

## BAYESIAN UPDATE FOR DESCRIPTIVE STATISTICS IN FISHERIES SCIENCE

Sedat GÜNDOĞDU \* and Mustafa AKAR \*\*

\* Cukurova University, Faculty of Fisheries, Department of Basic Sciences, Adana, Turkey, TR-01330, sgundogdu@cu.edu.tr

\*\* Osmaniye Korkut Ata University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Osmaniye, Turkey, TR-01330, makar@osmaniye.edu.tr

DOI: 10.1515/trser-2015-0027

**KEYWORDS:** Bayes Theory, descriptive statistics, *Upeneus pori*, Iskenderun Bay.

### ABSTRACT

In the present paper we have examined Bayesian update for descriptive statistics for a sample of 730 Por's Goatfish (*Upeneus pori*) (Ben-Tuvia and Golani, 1989), collected from Iskenderun Bay, in the northeast Mediterranean Sea. The computational approach uses the Markov Chain Monte Carlo simulation to draw samples from the posterior distributions of model parameters implementing the simulation in OpenBUGS software. We assigned the results of previous studies as a prior distribution. The posterior distribution for mean length and variance were found to be 11.1 cm and 0.003, while for weight, they were 15.7 g and 0.026. The 95% confidence limits of length and weight were 10.99-11.21 and 15.42-16.05 respectively. The key aspect of this research is that when previous studies are included in the estimation, this significantly reduces the variance and uncertainty, leading to a more sufficient and reliable estimation.

**RÉSUMÉ:** Mise à jour bayésienne pour le descriptif des sciences appliquées à la pêche.

Dans le présent article nous avons examiné les derniers aspects de la statistique descriptive bayésienne pour un échantillon de 730 individus de l'espèce *Upeneus pori* (Ben-Tuvia and Golani, 1989) de la baie de Iskenderun, au nord-est de la Méditerranée. La simulation Monte Carlo Markov Chain a été utilisée pour ploter les distributions postérieures des paramètres qui ont été implémentés en suite dans le logiciel OpenBUGS. Les enregistrements passés ont été assignés en tant que distributions antérieures. Les distributions postérieures de la moyenne et de la variance pour la longueur ont été de 11,1 cm et 0,003, et respectivement de: 15,7 et 0,026, pour le poids. L'intervalle de confiance de 95% pour la longueur a été de 10,99-11,21 et de 15,42-16,05, respectivement, pour le poids. Un résultat-clé a été le fait que l'utilisation des études précédentes pour les estimations baisse la variance et l'incertitude. Ceci rend à son tour l'estimation suffisante et plus fiable.

**REZUMAT:** Aplicații ale metodei bayesiene la statistica descriptivă în pescuitul științific.

În acest articol examinăm ultimele noutăți ale metodei bayesiene pentru statistica descriptivă aplicată la un eșantion de 730 pești din specia *Upeneus pori* (Ben-Tuvia and Golani, 1989) din Golful Iskenderun în nord-estul Mediteranei. S-a utilizat simularea Monte Carlo Markov Chain pentru a extrage eșantioane din distribuțiile posterioare ale parametrilor modelului, apoi a fost implementată în programul OpenBUGS. Distribuția anterioară a fost cea a înregistrărilor precedente. Distribuțiile posterioare pentru media și pentru varianța lungimii au fost 11,1 cm și 0,003, respectiv 15,7 și 0,026 pentru greutate. Intervalul de confidență de 95% este 10.99-11.21 pentru lungime, respectiv 15.42-16.05 pentru greutate. Una din concluziile studiului a fost că includerea în estimare a studiilor precedente reduce semnificativ varianța și incertitudinea, ducând la o estimare suficientă și mai fiabilă.

## INTRODUCTION

The Bayesian inference and decision making has experienced a fast growth over the last thirty years in fisheries modeling. The reason for this is that inference supplies an alternative path to analyze data that is likely to be more conducive to fisheries sciences difficulties than frequentist methods. Fisheries scientists bring together and analyze data with the aim of enhancing nature management. Hence, the analysis of data should arrive to results that are easy to understand and useful for fisheries management decisions (Ellison, 1996; Wade, 2000; Kinas and Andrade, 2007). Generally, fisheries scientists analyze their data in a classical statistical way that tests hypothesis. However, these ways may not explore what the data could possibly tell us about populations. For this reason, Bayesian inference provides an alternate way to analyze data that redresses many of the problems in the frequentist way of calculating descriptive statistics, and most importantly, allows the integration of uncertainty.

Bayesian inferences have similarities to likelihood based methods. However, in practice, they differ from likelihood by weighting the likelihood values by the prior probabilities to acquire posterior probabilities. The methods update estimates by combining the prior probabilities. This is the key difference between the frequentist way and Bayesian methods. The other difference is asking; “What is the probability in observing that for the given data the various hypotheses are true?”. Bayesian asks this question in a different way. Bayesian methods are interested in the probability of the hypotheses being true given the observed data (Wade, 2000; McCarthy, 2007). Therefore Bayesian methods have two main advantages for fisheries scientists. The first one is that Bayesian inferences are easy to present and automatically include the uncertainty of the estimate and probability statements, better representing the state of a population. The second one is that Bayesian theory allows unknown parameters to be included, which allows taking into account the relative consequences of making wrong decisions and the uncertainty from the significant state (Wade, 2000; Mantyniemi, 2006).

In fisheries sciences, calculation of descriptive statistics is most important. If researchers use the Bayesian way to make decisions about fish catchability size of a fish species, they might take into account prior knowledge. However, this cannot be possible in a frequentist way. Therefore, because of the Bayesian way for determining the descriptiveness of a fish species gives more details about the species, Bayesian inferences are more suitable than frequentist methods.

## METHODS

### What is Bayes Theory?

Bayes theory calculates probability of the value of a parameter given the observed data. The data is what is known, the value of the parameter is what is unknown, and Bayesian therefore focus on what the data tell about the parameter (Lindley, 1972; Wade, 2000; Kinas and Andrade, 2007; McCarthy, 2007). Prior distributions are combined with the information obtained from sample data and update to posterior distributions. The problem of this process is called Bayes Theorem stated by Thomas Bayes in 1764 (Lindley, 1972; DeGroot, 1989; Box and Tiao, 1992; Congdon, 2003; Lee, 2004; McCarthy, 2007; Link and Barker, 2010; Savchuk and Tsokos, 2011). The theorem which is given as follows:

$$P(\theta|y) = \frac{P(\theta)P(y|\theta)}{\int P(\theta)P(y|\theta)d\theta} \quad (1)$$

Consequently, equation (1) is often expressed as:

$$P(\theta|y) \propto P(\theta)P(y|\theta) \quad (2)$$

The symbol  $\propto$  being read as “is proportional to”. This also means that, for example when the data and prior have normal distributions, the posterior distribution also has a normal distribution. Here  $P(\theta)$  denotes the prior of model parameters and the term  $P(y|\theta)$  denotes the probability of data given the parameters.

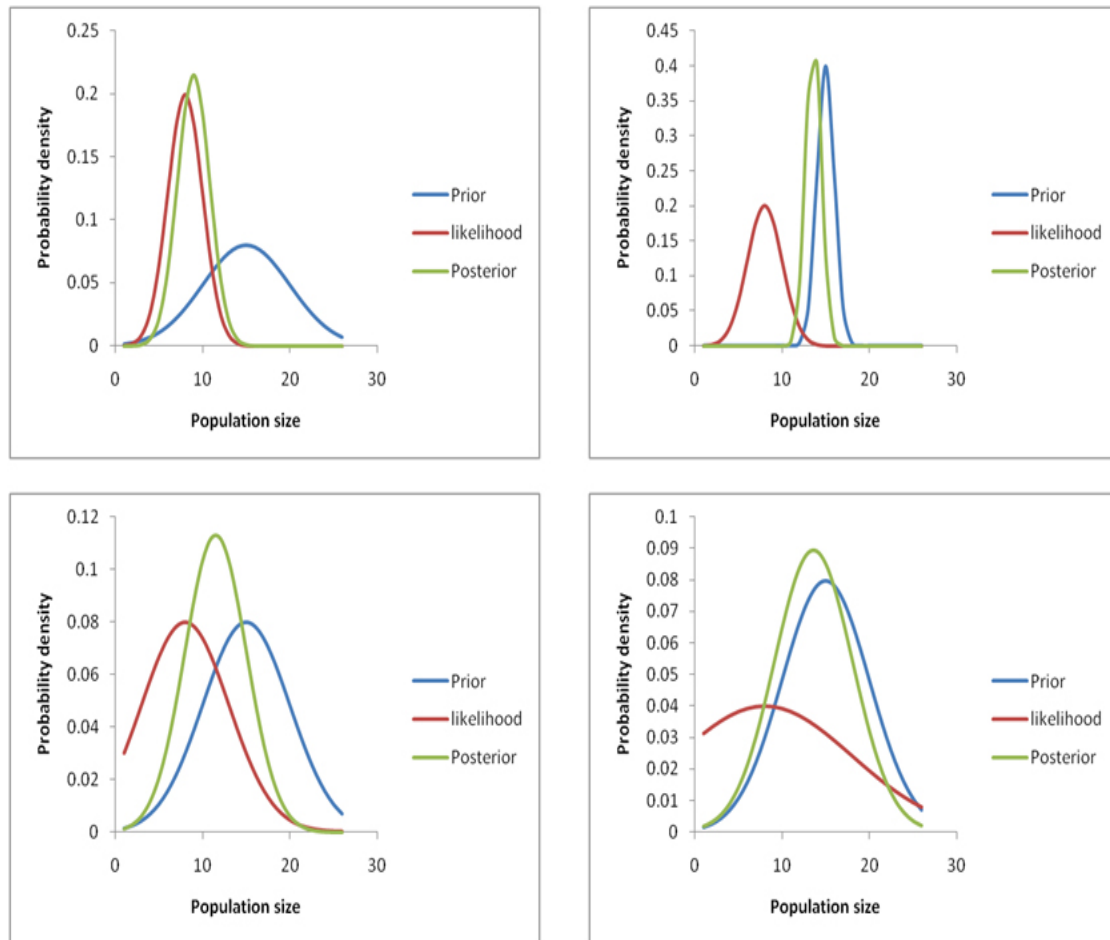


Figure 1: Effect of prior knowledge on posterior distribution  
(<http://mantyniemi.avaruus.net>).

The main idea of Bayesian inference is set at nothing for some by the essential of specifying priors for unknown parameters. The fragrance of subjectivity, connected to chosen priors is the biggest limitation of the widespread use of Bayesian inference by researchers today. Occasionally, Bayesian inference is desirable, because the prior distribution may have an effect on inference (Wade, 2000; Gelman et al., 2003; McCarthy, 2007; Millar, 2002). Prior knowledge of a parameter, for example, fish length, whether from previous knowledge or informed discretion, might be quantified in terms of a probability distribution. Additionally, we should take into account prior knowledge, which affects the posterior distribution (Fig. 1).

Presence of prior information does not remove the potential argument connected to the specification of a prior distribution. There are likely to be challengeable suppositions made concerning the comprehension and exclusion of the existing information, and in its transformation from prior into a prior distribution (Millar, 2002).

### Bayesian Update for Descriptive Statistics

In this study we focus on descriptive statistics of a normal model. In the easiest case, where the data and prior both have normal distributions, Bayesian methods supply an analytical solution for the posterior. The posterior depends on the sample size, mean, and variance of the data (Gelman et al., 2003). Given  $\sigma^2$ , we just have normal data with a normal prior, so the posterior is normal. According to Gelman et al., (2003) and Box and Tiao (1992), analytical solution for the posterior distribution is that:

$$\vartheta_i = \frac{\frac{\theta_{0i} + \bar{x} \frac{n}{S^2}}{\frac{1}{\varphi_{0i}} + \frac{n}{S^2}}}{\frac{1}{\varphi_{0i}} + \frac{n}{S^2}}, \quad \varphi_i = \frac{\frac{\varphi_{0i} S^2}{n}}{\frac{S^2}{n} + \varphi_{0i}}, \quad i = 1, 2 \quad (3)$$

where  $n$  is the size of sample,  $\theta_{0i}$  is the mean of prior,  $\varphi_{0i}$  is the variance of prior,  $\bar{x}$  is the mean of sample,  $S^2$  is the variance of sample,  $\vartheta_i$  is the mean of posterior and,  $\varphi_i$  is the variance of the posterior. For his study subscript  $i$  indicates length (= 1) and weight (= 2).

These two formulas provide useful perception into Bayesian inference. The mean of the posterior is a weighted average of the means of the prior and data. The weights are the precision of the prior ( $1/\varphi_{0i}$ ) and the data ( $n/S^2$ ). The effect of the data and prior on the posterior mean depend on which is more informative (Fig. 1). Therefore there is an approximate 95% chance that the mean of posterior plus or minus 1.96 times the standard deviation of the posterior.

Model entire OpenBUGS program is that:

```
model{
   $\vartheta_i \sim \text{dnorm}(\theta_{0i}, \varphi_{0i})$ 
  stdev <- sd(y[])
  prec <- 1/(stdev * stdev)
  for(i in 1:n){
     $y[i] \sim \text{dnorm}(\vartheta_i, \text{prec})$ 
  }
}
```

### RESULTS AND DISCUSSION

The Bayesian method described above is applied to a real sample of 730 Por's Goatfish (*Upeneus pori*) observations. A summary of this data is provided in figure 2. The computational approach uses the Markov Chain Monte Carlo simulation (Gilks et al., 1995) to draw samples from the posterior distributions of model parameters by implementing the simulation in OpenBUGS software (Spiegelhalter et al., 2012).

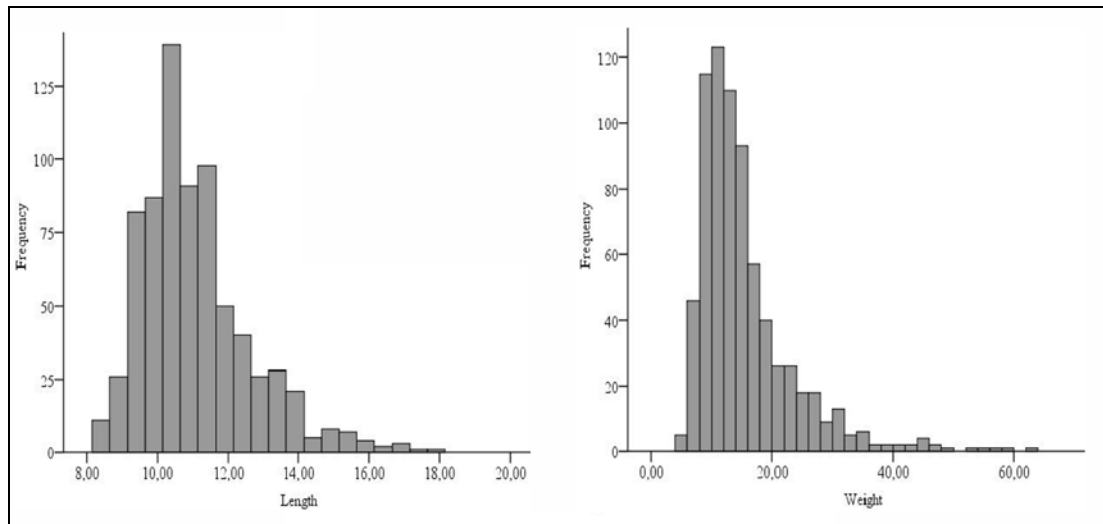


Figure 2: Summary of *Upeneus pori* sample.

Figure 2 shows histograms for total length and total weight. The total length is distributed between eight and 18 cm for the studied species. Most length spans are between nine and 12 cm. The total weight is distributed between five and 70 g. Most weight spans are between 10 and 20 g. Figure 2 shows that the empirical distribution for the lengths of the most spanned subjects (between nine and 12 cm) and for the weights of the most spanned subjects (between 10 and 20 g) are symmetric with light tails.

We assign the following prior distribution for the mean length and weight used in normal model and this information comes from various previous studies as indicated in table 1. These priors are informative, effectively saying that we have information about the model parameters.

Priors are  $\theta_1 \sim N(11.843, 1.714)$  and  $\theta_2 \sim N(15.815, 27.321)$ .

Previous studies shown in table 1, were conducted in the same area with samples used in this study. Here we incorporate the above prior distribution for means, consistent with our belief that the point spread is approximately the mean of the length and weight, but with nonzero variance, indicating some degree of uncertainty. The marginal posterior distribution of the mean and the variance of length and weight, calculated from equation (3), displayed in table 2.

Table 2 also shows that 95% (two-side) credible mean intervals.

The observed average length and weight of the samples were 11.09 cm and 15.53 g. The posterior distribution for the mean and the variance of length is 11.1 cm and 0.003, for weight is 15.7 g and 0.0259. The 95% credible interval of length is (10.99-11.21), for weight is (15.42-16.05) and the most probable fish length is about 11.1 cm and weight is about 15.7 g (Tab. 2). The posterior distribution for the mean weight and length is also informative compared to their priors (Fig. 3).

Table 1: Previous studies which are used for prior knowledge.

Title of study	Autor(s)	Mean		N
		Length	Weight	
Growth and reproduction of Por's Goatfish ( <i>Upeneus pori</i> ) (Ben-Tuvia and Golani, 1989) in Iskenderun Bay, the Eastern Mediterranean	İşmen A.	13.68	-	616
Weight-length relationships for 20 Lessepsian fish species caught by bottom trawl on the coast of Iskenderun Bay (NE Mediterranean Sea, Turkey)	Ergüden et al.	11.98	18.78	210
Distribution of trace elements in the tissues of <i>Upeneus pori</i> and <i>Upeneus mollucensis</i> from the Eastern Coast of Mediterranean, Iskenderun Bay, Turkey	Dural M., Bickici E.	12.68	22.79	20
Age, growth and mortality of <i>Upeneus pori</i> (Ben-Tuvia and Golani, 1989) off the Karatas Coasts of Iskenderun Bay	Çiçek E., Avsar D.	9.83	9.74	247
Length-weight relationships for 31 teleost fishes caught by bottom trawl net in the Babadillimani Bight (NE Mediterranean, Turkey)	Çiçek et al.	10.49	11.95	1225
Evaluation of the demersal fish assemblages of the Northeastern Levant Sea	Ok M.	12.4	-	9271

Table 2: Result of Bayesian analysis.

Parameter	Mean	Std. Deviation	Monte Carlo Error	25%	Median	97.5%	Chain	
							Start	Sample
$\theta_1$	11.1	0.057	0.00017	10.99	11.1	11.21	10000	100001
$\theta_2$	15.7	0.161	0.00049	15.42	15.7	16.05	10000	100001

The posterior marginal probability density graphics are shown in figure 3. As we can expect, both figure 3 and table 2, show us that the median and the mean of both parameters is the same.

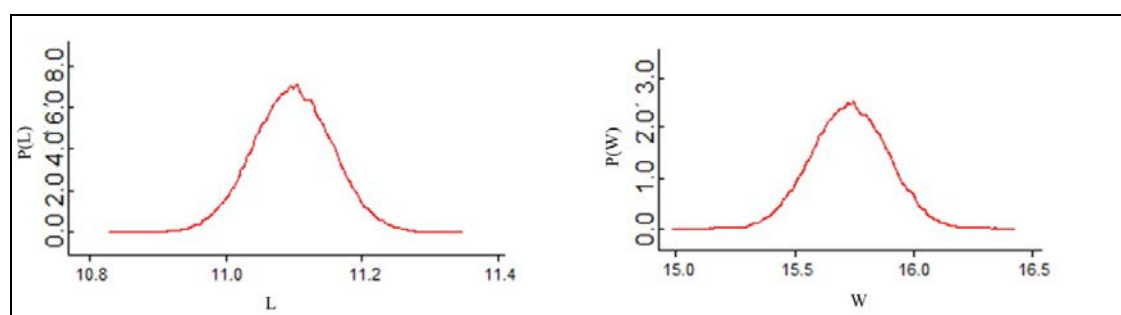


Figure 3: Probability density graph of posterior means.

## CONCLUSIONS

The Bayesian model for the estimation of population was developed by Box and Tiao (1992), Gelman et al. (2003), Lee (2012) and McCarthy (2007). Here, we have taken a step further towards fisheries data. Moreover, this paper attempts to answer a simple question: "Giving my past experience and samples obtained, what should I think about the population mean and variance?" For this idea, that is based on to use the probability concept as a measure of belief, the Bayesian method, is suitable to answer this kind of question. If we compare our posterior results to distribution of length and weight sample, in this study, we should understand how to answer the question (Figs. 2 and 3). On the one hand, the frequentist methods cannot provide a quantitative answer to this question. It is well known that the frequentist approach deals only with the conditional distribution of given observations that the parameter values were known. On the other hand, according to Lee (2004), direct comparison between the result of Bayesian and Frequentist analysis is pointless. Despite the similar values, they are answers to different questions. Descriptive statistics is the most important topic in statistics. In this paper, we examined Bayesian update for descriptive statistics of a random sample with the idea explained above. Of course, our study is not suggesting a new method, however, we tried to show how to calculate descriptive statistics in a Bayesian way.

## ACKNOWLEDGEMENTS

This study was supported by the Scientific Research Project (BAP), Cukurova University, Turkey with grant no. SUF2011BAP7.

# REFERENCES

1. Ben-Tuvia A. and Golani D., 1989 – A new species of goatfish (Mullidae) of the genus *Upeneus* from the Red Sea and the eastern Mediterranean, *Israel Journal of Zoology*, 36, 2, 103-112.
2. Box G. E. P. and Tiao G. C., 1992 – Bayesian Inference in Statistical Analysis, Wiley Classic Library Edition, Jon Wiley and Sons, New York, ISBN-10: 0471574287, 608.
3. Congdon P., 2003 – Applied Bayesian Modelling, John Wiley and Sons Ltd. ISBN: 10-471-48695-7, 478.
4. Dural M. and Bickici E., 2010 – Distribution of trace elements in the tissues of *Upeneus pori* and *Upeneus molucensis* from the Eastern coast of Mediterranean, Iskenderun Bay, Turkey, *Journal of Animal Veterinary Advances*, 9, 1380-1383.
5. Ellison A. M., 1996 – An Introduction to Bayesian Inference for Ecological Research and Environmental Decision Making, *Ecological Applications*, 6, 4, 1036-1046.
6. Gelman A., Carlin J. B., Stern H. S. and Rubin D. B., 2003 – Bayesian Data Analysis, Chapman and Hall/CRC Texts in Statistical Science, ISBN 9781584883883, 690.
7. Gilks W., Richardson S. and Spiegelhalter D., 1995 – Introduction Markov Chain Monte Carlo, in Markov Chain Monte Carlo in Practice, Chapman and Hall, London, UK, ISBN: 9780412055515, 504.
8. De Groot M. H., 1989 – Probability and Statistics, Second Edition, Addison Wesley Publishing Group, Canada, ISBN: 978-0-2011-1366-2, 895
9. İşmen A., 2006 – Growth and Reproduction of Por's Goatfish (*Upeneus pori* Ben-tuvia and Golani, 1989) in Iskenderun bay, the Eastern Mediterranean, *Turkish Journal of Zoology*, 30, 91-98.
10. Kinas P. G. and Andrade H. A., 2007 – Bayesian Statistics for Fishery Stock Assessment and Management: A Synthesis, *Pan American Journal of Aquatic Sciences*, 2, 103-112.
11. Lee P. M., 2012 – Bayesian Statistics an Introduction, Arnold Publication, London, ISBN-10: 1118332571, 486.
12. Lindley D. V., 1972 – Bayesian Statistics a Review, CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Montpelier, ISBN: 978-0-89871-002-1, 83.
13. Link W. A. and Barker R. J., 2010 – Bayesian Inference: with Ecological Applications, Elsevier Academic Publication, California, 339.
14. Mantyniemi S., 2006 – Bayesian Fisheries Stock Assesment: Integrating and Updating Knowledge, Academic Dissertation, Department of Mathematics and Statistics, Faculty of Science, University of Helsinki, ISBN 952-10-3136-0, 24.
15. Millar R. B., 2002 – Refrence Priors for Bayesian Fisheries Models, *Canadian Journal of Aquatic Sciences*, 59, 9, 1492-1502.
16. McCarthy M. A., 2007 – Bayesian Methods for Ecology, Cambridge University Press, London, ISBN: 9780511802454, 296.
17. Ok M. 2012 – Evaluation of the demersal fish assemblages of the Northeastern Levant Sea, Doctoral Dissertation, Middle East Technical University, 227.
18. Savchuk V. P. and Tsokos C. P., 2011 – Bayesian Theory and Methods and Applications, Atlantis Studies in Probability and Statistics, Paris, ISBN: 978-94-91216-13-8, 317.
19. Spiegelhalter D. J., Thomas A., Best N. and Lunn D., 2012 – The OpenBUGS User Manual Version 3.2.2, MRC Biostatistics Unit. Cambridge, UK.
20. Wade P. R., 2000 – Bayesian Methods in Conservation Biology, *Conservation Biology*, 14, 5 1308-1316.