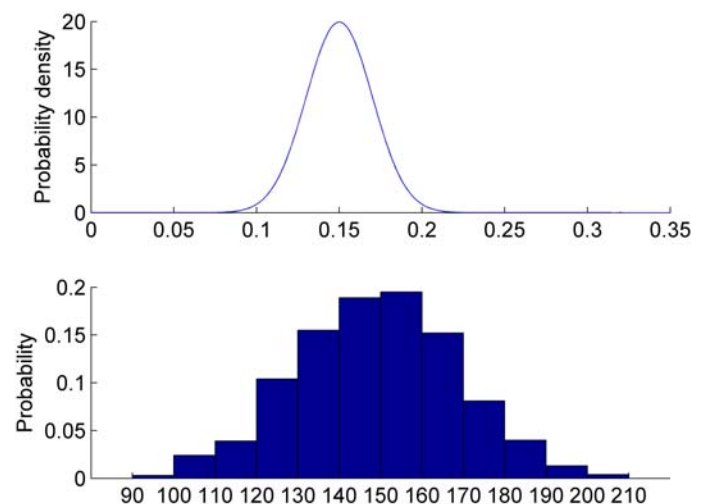
Recently, Bayesian statistics have become a common analytical tool in ecological and biological studies. In 1996, the journal *Ecological Applications* had a special section on Bayesian Inference (Vol. 6, Issue 3). In the issue, one introductory note and seven articles were published, of which one was a technical introduction (Ellison 1996), four were applications (Ludwig 1996, Taylor *et al.* 1996, Ver Hoef 1996, Wolfson *et al.* 1996), and two were discussions of the Bayesian inferential process (Dennis 1996, Edwards 1996). The series of papers clearly describes how Bayesian statistics can be used in ecological studies, as well as the opposite viewpoint expressed by Dennis (1996). Others also have contributed materials to introduce the Bayesian concept to biologists and ecologists (Clark 2007, Reckhow 1990, Wade 2000, Wade 2001). To keep this note short and succinct, I refrain from reviewing the differences between Bayesian and traditional statistical approaches. Readers who are not familiar with the differences may consult the paper by Ellison (1996) or other Bayesian textbooks (e.g., Bernardo & Smith 1993, Clark 2007, Gelman *et al.* 2004, Lee 1997).

The rapid spread of the Bayesian approach among some ecological statisticians (or statistical ecologists) in the past few years has resulted in a bimodal trend of data analysis as some “traditional” ecologists, who are not well versed in mathematics, remain in the comfort zone of the traditional approaches, such as hypothesis testing, learned in introductory statistics classes. The mystery of Bayesian statistics is exacerbated by the lack of user-friendly software, which would facilitate the teaching of the principles to a general audience. Furthermore, publications with Bayesian statistics often contain mathematical equations and the terminology, which are not thoroughly explained in many instances. As Mrosovsky (2006) pointed out, methods of a scientific study should be explicitly stated in a publication so that readers can judge the validity of the reported findings. Simultaneously, in order for the readers to be a critical audience, they must learn the language and concepts of Bayesian statistics.

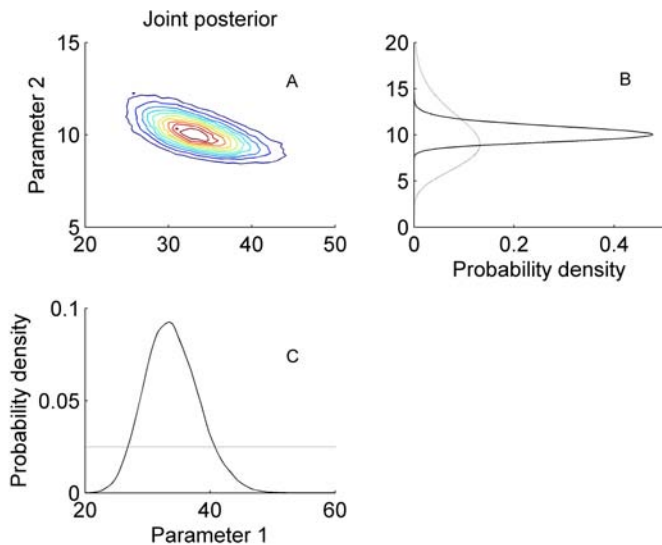
The goals of this paper are (1) to introduce the concepts of Bayesian statistics and (2) to point out some critical features of Bayesian statistics such that publications with Bayesian statistics can be read critically by a wider audience. Although many papers and books have been written to introduce Bayesian statistics in the past, all of them contained equations. In this short note, the basic concepts of Bayesian statistics are presented and explained in the absence of equations. This note is neither a comprehensive summary of Bayesian statistics in ecological studies nor a how-to guide to be used for data analysis. Rather, it explains the basic Bayesian principles in plain English. For those who seek a more comprehensive understanding of Bayesian statistics, consult the cited papers and reputable textbooks. Finally, coursework in Bayesian statistics classes ensures an understanding of and comfort with the Bayesian analytical approach in ecological studies.

**Concepts of Bayesian statistics.** The basic concepts of Bayesian statistics has been explained eloquently by some prominent ecologists, e.g., Dixon & Ellison (1996), Reckhow (1990), and Wade (2000). In a nutshell, Bayesian statistical methods are used to compute a probability distribution of parameters in a statistical model, using data and the previous knowledge about the parameters. Parameters are the unknown quantity of interest in a study. For example, if you are interested in determining the trend of a population, the population growth rate is the parameter of interest. For another example, to assess the size of a population, the abundance is the parameter of interest. In a simple linear regression analysis, the slope and intercept of the regression line are the two parameters of interest.

In a non-Bayesian statistical analysis, we have learned to treat a parameter as a fixed quantity without a distribution. We also have learned not to use prior knowledge about the parameter in the analysis. In Bayesian statistics, we consider a range of parameter values, where the possibility of any parameter value given the observed data is expressed with a probability, whereas the uncertainty about the parameter value is described with a “width” of the distribution. Some values are more likely than others, which is depicted by the difference in the height of the distribution (Figure 1). If all parameter values are equally likely, the distribution has an equal height over all possible parameter values. The curvature



**Figure 1.** Examples of continuous (top) and discrete (bottom) probability distributions. The continuous distribution is a normal distribution with mean 0.15 and standard deviation 0.02. The discrete distribution was created from 1000 random numbers generated from a normal distribution with mean 150 and standard deviation 20. The 1000 random numbers were grouped into 12 equal-width bins.



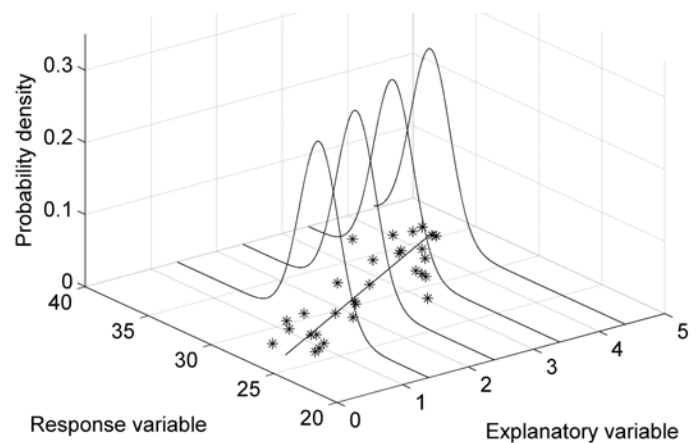
**Figure 2.** An example of numerically obtained joint and marginal distributions of a posterior distribution with two parameters. The top left figure (A) is a contour plot of random samples from the joint posterior distribution, where the height of the distribution is described by the contour lines. The top right figure (B) is the marginal distribution of one parameter (Parameter 2). This can be seen by looking at the joint posterior (A) from the y-axis in (A), ignoring the x-axis. The dashed line indicates the prior distribution for the parameter (informative prior). The bottom left figure (C) is the marginal distribution of the other parameter (Parameter 1). This can be seen by looking at the joint posterior (A) from the x-axis, ignoring the y-axis. The dashed line is the prior distribution (flat prior).

or shape of the distribution is defined by the functional form, i.e., a mathematical equation of the probability distribution, which is governed by one or more parameters, e.g., mean and variance.

At the end of a Bayesian analysis, a probability distribution of parameters, called the **posterior distribution** (or simply posterior), is obtained. This distribution is a result of the combination of the prior knowledge about the parameters (called the **prior distribution** or prior), newly collected data, and a statistical model. The statistical model describes, in a mathematical form, the relationship between the parameters and the data. When there are multiple parameters in an analysis, the posterior distribution of the parameters is called the **joint posterior distribution** (or joint posterior). This part of the analysis also informs about the relationship between the parameters, i.e., correlations between parameters. To make inference on just one parameter, we disregard the other parameters from the posterior distribution (Figure 2). This process is completed by the method of integration. Imagine a two-parameter situation as in Figure 2. If you tilt and rotate the joint posterior distribution (looks like a hill) to such an angle that you can see only one axis, you have just integrated the joint posterior along the other axis. Such distribution is called a **marginal posterior distribution** (or marginal). For a two-parameter model, a marginal posterior distribution can be visualized by changing the angle of the posterior distribution (Figure 2). For a higher dimension, analytical or numerical integration is necessary to obtain a marginal distribution.

**Statistical model:** A statistical model describes the relationship between the parameters and data. It summarizes the stochastic process which produced the data. It is also called a **likelihood function**, because it is used to compute which values of the parameters of the model are most likely to have produced the data we have observed. For example, you may conduct a capture-mark-recapture study to estimate the abundance of a population. The data, consequently, are series of capture/non-capture data for identified individuals. Such data need to be linked to your parameters, such as abundance, capture probabilities, and survival rates. Often, statistical models already exist for standard sampling protocols, such as mark-recapture. Make sure to follow the required sampling protocol so that assumptions in the chosen statistical model are met. With non-standard sampling, especially opportunistically collected data, more creativity may be necessary to build your own statistical model(s). The old adage about the importance of proper sampling to obtain valid results, “garbage in, garbage out” (Krebs 1989, p. 9), applies also to the Bayesian approach. Bayesian statistics, or any other analytical tools, should not be considered as a replacement for well-designed studies. Regardless of the use of Bayesian statistics, modeling is not a replacement for well-thought sampling and experimental designs and meticulous data collection practices. Statistical modeling does not and cannot save a poorly designed and executed study.

Although they may not be explicitly stated, statistical models are used in many analyses. For example, the normal distribution is used as its statistical model in a simple linear regression analysis. It is assumed that the response (or dependent) variable is distributed normally with a mean and variance, given a specific value of the predictor (or independent) variable. The implicit assumption is that the mean changes linearly with the independent variable, whereas the variance remains constant across all observed values of the independent variable (Figure 3).

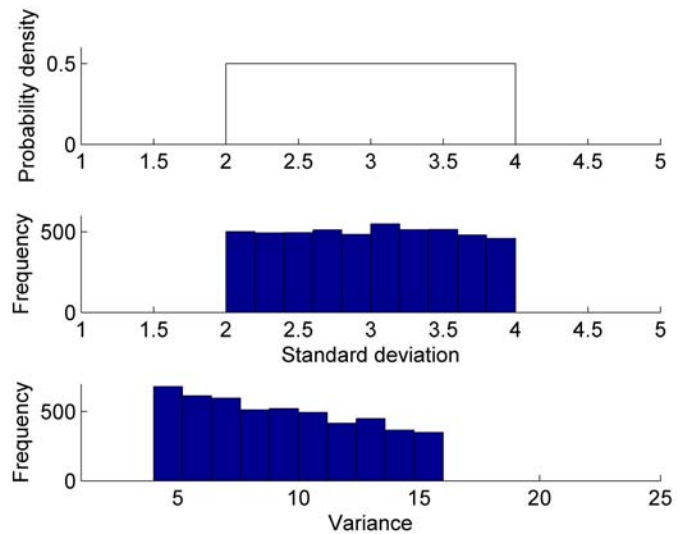


**Figure 3.** A schematic diagram of a linear regression analysis, where data ( $n = 30$ , indicated by \*) were generated from the linear function with the slope = 2.5, intercept = 25.0, and standard deviation = 1.5. The estimated parameters using a linear regression were slope = 2.45, intercept = 25.3, and standard deviation = 1.48. The estimated regression line is shown as a straight line, whereas four normal distributions indicate the assumed normal distributions of data at four values of the explanatory variable (1.4, 2.4, 3.4, and 4.4)

**Previous knowledge about parameters:** The Bayesian philosophy about the distributions of parameters allows us to use the previous knowledge of the parameters to construct a distribution before collecting and analyzing data. The previous knowledge is updated with new data and a likelihood function via Bayes' theorem, which is a well-known theorem in mathematical statistics. Bayes' theorem guarantees that the probability of a specific value of the parameter is proportional to the product of the probability of the parameter value before seeing and analyzing the data and the probability of obtaining the observed data if the parameter value is true. The probability distribution of the parameter prior to analyzing your current data is called the prior distribution. In a Bayesian statistical analysis, each parameter requires a prior distribution. Consequently, you will have multiple prior distributions in multi-parameter situations. Because a prior distribution can be selected arbitrarily, this term of the Bayesian approach is a contentious issue between Bayesians and anti-Bayesians. Anti-Bayesians point out the potential subjectivity of prior distributions (e.g., Dennis 1996, Dennis 2004). In other words, prior distributions may be very different among researchers, even if the same parameters, likelihood function, and data are used in an analysis. The anti-Bayesians have made a good argument that one's belief should not influence a scientific study. Many Bayesian practitioners would agree with the idea of not incorporating someone's belief in a data analysis. Avoiding one's belief in a scientific study, however, is not limited to Bayesian statistics. Any scientist should avoid being influenced by his/her belief when data are analyzed and results are interpreted.

How can we build an objective prior distribution? Some have used **non-informative prior distributions**. A non-informative prior distribution represents the equal probability of all possible parameter values (Figure 4). The phrase 'non-informative' is not really an appropriate term because a 'non-informative' prior distribution provides information about the equal probability of all parameter values within a certain range. This terminology, however, is conventional. We must be cautious that a non-informative (or flat) prior distribution may be not so non-informative when the parameter is transformed. For example, you may decide to set a non-informative prior distribution on the standard deviation. We know that the square of the standard deviation is the variance, which should provide the same information about the variability as the standard deviation. However, when you take the square of a non-informative (flat) distribution, you obtain a "not-so-flat" distribution (Figure 4). The original intention of "non-informativeness" on the standard deviation disappears when it was transformed into variance, even though these two quantities provide the same information. Consequently, the use of a non-informative prior distribution is not the cure-all answer to providing equal probability to all possible values of a parameter.

One way to avoid this problem is to build a **hierarchical model**, in which each of the parameters is assumed to come from another distribution, called a **hyper-distribution**. We then set flat prior distributions for the parameters of the **hyper-distributions (hyper-parameters)**. Because these hyper-distributions reside in the abstract parameter space, which you do not observe directly, the flatness of the distribution may be justified. The hierarchical approach is becoming more common because of its versatility to a wide variety of data structures. The hierarchical approach, for example, allows us to easily model mixed effects models, where



**Figure 4.** An example of a uniform distribution and its transformation. The top figure presents the continuous uniform distribution between 2 and 4. The middle figure is a histogram of 5000 random samples drawn from the uniform distribution. The bottom figure is the histogram of the squared transformation of the 5000 random samples from the uniform distribution.

some factors are considered fixed whereas others are considered random factors. For further discussion and examples of these models, which are somewhat advanced topics, see Clark (2007) and Gelman *et al.* (2004).

A prior distribution can be built from the historical data, a pilot study, and other sources. You may also use information from other closely related species and systems to construct prior distributions. For example, you may be interested in estimating the population growth rate of the snapping turtle (*Chelydra serpentina*). Even though you may not have any information about the population growth rate of your study population, a prior distribution can be constructed from the knowledge of the species from elsewhere. Construction of an informative prior distribution from historical and other systems provide an opportunity to search for the relevant information about the parameter of interest.

**Presenting Bayesian data analysis:** In this section, I will discuss how you may present your results when you use a Bayesian analysis. I will highlight the key features of Bayesian statistics which should not be overlooked when presenting results. The same features should be looked for when reading a research paper that contains Bayesian statistical analysis. First, prior distributions for all parameters should be stated clearly. The justifications for using particular prior distributions should also be included. It is not adequate to say 'non-informative prior distributions were used for all parameters.' Similarly, reasons for selecting certain prior distributions should be provided even if flat priors are used. The effects of prior distributions on the posterior should also be determined. In other words, several prior distributions may be used to compare how they affect the shape of posterior distributions and how sensitive the results of the analysis are to the choice of prior distributions. If the shape of the marginal posterior distribution for a parameter is affected by the prior distribution, the data provide little information on the

parameter. In other words, the knowledge about the parameter is not updated without new information. The important point is that no new inference should be made on the parameters whose posterior distributions are very similar to the prior distributions. Perhaps, those parameters should not be in the model to begin with.

Some things to consider about the prior distributions are: (1) Are the limits of each parameter appropriate? For example, the lower bounds of prior distribution for abundance should be zero or greater. (2) Is the shape of each prior distribution defensible? In other words, does the prior distribution adequately describe the uncertainty about the parameter? (3) Are there enough references given to defend the form of each prior distribution? (4) How does the prior distribution affect the posterior distribution? Do the priors affect posteriors? (5) Are the correlations between parameters considered, especially when the priors affect the posterior?

The second component of Bayesian analysis, the likelihood function, should be scrutinized as well. This is the connection between the data and the parameters of interest. Because statistical models are simplifications of the real ecological processes, multiple models are often applicable to the data. Consequently, the fit of each model to the data should be examined and the best model should be used for the inference. Gelman *et al.* (2004) recommend using a simulation approach. Using the posterior distribution(s) and the likelihood function, one can simulate "data" that could have been observed, given estimated parameters. By simulating a large number of possible "datasets," one can determine whether or not the observed real data are within the plausible range of simulated data based on the assumed underlying model. If there is a discrepancy between the real and simulated data, the model probably was not appropriate for the real data.

In a theoretical approach, Spiegelhalter *et al.* (2002) introduced a measure (deviance information criterion or DIC) that can be used to compare multiple models, similar to AIC (Akaike's Information Criterion; Akaike 1974). DIC can be used to select the best model among candidate models. Alternatively, the uncertainty among possible models can be incorporated into an analysis instead of selecting a simple model. Green (1995) introduced a method called the reversible jump Markov chain Monte Carlo (RJMCMC; Green 1995), which can be used to compute posterior probabilities for multiple models. The posterior probabilities of models, then, are included into the inference process of parameters. This, however, is an advanced feature of Bayesian analysis and beyond the scope of this short note. Although not all in the Bayesian framework, very good discussion about model selections in ecological studies can be found in Burnham & Anderson (2002), Clark (2007), and Hilborn & Mangel (1997).

Some issues to consider about likelihood functions are: (1) Are the likelihood functions defensible? As I mentioned previously, more than one statistical model may be appropriate for the data. In this situation, the selection of a particular model should be explained. (2) Did the model fit to the data? (3) Is it necessary to compare multiple models or is a single model sufficient?

The last component of a Bayesian analysis is the posterior distribution. Conclusions of the study depend on the posteriors. The main feature to look for is the difference between the prior and posterior. If they are very similar, the data did not have much information about the parameter. The conclusions drawn in these cases should be treated as such: the data did not provide new

information about the parameters. Correlations among parameters should also be examined. Summary statistics should be scrutinized. For example, when a posterior distribution is skewed, the mean is not a good statistic of the central tendency. The median or mode may be a better choice. Also, a point estimate, let it be mean, median, or mode, should not be presented without some measure of uncertainty, such as a standard deviation or posterior interval, the Bayesian analogue to a confidence interval. The practice of providing an error measurement in a scientific analysis is not unique to Bayesian statistics. It has been pointed out in many textbooks that no ecological estimate should be presented without some measure of errors (e.g., Krebs 1989). When providing the error of a point estimate in a Bayesian analysis, a posterior interval is preferred over standard deviation because it provides an understandable measure of uncertainty in a probability statement. For example, you may provide a 95% posterior interval from a posterior distribution. The limits of this interval can be interpreted with a simple probability statement; the true parameter is between the limits with probability 0.95, given the data, the statistical model, and the prior.

The difference between the definitions of confidence (CI) and posterior intervals (PI) needs an attention. The true definition of a confidence interval is often overlooked. An  $x\%$  CI should be interpreted as the following: "we are  $x\%$  confident that the true value will be between the two limits." Note that this is not a probabilistic statement. On the other hand, an  $x\%$  PI of a parameter may be interpreted as "the true parameter value is in the interval with probability  $x/100$ ." The practical difference between the two intervals, however, may be trivial. Either interval would provide a measure of uncertainty about the estimate of the parameter. You, as a critical reader or analyst, need to know if the result is interpreted correctly, especially when a manuscript is reviewed for publication.

**A few last words:** I hope I have successfully introduced the concepts of Bayesian statistics without using equations. I also hope this brief note has successfully explained the Bayesian analytical method in such a way that you, the reader, are better equipped to approach your own and other's research more critically. I have left out many topics of Bayesian statistics that cannot be covered in this introductory note. The intention of this paper was to provide insight on the basics of Bayesian statistics and encourage a pursuit of deeper understanding such as how to use and manipulate mathematical equations. Even though we, as biologists, tend to avoid equations, they should be considered as a tool or language that serves to improve our research methods and analyses. One line of an equation can provide as much information as a paragraph in English. You will gain clarity by first understanding some of the nuts and bolts of the mathematical statistics that lend themselves to analyses using either Bayesian statistics or other statistical modeling.

The basic concepts of Bayesian statistics are simple. Bayesian statistics take into account your prior knowledge or experience. Bayesian statistics are not magic. If something does not fit well in your mind, it is likely some things are askew. Also, never hesitate to ask questions about the terminology. If you do not understand something in a paper, it may be that methods are not fully thought out or that you require more information to fully comprehend the analytical technique to critically evaluate the findings. To enhance our knowledge in science, we should strive to use the best tools

available for analyzing data. Although some may argue against it (e.g., Dennis 1996), I think the Bayesian approach should be another set of tools in a biologist's analytical toolbox.

*Acknowledgments:* I thank Nicholas Mrosovsky for encouraging me to write this note. I am grateful to Mandy Karch, Megan Ferguson, and Tim Gerrodette who provided me with constructive comments on early versions of the manuscript that improved the presentation tremendously. Marc Girondot and an anonymous reviewer also provided invaluable comments that increased the quality of the manuscript.

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

BERNARDO, J.M. & A.F.M. SMITH. 1993. *Bayesian theory*. John Wiley & Sons, Ltd. New York, NY.

BURNHAM, K.P. & D.R. ANDERSON. 2002. Model selection and multimodel inference – A practical information theoretic approach. Second edition. Springer, New York, NY.

CLARK, J.S. 2007. Models for ecological data. Princeton University Press. Princeton, NJ.

DENNIS, B. 1996. Discussion: Should ecologists become Bayesians? *Ecological Applications* 6:1095-1103.

DENNIS, B. 2004. Statistics and the scientific method in ecology. In: Taper, M.L. & S.R. Lele (Eds). *The Nature of Scientific Evidence*. The University of Chicago Press, Chicago, pp. 327-359.

DIXON, P. & A.M. ELLISON. 1996. Introduction: ecological applications of Bayesian inference. *Ecological Applications* 6:1034-1035.

EDWARDS, D. 1996. Comment: The first data analysis should be journalistic. *Ecological Applications* 6:1090-1094.

ELLISON, A.M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications* 6:1036-1046.

GELMAN, A., J.B. CARLIN, H.S. STERN & D.B. RUBIN. 2004. *Bayesian data analysis*. 2<sup>nd</sup> edition. Chapman and Hall/CRC. Boca Raton, FL. 696pp.

GREEN, P.J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82:711-732.

HILBORN, R. & M. MANGEL. 1997. The ecological detective – Confronting models with data. *Monographs in Population Biology* 28, Princeton University Press, Princeton, NJ. 338pp.

KREBS, C.J. 1989. *Ecological Methodology*. Harper Collins Publisher, New York, NY. 654pp.

LEE, P.M. 1997. *Bayesian statistics – An introduction*. Second edition. John Wiley & Sons, Inc. New York, NY. 344pp.

LUDWIG, D. 1996. Uncertainty and the assessment of extinction probabilities. *Ecological Applications* 6:1067-1076.

MROSOVSKY, N. 2006. The marginalization of methods. *BioScience* 56:790-791.

RECKHOW, K.H. 1990. Bayesian inference in non-replicated ecological studies. *Ecology* 71:2053-2059.

SPIEGELHALTER, D.J., N.G. BEST, B.P. CARLIN & A. VAN DER LINDE. 2002. Bayesian measure of model complexity and fit. *Journal of Royal Statistical Society* 64:583-639

TAYLOR, B.L., P.R. WADE, R.A. STEHN & J. F. COCHRANE. 1996. A Bayesian approach to classification criteria for spectacled eiders. *Ecological Applications* 6:1077-1089.

VER HOEF, J.M. 1996. Parametric empirical Bayes methods for ecological applications. *Ecological Applications* 6:1047-1055.

WADE, P.R. 2000. Bayesian methods in conservation biology. *Conservation Biology* 14: 1308-1316.

WADE, P. R. 2001. The conservation of exploited species in an uncertain world: novel methods and the failure of traditional techniques. In: Mace, G.M., K.H. Redford, J.G. Robinson & J.D. Reynolds (Eds). *Conservation of Exploited Species*. Cambridge University Press, Cambridge, U.K. 544pp.

WOLFSON, L.J., J.B. KADANE & M.J. SMALL. 1996. Bayesian environmental policy decisions: two case studies. *Ecological Applications* 6:1056-1066.

## Epibionts Associated with Green Sea Turtles (*Chelonia mydas*) from Cananéia, Southeast Brazil

Bárbara Oliveira De Loreto<sup>1</sup> & Ana Cristina Vigliar Bondioli<sup>2</sup>

<sup>1</sup>UNESP, Botucatu, São Paulo, Brazil (E-mail: badeloreto@hotmail.com);

<sup>2</sup>Instituto de Pesquisas Cananéia; Museu de Zoologia, Universidade de São Paulo, São Paulo, Brazil

Sea turtles are frequently colonized by epibiotic organisms. This type of association involves all sea turtle species in several developmental life stages. Gathering knowledge on the species composition, distribution patterns and abundance of epibiotic species represents the first stage in understanding the nature of epibiotic relationships (Frazier *et al.* 1991, Miranda & Moreno 2002). It is also possible to obtain information about pre-reproductive migratory routes (Eckert & Eckert 1988) and population distribution and movements through more detailed studies about epibionts (Casale *et al.* 2004).

The majority of the studies on sea turtle epibionts refer to the epibiotic assemblages of adult sea turtles. However, recent studies involving juvenile loggerheads (*Caretta caretta*) (Frick *et al.* 2003a) present new associations that demonstrate the need for further

investigation and research projects concerning sea turtle epibiosis in feeding and development grounds. As a result, we initiated a study of the epibionts associated with a foraging population of juvenile green turtles (*Chelonia mydas*) in southeast Brazil.

Epibionts of fifty juvenile green turtles (curved carapace length = 39.6 cm ± 5.69 SD and curved carapace width = 36.4 cm ± 5.57 SD) from Cananéia, southeast Brazil, were collected during the activities carried out by Projeto Tartarugas – IPeC (Bondioli *et al.* 2005) in 2007, between January and September. Forty were captured by local fishermen that use artisanal fish traps, the “*cercos-fixos*”, where turtles remain alive and unharmed (Nagaoka *et al.* 2005). Ten of them were found dead on local beaches during beach monitoring activities. The biometric data of turtles were collected (Bolten 1999),