

Robust Bayesian model averaging for the analysis of presence–absence data

Giorgio Corani · Andrea Mignatti

Received: 2 July 2014 / Revised: 8 November 2014 / Published online: 7 February 2015
© Springer Science+Business Media New York 2015

Abstract When developing a species distribution model, usually one tests several competing models such as logistic regressions characterized by different sets of covariates. Yet, there is an exponential number of subsets of covariates to choose from. This generates the problem of model uncertainty. Bayesian model averaging (BMA) is a state-of-the-art approach to deal with model uncertainty. BMA weights the inferences of multiple models. However, the results yielded by BMA depend on the prior probability assigned to the models. Credal model averaging (CMA) extends BMA towards robustness. It substitutes the single prior over the models by a set of priors. The CMA inferences (e.g., posterior probability of inclusion of a covariate, coefficient of a covariate, probability of presence) are intervals. The interval shows the sensitivity of the BMA estimate on the prior over the models. CMA detects the prior-dependent instances, namely cases in which the most probable outcome becomes presence or absence depending on the adopted prior over the models. On such *prior-dependent* instances, BMA behaves almost as a random guesser. The weakness of BMA on the prior-dependent instances is to our knowledge pointed out for the first time in the ecological literature. On the prior-dependent instances CMA avoids random guessing acknowledging undecidability. In this way it stimulates the decision maker to convey

Handling Editor: Bryan F. J. Manly.

G. Corani (✉)

Istituto Dalle Molle di studi sull'Intelligenza Artificiale (IDSIA), Scuola universitaria professionale della Svizzera italiana (SUPSI), Università della Svizzera italiana (USI),
Galleria 2, Manno, Lugano, Switzerland
e-mail: giorgio@idsia.ch

A. Mignatti

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano,
Via Ponzio 34/5, Milan, Italy
e-mail: andrea.mignatti@polimi.it

further information before taking the decision. We provide thorough experiments on different data sets.

Keywords Bayesian model averaging · Credal model averaging · Imprecise probability · Logistic regression · Presence–absence · Robust Bayesian analysis · Species distribution model

1 Introduction

Species distribution models (Guisan and Thuiller 2005) quantify the relationship between the characteristics of the habitat and the presence of a species. They can support decisions regarding, for instance, the conservation (Wilson et al. 2005) or reintroduction of species (see Araùjo and Williams 2000; Olsson and Rogers 2009 and the references therein). They have also been used for landscape planning (Thomson et al. 2007), and for analyzing the risk of invasion of alien species (Goodwin et al. 1999; Peterson 2003).

From the statistical viewpoint, the problem is to predict the outcome of a binary variable (presence or absence) on the basis of different environmental covariates. This is a *classification* problem. For instance one can test different logistic regressions characterized by different sets of covariates and eventually selects among them according to a model selection criterion. However, given k covariates there are 2^k possible subsets of covariates. The choice of the most suitable subset of covariates is thus affected by considerable uncertainty (*model uncertainty*). Performing the analysis on the basis of the single best model overlooks model uncertainty; this can lead to overconfident conclusions.

Bayesian model averaging (BMA) (Hoeting et al. 1999) is a sound solution to model uncertainty. The BMA inferences are obtained by weighting the inferences yielded by different competing models. The weights are constituted by the models' posterior probabilities. *Multi-model inference* (Burnham and Anderson 2002) is another well-known approach to deal with model uncertainty. The tight link between multi-model inference and BMA is discussed by Link and Barker (2006).

The BMA inferences thus account for model uncertainty. BMA has been successfully adopted in the analysis of presence–absence data, showing better predictive performances than the best single model (Wintle et al. 2003; Thomson et al. 2007; St-Louis et al. 2012).

Yet, an open problem is that the BMA inferences depend on the prior probability assigned to the different models. The choice of any prior is necessarily subjective. The likelihood, which summarizes the information contained in the data, does not always overwhelm the prior. In this case different conclusions are drawn depending on the adopted prior.

In the field of *imprecise probability* (Walley 1991) and robust Bayesian analysis (Berger et al. 1994) it has been proposed to compute inferences based on a *set* of prior distributions. The set of prior distributions is called *credal set*. This allows for assessing how much the Bayesian answer changes when the prior varies within the credal set. Inspired by Walley's work, the area of imprecise probability focuses on probabilis-

tic reasoning using sets of probability distributions (Cozman 2000; Destercke et al. 2008).

Credal model averaging (CMA) extends BMA towards robustness; it was firstly developed for probabilistic graphical models (Corani and Zaffalon 2008a). CMA substitutes the prior over the models by a convex set of priors over the models. Such convex set of priors constitutes CMA's credal set. CMA analytically computes the range of posterior answers that would be obtained by letting the prior seamlessly vary within the credal set.

As for the prediction of presence, CMA checks *on each prediction* the sensitivity of the most probable class on the prior over the models. If the most probable class remains the same regardless the prior over the models, the instance is *safe*. If instead absence or presence become more probable depending on the adopted prior over the models, the instance is *prior-dependent*. CMA suspends the judgment on prior-dependent instances, returning both presence and absence as possible outcomes. In this way it acknowledges the undecidability of the instance. Experiments show that BMA usually behaves as a random guesser on the instances recognized as prior-dependent by CMA (Corani and Zaffalon 2008a).

In previous works we have designed CMA algorithms suitable to create ensemble of logistic regression models characterized by different sets of covariates (Corani and Mignatti 2013, 2015). Such papers have been published within the community of imprecise probabilities. In this paper we introduce CMA to the audience of environmental statistics. We extend the analysis of our previous works from an ecological viewpoint. We consider two ecological case studies. The first case study regards the presence data of Alpine marmot (*Marmota marmota*) burrows in an Alpine Valley in Italy. The data set has been collected by one of the authors (A. Mignatti) together with some collaborators, surveying a territory of about 95 ha. As a second case study we consider the greater glider (*Petauroides volans*) in south-eastern Australia, already analyzed by Wintle et al. (2003).

We consider four different priors over the models: the uniform, the beta-binomial (Ley and Steel 2009), the Occam (St-Louis et al. 2012) and the Kullback–Leibler (K–L, Burnham and Anderson 2002; Link and Barker 2006). We show that BMA achieves high accuracy on the safe instances but behaves almost as a random guesser on the prior-dependent instances, regardless the adopted prior over the models. We support this conclusion with additional results regarding synthetic data sets.

We also provide a thorough analysis of parameter estimates and the posterior probability of inclusions of the covariates. CMA yields intervals which show the sensitivity of the posterior estimate on the prior over the models.

We make available the marmot data set and the code from the url www.idsia.ch/~giorgio/cma/cma_code_and_marmot_dset.zip.

The paper is organized as follows: Sect. 2 describes BMA for logistic regression and the most common priors over the models; Sect. 3 describes CMA; Sect. 4 and 5 present the case studies and the experimental results.

2 Logistic regression and Bayesian model averaging

Logistic regression predicts the outcome c_0 or c_1 of the binary variable C . When dealing with species distribution models, c_0 and c_1 respectively denote the absence and the presence of a species. The prediction is based on the observation of one or more *covariates*. Given k covariates ($\{X_1, X_2, \dots, X_k\}$), 2^k different *models* ($\{m_1, m_2, \dots, m_{2^k}\}$) can be defined, each corresponding to a different subset of covariates. We denote by \mathcal{X}_i the set of covariates included by model m_i . The training set is denoted by D . It contains n *instances*, which are joint observations of covariates and class. The posterior probability of model m_i is thus $P(m_i|D)$.

A generic observation of the set of covariates is denoted as $\mathbf{x} = \{x_1, \dots, x_k\}$. Thus $P(c_1|D, \mathbf{x}, m_i)$ is the posterior probability of *presence* given the observation of the covariates \mathbf{x} , estimated by model m_i which has been trained on data D . The logistic regression model is:

$$\eta_{D, \mathbf{x}, m_i} = \log \left(\frac{P(c_1|D, \mathbf{x}, m_i)}{1 - P(c_1|D, \mathbf{x}, m_i)} \right) = \beta_0 + \sum_{x_l \in \mathcal{X}_i} \beta_l x_l$$

where $\eta_{D, \mathbf{x}, m_i}$ is the *logit* of the posterior probability of presence; x_l and β_l are respectively the observation and the coefficient of l -th covariate. The summation extends over the set of covariates \mathcal{X}_i which are included by model m_i .

Our algorithms extend those of the BMA package¹ for R, which estimates the model parameters via maximum likelihood. The same package has been used for previous applications of BMA in ecology (Wintle et al. 2003; St-Louis et al. 2012). We denote by the hat the maximum likelihood estimates; for instance, $\hat{\beta}_0, \hat{\beta}_1$ etc. It is possible to perform Bayesian inference starting from maximum likelihood estimates of the parameters. In particular the posterior distribution of parameter β_j given model m_i is Clyde (2000), Wintle et al. (2003):

$$Pr(\beta_j|m_i, D) = N(\hat{\beta}_j, I(\hat{\beta}_j)^{-1})$$

where $I(\hat{\beta}_j)$ is the observed Fisher information evaluated at the maximum-likelihood estimate $\hat{\beta}_j$.

BMA combines the inferences of multiple models; the weights of the combination are the posterior probability of the models. The posterior probability of model m_i is:

$$P(m_i|D) = \frac{P(m_i)P(D|m_i)}{\sum_{m_i \in \mathcal{M}} P(m_i)P(D|m_i)} \quad (1)$$

where $P(m_i)$ and $P(D|m_i)$ are respectively the prior probability and the *marginal likelihood* of model m_i . Dealing with logistic regression, \mathcal{M} contains the 2^k models obtained considering all the possible subsets of features. When k is large, the model space becomes huge and the enumeration of all models becomes unfeasible. In this

¹ <http://www.cran.r-project.org/web/packages/BMA/index.html>.

case the posterior model probabilities can be estimated by sampling to model space (Clyde and George 2004).

The marginal likelihood integrates the likelihood with respect to the prior of the model parameters: $P(D|m_i) = \int P(D|m_i, \beta_i)P(\beta_i|m_i)d\beta_i$ where β_i is the set of parameters of model m_i . The marginal likelihood can be conveniently approximated using the Bayesian information criterion (BIC) (Raftery 1995). The BIC of model m_i is

$$BIC_i = -2LL(\hat{\beta}_i) + |\beta_i| \log(n)$$

where $LL(\hat{\beta}_i)$ is the log-likelihood of m_i in correspondence of the maximum likelihood estimate of the parameters and $|\beta_i|$ is the total number of the model parameters.

The marginal likelihood of model m_i is approximated as:

$$P(D|m_i) \approx \frac{\exp(-BIC_i/2)}{\sum_{m_i \in \mathcal{M}} \exp(-BIC_i/2)}$$

This approximation is often adopted to compute BMA (Wintle et al. 2003; Clyde and George 2004; Link and Barker 2006; St-Louis et al. 2012). The error involved by the approximation decreases with $1/\sqrt{n}$ and goes to zero as n tends to infinity (Raftery 1995, p. 132).

The posterior probability of model m_i can then be approximated as:

$$P(m_i|D) \approx \frac{\exp(-BIC_i/2)P(m_i)}{\sum_{m_i \in \mathcal{M}} \exp(-BIC_i/2)P(m_i)} \tag{2}$$

2.1 Inferences

The three most common BMA inferences under the logistic model are: the posterior probability of inclusion of covariate X_j ; the posterior expected value of coefficient β_j , referring to covariate X_j ; the posterior probability of presence given a specific observation of the covariates. The common trait of the BMA inferences is that they are carried out by marginalizing out the model variable (Hoeting et al. 1999). Thus BMA weights the inferences of the single models according to models' posterior probabilities.

The posterior probability of inclusion of covariate X_j is the sum of the posterior probabilities of the models which include X_j :

$$P(\beta_j \neq 0) = \sum_{m_i \in \mathcal{M}} \rho_{ij} P(m_i|D) \tag{3}$$

where the binary variable ρ_{ij} is 1 if model m_i includes the covariate X_j and 0 otherwise.

The posterior expected value of parameter β_j referring to covariate X_j is:

$$E[\beta_j|D] = \frac{\sum_{m_i \in \mathcal{M}} \hat{\beta}_{ij} P(m_i|D)}{\sum_{m_i \in \mathcal{M}} \rho_{ij} P(m_i|D)} \tag{4}$$

where $\hat{\beta}_{ij}$ is the maximum likelihood estimate of β_j within model m_i if model m_i includes the covariate X_j , and 0 otherwise. The summation at the denominator extends only over the models which include X_j .

The posterior probability of presence given the observation \mathbf{x} of the covariates is computed as:

$$P(c_1|D, \mathbf{x}) = \sum_{m_i \in \mathcal{M}} P(c_1|D, \mathbf{x}, m_i)P(m_i|D). \tag{5}$$

2.2 Prior over the models

We present in the following the four most common priors over the models for BMA. A graphical comparison of the priors is given in Fig. 1.

The independent Bernoulli (IB) prior assumes that each covariate is independently included in the model with identical probability θ . Parameter θ is thus the prior *probability of inclusion*. The prior probability of model m_i which contains k_i covariates is Clyde and George (2004):

$$P(m_i) = \theta^{k_i} (1 - \theta)^{k-k_i}. \tag{6}$$

The IB prior requires specifying the value of θ . A common choice is $\theta = 1/2$, which yields the *uniform prior* over the models.

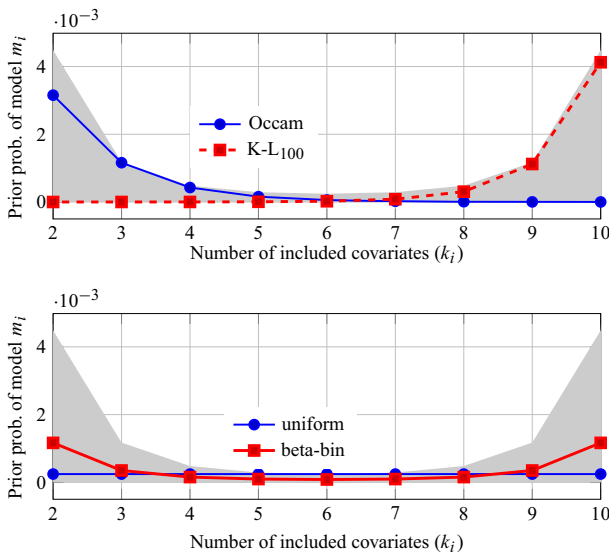


Fig. 1 Prior probability of model m_i containing k_i covariates under different priors. We assume the number of available covariates to be $k = 12$; for the K–L prior we assume the sample size to be $n = 100$. For CMA we set $\theta^{up} = 0.95$ and $\theta^{low} = 0.05$. The gray area shows the interval within which the probability of the model varies according to CMA. We limit the X axis between 2 and 10 to improve readability

Alternatively to the IB prior, the *beta-binomial* (BB) prior has been recommended (Clyde and George 2004; Ley and Steel 2009) since its inferences are less sensitive to the choice of θ . The BB prior treats θ as a random variable. Assuming a uniform prior on θ , the prior probability of model m_i which includes k_i covariates is (Ley and Steel 2009):

$$P(m_i) = \frac{1/(k + 1)}{\binom{k}{k_i}}$$

Under the BB prior each model size (number of covariates included in the model) is equally probable a priori.

The Kullback–Leibler prior (K–L, Burnham and Anderson 2002) yields posterior model probabilities which correspond to the AIC weights used in multi-model inference (Burnham and Anderson 2002; Link and Barker 2006). It is defined as follows:

$$P(m_i) = \frac{\exp[k_i \log(n)/2 - k_i]}{\sum_{m_j \in \mathcal{M}} \exp[k_j \log(n)/2 - k_j]} \tag{7}$$

The K–L prior has been criticized (Link and Barker 2006) since it assigns higher probability to models including a large number of covariates.

Contrarily to the K–L prior, the Occam prior assigns higher probability to models which include few covariates:

$$P(m_i) = \frac{\exp[-k_i]}{\sum_{m_j \in \mathcal{M}} \exp[-k_j]} \tag{8}$$

St-Louis et al. (2012) reports that BMA achieves higher predictive accuracy when trained using the Occam prior rather than the uniform or the K–L prior.

Figure 1 shows how the different priors vary as a function of the number of included covariates.

3 Credal model averaging (CMA)

CMA extends BMA towards robustness, considering a convex set of priors (credal set) over models instead of a single prior. We design a convex credal set of IB priors. While the IB prior is specified by the value of θ (Eqn. 6), the credal set is specified by an upper and a lower value of θ , denoted as θ^{low} and θ^{up} . Such two values are respectively the *lower* and the *upper* prior probability of inclusion. Given an inference (posterior probability of presence, posterior expected value of the parameter, etc), CMA computes the upper and lower value of the inference which would obtained running BMA with IB prior, whose parameter θ seamlessly varies within $[\theta^{low}, \theta^{up}]$.

Often prior knowledge is *not* available; hence, one is *ignorant* a priori. Prior ignorance can be appropriately represented by letting the credal set contain a wide set of prior beliefs. This imply letting θ vary within a large interval. To represent ignorance, we set $\theta^{low} = .05$ and $\theta^{up} = .95$. Under this choice the credal set of CMA accommodates very different prior beliefs about the different models. For instance it includes

the four prior over the models discussed in the previous section. See [Walley \(1996\)](#) for a technical discussion on models of prior-ignorance.

Assuming that θ can vary between 0.05 and 0.95 in case of prior-ignorance, the traditional approach to check the BMA sensitivity on the parameter θ is to re-run BMA with different values of θ , such as for instance: (0.05; 0.35; 0.65; 0.95). This is a coarse discretization of θ . CMA lets theta vary seamlessly between 0.05 and 0.95. It exactly identifies the value of theta which minimizes or maximizes the result of the current inference (posterior probability of presence, posterior expected value of a parameter, etc.). CMA thus addresses sensitivity analysis w.r.t. θ in a continuous fashion, being equivalent to run infinitely many BMAs.

CMA solves optimization problems in order to identify the values of θ which minimizes or maximizes the value of the inference. We formalize in the following the optimization problems for the difference inferences. The solving algorithms are discussed by [Corani and Mignatti \(2015\)](#).

3.1 Covariates: probability of inclusion and parameters

The lower probability of inclusion of X_j is computed as:

$$\begin{aligned}
 P^{low}(\beta_j \neq 0) &= \min_{\theta \in [\theta^{low}, \theta^{up}]} \sum_{m_i \in \mathcal{M}} \rho_{ij} P(m_i | D) \\
 &= \min_{\theta \in [\theta^{low}, \theta^{up}]} \sum_{m_i \in \mathcal{M}} \rho_{ij} \frac{P(D|m_i)P(m_i)}{\sum_{m_j \in \mathcal{M}} \rho_{ij} P(D|m_j)P(m_j)} \\
 &= \min_{\theta \in [\theta^{low}, \theta^{up}]} \sum_{m_i \in \mathcal{M}} \rho_{ij} \frac{P(D|m_i)\theta^{k_i}(1-\theta)^{k-k_i}}{\sum_{m_j \in \mathcal{M}} \rho_{ij} P(D|m_j)\theta^{k_j}(1-\theta)^{k-k_j}} \quad (9)
 \end{aligned}$$

where we recall that ρ_{ij} is 1 if model m_i includes covariate X_j and 0 otherwise. The upper probability is found maximizing rather than minimizing Eq. (9).

The lower expected value of parameter β_j^{low} of covariate X_j , conditioned on its inclusion, is computed as follows:

$$\begin{aligned}
 \beta_j^{low} &= \min_{\theta \in [\theta^{low}, \theta^{up}]} \frac{\sum_{m_i \in \mathcal{M}} \hat{\beta}_{ij} P(m_i | D)}{\sum_{m_i \in \mathcal{M}} \rho_{ij} P(m_i | D)} \\
 &= \min_{\theta \in [\theta^{low}, \theta^{up}]} \sum_{m_i \in \mathcal{M}} \hat{\beta}_{ij} \frac{P(D|m_i)\theta^{k_i}(1-\theta)^{k-k_i}}{\sum_{m_j \in \mathcal{M}} \rho_{ij} P(D|m_j)\theta^{k_j}(1-\theta)^{k-k_j}} \quad (10)
 \end{aligned}$$

where $\hat{\beta}_{ij}$ is the estimate of the expected value of the parameter of covariate X_j within model m_i . The upper parameter β_j^{up} is obtained by maximizing instead of minimizing the same expression.

Given the observation of the covariates, CMA computes an *interval* for the posterior probability of presence. The *lower* probability of presence is computed as follows:

$$\begin{aligned}
 P^{low}(c_1|D, \mathbf{x}) &= \min_{\theta \in [\theta^{low}, \theta^{up}]} \sum_{m_i \in \mathcal{M}} P(c_1|D, \mathbf{x}, m_i)P(m_i|D) \\
 &= \min_{\theta \in [\theta^{low}, \theta^{up}]} \sum_{m_i \in \mathcal{M}} P(c_1|D, \mathbf{x}, m_i) \frac{P(D|m_i)\theta^{k_i}(1-\theta)^{k-k_i}}{\sum_{m_j \in \mathcal{M}} P(D|m_j)\theta^{k_j}(1-\theta)^{k-k_j}}.
 \end{aligned}
 \tag{11}$$

The upper probability of presence $P^{up}(c_1|D, \mathbf{x})$ is obtained by maximizing rather than minimizing expression (11).

3.2 BMA credible intervals versus CMA intervals

It is worth discussing the difference between these two kinds of intervals.

Consider for instance the posterior expected value of parameters. The BMA credible intervals represent uncertainty on the parameter according to a single posterior distribution, derived starting from a single prior. Its interpretation is that e.g. 95 % of the posterior distribution over the parameter lies within the credible interval.

The CMA interval shows how the posterior expected value of the parameter varies when the prior changes (actually, when θ varies between θ^{low} and θ^{up}). It is based on a set of posteriors, derived from a set of priors. The CMA interval thus represents a automatic sensitivity analysis for the expected value computed by BMA.

One could also compute CMA credible intervals as follows. Identify the two values of θ which respectively minimize and maximize the parameter estimate. Compute the credible interval corresponding to the two values of θ and merge them. See [Benavoli and Zaffalon \(2012\)](#) for a discussion of computing credible intervals with imprecise probability. Thus the CMA credible intervals would account for parameter uncertainty, model uncertainty and sensitivity to the prior over the models. To avoid putting excessive material in the paper, we do not further discuss the CMA credible intervals.

3.3 Safe instances and prior-dependent instances

CMA predicts *presence* if both upper *and* lower probability of presence are *greater* than 1/2. CMA predicts *absence* if both upper *and* lower probability of absence are *greater* than 1/2. Such instances are *safe*: the most probable class (presence or absence) remains the same for any value of $\theta \in [\theta^{low}, \theta^{up}]$. An instance is instead *prior-dependent* if the most probable outcome (absence or presence) varies with θ . In this case the probability interval of presence and absence overlap. Both intervals contain the point 1/2.

Previous work dealing with credal classification have shown that traditional classifiers are almost random guessing on the prior-dependent instances ([Corani and Zaffalon 2008a, b](#)). Instead CMA suspends the judgment on the prior-dependent instances, returning both presence and absence as outcomes. In this way, it warns the

decision maker that the instance cannot be safely assigned to any class. A behavior of this type is for instance advocated by Berger et al. (1994, p. 7).

The CMA interval for θ contains the value $\theta = 0.5$, which corresponds to the uniform prior over the models. On the safe instances, the most probable outcome (presence or absence) does not change when θ varies within $[\theta^{low}, \theta^{up}]$. Therefore, on the safe instances CMA predicts the same outcome as BMA trained using the uniform prior.

3.4 Comparing CMA and BMA predictions

We train BMA using four prior over the models: uniform (IB prior with $\theta = 0.5$), beta-binomial, Occam and K–L. For CMA we assume that prior knowledge is not available and thus we set $\theta^{up} = 0.95$ and $\theta^{low} = 0.05$. In this way we represent a condition of ignorance before analyzing the data.

We compare CMA and BMA predictions considering training sets of different dimension. For each sample size we repeat 30 times the procedure of (i) building a training set by randomly down-sampling the original data set; (ii) training BMA and CMA; (iii) assessing the model predictions on the test set, constituted by the instances not included in the training set. We stratify the data set so that each training has the same proportion of presence of the original data set.

The most common indicator of performance for classifiers is the *accuracy*, namely the proportion of correctly classified instances. As already discussed, CMA divides the instances into two groups: the *safe* and the *prior-dependent* ones.

On the safe instances, CMA is as accurate as the BMA trained using the uniform prior (or any other prior contained in the credal set). This accuracy is generally very high. On the prior-dependent instances, CMA acknowledges undecidability returning both presence and absence. On such instances BMA behaves almost as a random guesser, regardless the adopted prior over the models.

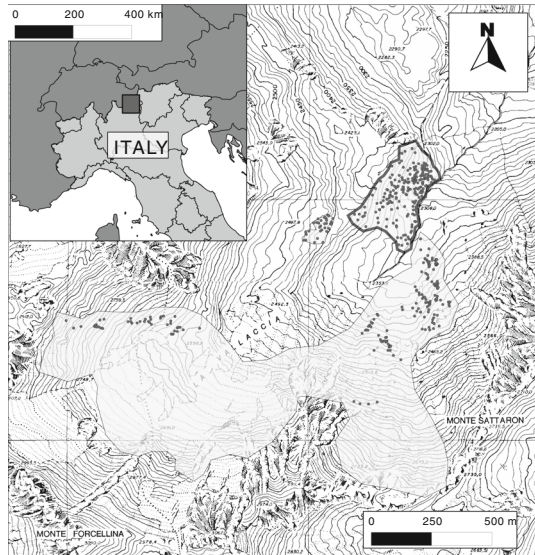
4 Case studies

We analyze the spatial distribution of two different species: the Alpine marmot (*Marmota marmota*) and the greater glider (*Petauroides volans*). One of us (AM) has been collecting with other collaborators the data of marmot, which is therefore our main case study; for the Greater glider we instead re-analyze the data by Wintle et al. (2003). Analyzing both datasets, we deal with a limited number of covariates and, thus, we exhaustively sample the model space.

Alpine marmot. Alpine marmot is a rodent endemic to Europe and mainly distributed in the Alps (Herrero et al. 2008). It inhabits burrow systems, usually on south-facing meadows covered with grass or shrubs (Cantini et al. 1997; Borgo 2003; Lóopez et al. 2010). The altitude range is comprised between 1,000 and 3,000 m a.s.l.; intermediate altitudes ($\sim 1,650$ – $1,900$ m a.s.l., Cantini et al. 1997) are the most suitable ones.

The study area is a high altitude undisturbed Alpine valley in Northwestern Italy (Fig. 2). The altitude of the valley is in the range 2,100–3,100 m a.s.l. . We chose

Fig. 2 Map of the marmot censuses. The censused areas are shown with a transparent mask. The masks of the areas censused in 2010 and 2011 have, respectively, a *thick* and a *thin* border. The burrows are shown as *circles*



a valley at the upper limit of the marmot distribution because the study also aimed at studying the relationship between marmot burrows, summit flora and permafrost. Such further analysis is however out of the scope of this paper.

We collected the presence of marmot burrows during two consecutive summers (2010 and 2011) by exhaustively exploring three census areas within the valley. Census areas were chosen in order to explore habitats characterized by different altitude, slope, aspect, vegetation associations and geological formations. Each burrow was geo-referenced using a GPS receiver, and positional data were post-processed to obtain sub-metric precision. Census areas and position of the burrows are reported in Fig. 2. To develop the species distribution model we divide the area into square cells with 10 m sides. As highlighted by several authors (e.g. see Graf et al. 2005; Guisan et al. 2007; Elith and Leathwick 2009) the grain size should be chosen according to the goals of the study and the precision of the available data. However, the performances of species distribution models change only slightly with the grain size (Guisan et al. 2007). The fine grain adopted for the marmot case study is coherent with the objective of identifying the habitat characteristics preferred by the species for locating burrows (Graf et al. 2005; Schweiger et al. 2012). The dataset contains observations regarding 9429 cells. The fraction of presence (*prevalence*) is 4.6 %.

Covariates were retrieved from a digital terrain model (DTM) with a resolution of 10×10 m, and from a digital map of land use.² The covariates obtained from the DTM are altitude, slope, aspect (the direction in which the slope faces) topographic ruggedness index (Riley et al. 1999), hillshade, curvature and soil cover.

² We used the database DUSAF2.0, which is a product of Regione Lombardia—Infrastruttura per l'Informazione Territoriale. It was retrieved at: <http://www.cartografia.regione.lombardia.it/geoportale>.

The aspect is usually reported as the angle from the North, thus as a periodical variable. To obtain a continuous gradient we split the information contained in the aspect variable into two covariates: *northness* and *eastness*, corresponding respectively to the cosine and the sine of the angle from North. The *northness* is a proxy for the attitude of the slope at receiving the sunlight in the hottest hours of the day: it varies between -1 (southerly exposed slope) and $+1$ (northerly exposed slope). The *eastness* measures the distribution of the sunlight during the day; it varies between -1 (the slope is sunny during the sunset) and $+1$ (the slope is sunny during the sunrise). Hillshade and curvature were calculated using ArcGIS9.2[®]. The categories of soil cover present in the census area are: (i) high altitude alpine meadows without trees or bushes and (ii) debris and lithoidal outcrops without vegetation.

The alpine marmot is a highly mobile species that uses a wide territory for its activities. Its home range is comprised between 1 and 3 ha (Perrin and Berre 1993; Lenti Boero 2003). The decision of digging a burrow in a given cell is therefore made also on the basis of the conditions of the surrounding cells. We therefore averaged the value of each environmental variable over a circular area around each cell (*buffer area*). Using the buffer area, the soil cover variable is redefined as the fraction of buffer area in the category debris and outcrops. We considered buffer areas of 1, 2 or 3 hectares. The results are consistent when different buffer areas are adopted. To avoid redundancies, we report only on the results obtained using the buffer area of 2 ha. As a pre-processing step we removed two highly cross-correlated covariates: topographic ruggedness index (correlation 0.99 with slope) and hillshade (correlation 0.94 with northness).

Greater glider The greater glider is a marsupial glider endemic to eastern Australia. We use the data already analyzed by Wintle et al. (2003), consisting of 405 observations with 74 presence. The prevalence is $74/405 = 0.18$. B.A. Wintle has kindly provided us with the dataset, which includes four covariates: the foliar nutrient index, the mean annual temperature, the solar radiation index and the wetness index. Other covariates used in Wintle et al. (2003) are instead not available. Greater glider surveys were conducted in Southeastern Australia in 1992 and 1994 by Kavanagh and Bamkin (1995).

5 Results

5.1 Alpine Marmot

5.1.1 Probability of inclusion

To show how the posterior probability of inclusion of covariate X_j varies with the prior probability of inclusion θ we devise the following approach. Considering BMA trained under the IB prior, we plug the expression of marginal likelihood (2) and prior probability of the models (6) into the formula of the posterior probability of inclusion (3). In this way we get a function showing how the posterior probability of inclusion of a covariate varies with the prior probability of inclusion θ . The posterior probability of inclusion has a minimum value of 0 for $\theta = 0$ and a maximum value of 1 for $\theta = 1$; it smoothly varies between these two points.

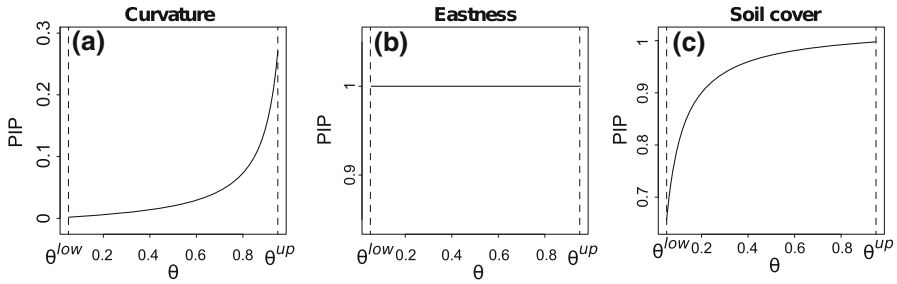


Fig. 3 Posterior probability of inclusion (PIP) as a function of θ for selected covariates. We show the functions within the interval [$\theta^{low} = 0.05, \theta^{up} = 0.95$]

Table 1 *Marmot case study*: posterior probabilities of inclusion of each covariate according to BMA and CMA

Covariate	BMA				CMA
	<i>Uniform</i>	<i>Beta-bin.</i>	<i>Occam</i>	<i>K–L</i>	<i>Interval</i>
altitude	1.00	1.00	1.00	1.00	[1.00, 1.00]
slope	1.00	1.00	1.00	1.00	[1.00, 1.00]
curvature	0.02	0.01	0.01	0.41	[0.00, 0.27]
northness	1.00	1.00	1.00	1.00	[1.00, 1.00]
eastness	1.00	1.00	1.00	1.00	[1.00, 1.00]
soil cover	0.97	0.94	0.93	0.99	[0.66, 0.99]

We show some results in Fig. 3. The posterior probability of inclusion of both curvature and soil cover is quite sensitive to θ despite the huge size of the data set. The posterior probability of inclusion of eastness is instead not sensitive to θ (Fig. 3b). Also the posterior probability of inclusion of altitude, slope and northness are substantially not sensitive to θ .

The CMA interval shows the minimum and the maximum value achieved by the posterior probability of inclusion when θ varies within the interval [$\theta^{low}, \theta^{up}$]. Table 1 reports the estimated probability of inclusion of each covariate, according to BMA (four different priors over the models) and CMA. The estimates are computed using the *whole* data set. The posterior probability of inclusion of four covariates (altitude, slope, eastness and northness) is around one. Such estimates are not sensitive to θ . In these cases the upper and the lower bound of the CMA interval have almost the same value and the CMA interval becomes practically a point.

The CMA intervals regarding curvature and soil cover are instead much larger. The CMA intervals contain the corresponding BMA estimates obtained under the different priors. An exception regards the estimates yielded by the K–L prior. If the data set is large the K–L prior assigns an extremely high probability to the models containing more covariates. This extreme behavior implies that sometimes the estimates obtained under the K–L prior lie slightly outside of the CMA interval.

Table 2 *Marmot case study*: mean and standard deviation of the covariates; expected values of the parameters β_j 's according to BMA and CMA

Covariate	Statistics		BMA				CMA	
	Mean	Std	<i>Unif.</i>	<i>Beta-bin.</i>	<i>Occam</i>	<i>K–L</i>	<i>Lower</i>	<i>Upper</i>
altitude (m)	2,577.14	136.74	−1.05	−1.07	−1.07	−1.04	−1.24	−1.04
slope (°)	24.58	6.63	0.49	0.49	0.49	0.48	0.48	0.49
curvature ($\text{m}^{-1}/100$)	−0.15	0.25	0.06	0.07	0.06	0.06	0.06	0.08
northness	0.49	0.55	−1.38	−1.39	−1.39	−1.37	−1.45	−1.37
eastness	0.44	0.41	−0.55	−0.55	−0.55	−0.56	−0.56	−0.55
soil cover (%)	0.84	0.35	−1.16	−1.16	−1.16	−1.13	−1.16	−1.14

5.1.2 Parameter estimates

We compare the posterior expected values of the parameters yielded by BMA and CMA. The expected values are point estimates for BMA and intervals for CMA. The CMA interval shows the sensitivity of the BMA expected value to the prior over the models. We use the *whole* data set to train the models, standardizing the covariates before estimating the models. The estimates of the parameters are reported in Table 2. Given the huge data set and the limited number of covariates, the coefficients show only limited sensitivity, apart from the case of soil cover and altitude. The estimates obtained under the different priors are almost always included within the CMA interval. There is a single exception, corresponding to the K–L prior on soil-cover. We have already discussed why on very large data sets the K–L estimates might lie outside the CMA interval.

The signs of the parameters are coherent using all priors and mostly confirm what is reported in literature. The probability of burrow presence decreases with the altitude. The most suitable altitude ranges between about 1650 m a.s.l. and 1950 m a.s.l. (Cantini et al. 1997; Borgo 2003) with a maximum of about 3,000 m a.s.l.. Since the valley ranges between 2,200 and 3,100 m a.s.l., at the higher limit of the marmot altitude range, the decrease of the suitability with the altitude is coherent with past studies. The slope positively influences the presence of burrows. The *northness* negatively influences the presence of burrows, thus marmot prefers southerly exposed slopes, as previously reported in several studies (Borgo 2003). Differently from what is reported in literature, in our case the marmot shows a preference for the westerly exposed slopes, since the parameter of *eastness* is negative. This result can be partially due to the valley shape: easterly exposed areas are in fact mainly located at a high elevation, characterized by a low suitability. There is indeed a significant correlation (0.30) between altitude and eastness. The soil cover is measured as the *fraction of outcrops and debris cover* in the buffer area. A high percentage of outcrops and debris cover negatively influences the presence of marmot burrows, showing that the species avoid the use of areas without vegetation, favouring areas covered with alpine meadows, as also reported by many authors (e.g. Borgo 2003; Lóopez et al. 2009).

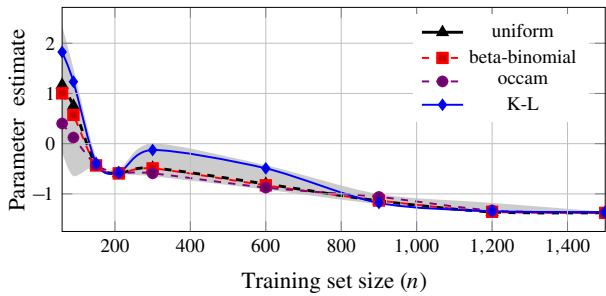


Fig. 4 *Marmot case study*: expected value of the parameter of altitude, as estimated by CMA and BMA

In Fig. 4 we show how the parameter estimates vary with the sample size, for two selected covariates. We start by randomly selecting $n = 60$ instances; we then increase the data set size progressively. At each iteration we add some randomly selected instances to the training set and we correspondingly update the parameter estimates. The parameter estimates obtained with BMA (uniform, beta-binomial, Occam and K–L prior) are consistently contained by the CMA interval. We recall that the CMA interval shows how the expected posterior value of the coefficient varies with the prior over the models.

5.1.3 Prediction of presence

We let the size n of the training set vary between 30 and 1,500; beyond this value the amount of prior-dependent instances is negligible (below 5% of the instances). For each value of n we perform 30 training—test experiments, as described in Sect. 4.

BMA

The fraction of correctly classified instances (accuracy) increases for all priors with n . Larger training sets allow to better estimate both the models’ parameters and the models’ posterior probabilities. Thus larger training sets are associated with better predictions (Fig. 5). In general the Occam prior achieves slightly higher accuracy than the other priors. The difference of performance between priors vanishes on large sample sizes, when the likelihood overwhelms the prior. The BMA accuracy (% of correctly predicted instances) is consistently above 0.9 (Fig. 5).

We analyze the spatial correlation of the BMA predictions through the *Moran’s I* index (Li et al. 2007). This index varies between -1 and $+1$; a zero value indicates a random spatial pattern. Spatial correlation of the BMA residuals is consistently lower than 0.01 under any prior over the models.

CMA

The percentage of indeterminate classifications decreases with the training set dimension; it is 7.5 % for $n = 30$ and 1 % for $n = 450$. BMA is highly accurate on the safe

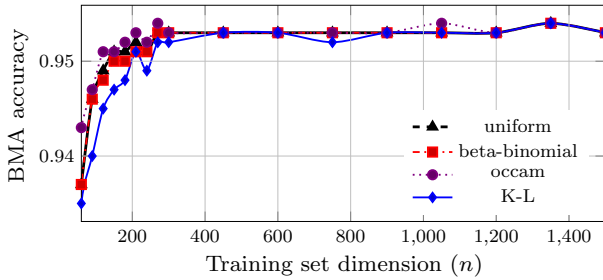


Fig. 5 *Marmot case study*: accuracy of BMA trained under different priors. Each point represents the mean accuracy obtained on the test sets over the 30 experiments

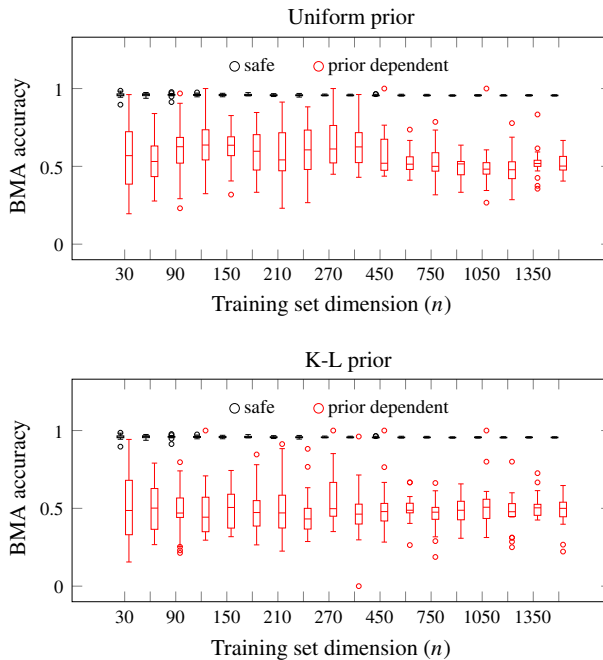


Fig. 6 *Marmot case study*: accuracy of BMA trained under different priors on the safe and prior-dependent instances. For the sake of space, we only show results obtained with the uniform and the K–L prior

instances; conversely, on the prior dependent instances its accuracy drops to a value close to 0.5, behaving almost as a random guesser (Fig. 6). This result is consistently found regardless the prior over the models adopted to train BMA. For instance for a training set of size $n = 300$ all priors yield accuracy of about 96%. On the prior-dependent instances the accuracy varies between 47% (K–L prior) and 65% (Occam prior).

CMA detects a set of instances over which BMA performs poorly because of prior-dependence. CMA warns the decision maker about the undecidability of such instances by returning both presence and absence as possible outcomes.

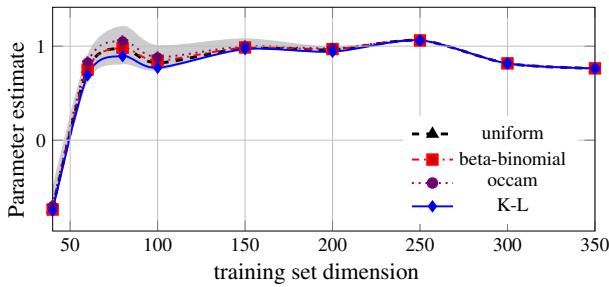


Fig. 7 *Glider case study*: expected value of the *Foliar nutrient* parameter according to BMA and CMA (gray)

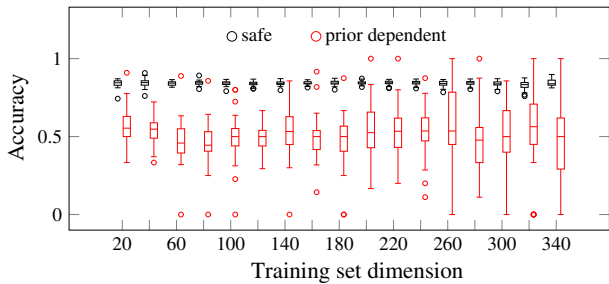


Fig. 8 *Glider case study*: accuracy of BMA calculated under the uniform prior and separately assessed on safe and prior-dependent instances. For each sample size, the boxplot refers to 30 experiments

5.2 Greater glider

5.2.1 Probability of inclusion and parameter estimate

The greater glider data set contains 405 instances. The BMA estimates obtained under all priors, including the K–L, are always contained within the corresponding CMA intervals.

In Fig. 7 we show an experiment regarding the expected value of the parameters. We progressively increase the sample size. Every time we increase the training set, we update the estimate of the parameters. The CMA interval consistently contains the estimates yielded by BMA developed under the four different priors.

5.2.2 Prediction of presence

We downsample the original data set generating training sets of size n comprised between 40 and 240. Beyond this sample size, the amount of indeterminate classifications is below 1%. For each sample size we perform 30 training—test experiments.

Depending on the prior and on the sample size, accuracy varies between 0.81 and 0.83. The Occam prior is slightly better than the others on the accuracy. The fraction of prior-dependent instances decreases from 12 to 7% as the sample size increases.

BMA is highly accurate on the safe instances, but again it behaves almost as a random guesser on the prior-dependent instances (Fig. 8). This result is found for every prior over the models. The results are thus consistent with those of the marmot data set.

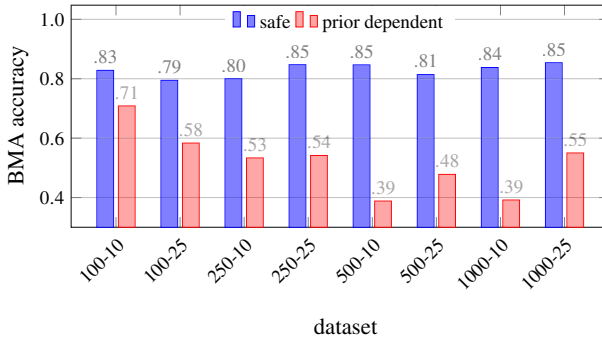


Fig. 9 Accuracy of BMA trained using the uniform prior on the datasets generated using the Friedman function. Each point represents the mean accuracy of the twofold cross-validation. The name of the data sets contains the sample size n (100, 250 or 1,000) and the number k of covariates (10 or 25)

5.3 Further experiments on synthetic data sets

We further test the robustness of the CMA approach on a collection of synthetic datasets. We consider 12 synthetic datasets available from the Weka website.³ They are generated using the so-called Friedman function (Friedman 1991):

$$y = 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5.$$

The data sets differ for the number of instances (100, 250, 500 or 1000) and covariates (5, 10 or 25). The target variable Y only depends on the first five covariates. In data sets with more than five covariates, the additional covariates are independent of Y ; thus they introduce noise making the estimation problem more difficult. We discretize Y setting a cut-point corresponding to its median. In this way we obtain a binary classification problem. Fig. 9 shows the accuracy of BMA on the prior and on the prior-dependent instances on various data sets generated by the Friedman function.

The synthetic datasets do not refer to ecological problems and thus we do not investigate covariates parameters or posterior probability of inclusion. However, we do investigate the robustness of CMA in detecting the prior dependent instances. For each dataset, we performed a twofold cross validation of CMA and of BMA trained under the uniform prior. Averaging over data sets, the accuracy of BMA is respectively 83 and 55 % on the safe and on the prior-dependent instances. The average fraction of prior-dependent instances is about 7 %.

5.4 Credal classification versus reject option

One could try mimicking the CMA behavior by suspending the judgment when BMA is most uncertain, namely when the posterior probability of presence and absence are closer. More formally, an instance is *rejected* (i.e., not classified) if the posterior

³ <http://www.cs.waikato.ac.nz/ml/weka/datasets.html>.

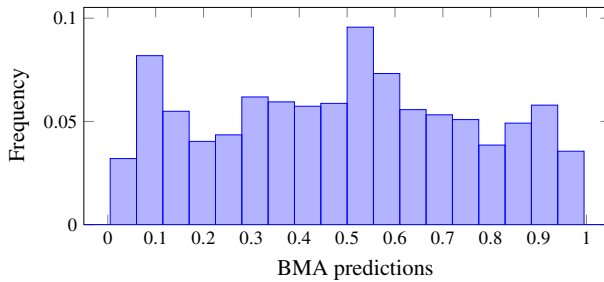


Fig. 10 Distribution of the posterior probability of presence estimated by BMA on the *prior-dependent* instances for the marmot case study. The distributions have been measured over test set predictions. The training set used to train BMA have size $n = 60$

probability of the most probable class falls *below* the threshold p^* . Such approach is known as *rejection option*: see Herbei and Wegkamp (2006) and the references therein.

It is thus worth pointing out the difference between BMA equipped with rejection option and CMA. First, the rejection option requires specifying the parameter p^* . This can be done only after having specified the *rejection cost* which is incurred into when the instance is not classified (Herbei and Wegkamp 2006). Eliciting this cost is far from trivial. Let us however assume that the elicitation has been accomplished, resulting in $p^* = 55\%$. Thus, BMA rejects an instance whenever the posterior probability of presence is comprised between 45 and 55%.

The posterior probabilities of presence estimated by BMA on the prior-dependent instances are *not* strongly concentrated around 0.5; see Fig. 10 for an example regarding the marmot case study. In some cases BMA draws *strong* but misleading conclusions on the prior-dependent instances, e.g. estimating a probability of presence far from 50%, i.e. considerably larger or smaller than 50%. Such conclusions are clearly overconfident, since prior-dependent instances cannot be safely assigned to any class. The problem is that BMA induced under a single prior has no way to detect the prior-dependent instances. Moreover, the sensitivity checks are usually performed by comparing the *average* BMA accuracy obtained under different priors. We are unaware of previous approaches assessing the sensitivity of each single prediction to the specification of the prior.

Let us consider now a case in which a large training set is available and the value of p^* is high. In that case, there would be only few prior-dependent instances, since on large data sets the choice of the prior is less critical. Given the high value of p^* BMA will however reject many instances, most of which not prior-dependent.

Summing up, there is only a partial overlap between the set of prior-dependent instances and set of the instances rejected by the reject option. The point is that CMA and BMA with reject option aim at different goals. CMA automates sensitivity analysis with respect to the choice of the prior over the models. CMA does so for all BMA inferences, which allows to detect the prior-dependent instances. CMA is thus very well suited to small and medium data sets where the choice of prior over the models can critically affect the conclusions. Equipping BMA with rejection option is a valuable approach when the following two conditions are met: a) there is a large training set available, so that analyzing sensitivity with respect to the choice of the prior is not

crucial; b) it is possible to reliably estimate the rejection costs and thus the value of p^* .

6 Conclusions

In species distribution models, data paucity and a possibly large number of covariates are responsible for model uncertainty. BMA is a sound solution to produce inferences which account for model uncertainty. However, the BMA results are sensitive to the prior specified over the models.

CMA extends BMA towards robustness, performing automatic sensitivity analysis regarding the prior over the models. The CMA inferences are intervals that show the range of answers which would be computed by BMA, letting seamlessly vary the prior probability of inclusion within a given interval. This kind of uncertainty is not included when credible intervals are built around a BMA inference; in fact, BMA credible intervals are computed according to a single posterior distribution, obtained starting from a single prior.

CMA automatically detects prior-dependent instances. On such instances, presence or absence is more probable depending on the adopted prior over the models. CMA suspends the judgment on such instances, warning the decision maker about their undecidability. On such instances BMA is almost a random guesser. This results is confirmed on each prior over the models and data set we analyzed. CMA can thus prevent taking decisions in cases whose outcome is in fact not predictable. In such cases the decision maker should try to convey further information before taking a decision. To our knowledge the BMA weakness on prior-dependent instances has not been yet pointed out in the ecological literature.

Acknowledgments The work has been performed during Andrea Mignatti's Ph.D., supported by Fondazione Lombardia per l'Ambiente (project SHARE—Stelvio). We are grateful to B.A. Wintle for providing us with the greater glider data set. We thank M. Gatto, R. Casagrandi, V. Brambilla, M. Cividini and F. Mattioli for the help provided in collecting marmot data. We also thank the anonymous reviewers for their valuable suggestions.

References

- Araújo MB, Williams PH (2000) Selecting areas for species persistence using occurrence data. *Biol Conserv* 96:331–345
- Benavoli A, Zaffalon M (2012) A model of prior ignorance for inferences in the one-parameter exponential family. *J Stat Plan Inference* 142:1960–1979
- Berger JO, Moreno E, Pericchi LR, Bayarri MJ, Bernardo JM, Cano JA, De la Horra J, Martín J, Ríos-Insúa D, Bettrò B et al (1994) An overview of robust Bayesian analysis. *Test* 3:5–124
- Borgo A (2003) Habitat requirements of the Alpine marmot *Marmota marmota* in re-introduction areas of the Eastern Italian Alps. Formulation and validation of habitat suitability models. *Acta Theriologica* 48:557–569
- Burnham KP, Anderson DR (2002) *Model selection and multi-model inference: a practical information-theoretic approach*. Springer, Berlin
- Cantini M, Bianchi C, Bovone N, Preatoni D (1997) Suitability study for the alpine marmot (*Marmota marmota*) reintroduction on the Grigne massif. *Hystrix—Ital J Mammal* 9:65–70
- Clyde M (2000) Model uncertainty and health effect studies for particulate matter. *Environmetrics* 11:745–763

- Clyde M, George EI (2004) Model uncertainty. *Stat Sci* 19:81–94
- Corani G, Mignatti A (2013) Credal model averaging of logistic regression for modeling the distribution of marmot burrows. In: Cozman F, Denoeux T, Destercke S, Seidenfeld T (eds) ISIPTA'13: proceedings of the eighth international symposium on imprecise probability: theories and applications, pp 233–243
- Corani G, Mignatti A (2015) Credal model averaging for classification: representing prior ignorance and expert opinions. *Int J Approx Reason* 56:264–277
- Corani G, Zaffalon M (2008a) Credal model averaging: an extension of Bayesian model averaging to imprecise probabilities. In: Proceedings of the ECML-PKDD 2008 (European conference on machine learning and knowledge discovery in databases), pp 257–271
- Corani G, Zaffalon M (2008b) Learning reliable classifiers from small or incomplete data sets: the naive credal classifier 2. *J Mach Learn Res* 9:581–621
- Cozman FG (2000) Credal networks. *Artif Intell* 120:199–233
- Destercke S, Dubois D, Chojnacki E (2008) Unifying practical uncertainty representations I: generalized p-boxes. *Int J Approx Reason* 49:649–663
- Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst* 40:677–697
- Friedman JH (1991) Multivariate adaptive regression splines. *Ann Stat* 19:1–67
- Goodwin B, McAllister A, Fahrig L (1999) Predicting invasiveness of plant species based on biological information. *Conserv Biol* 13:422–426
- Graf RF, Bollmann K, Suter W, Bugmann H (2005) The importance of spatial scale in Habitat models: capercaillie in the Swiss Alps. *Landsc Ecol* 20:703–717
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecol Lett* 8:993–1009
- Guisan A, Graham CH, Elith J, Huettmann F (2007) Sensitivity of predictive species distribution models to change in grain size. *Divers Distrib* 13:332–340
- Herbei R, Wegkamp MH (2006) Classification with reject option. *Can J Stat* 34:709–721
- Herrero J, Zima J, Coroiu I (2008) *Marmota marmota*. In: IUCN Red List of Threatened Species. Version 2013.1. www.iucnredlist.org. Downloaded 19 July 2013
- Hoeting J, Madigan D, Raftery A, Volinsky C (1999) Bayesian model averaging: a tutorial. *Stat Sci* 44:382–417
- Kavanagh RP, Bamkin KL (1995) Distribution of nocturnal forest birds and mammals in relation to the logging mosaic in south-eastern New South Wales, Australia. *Biol Conserv* 71:41–53
- Lenti Boero D (2003) Long-term dynamics of space and summer resource use in the alpine marmot (*Marmota marmota* L.). *Ethol Ecol Evol* 15:309–327
- Ley E, Steel MF (2009) On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *J Appl Econom* 24:651–674
- Li H, Calder CA, Cressie N (2007) Beyond Moran's I : testing for spatial dependence based on the spatial autoregressive model. *Geogr Anal* 39:357–375
- Link W, Barker R (2006) Model weights and the foundations of multimodel inference. *Ecology* 87:2626–2635
- López B, Figueroa I, Pino J, López A, Potrony D (2009) Potential distribution of the alpine marmot in Southern Pyrenees. *Ethol Ecol Evol* 21:225–235
- López B, Pino J, López A (2010) Explaining the successful introduction of the alpine marmot in the Pyrenees. *Biol Invasions* 12:3205–3217
- Olsson O, Rogers DJ (2009) Predicting the distribution of a suitable habitat for the white stork in Southern Sweden: identifying priority areas for reintroduction and habitat restoration. *Anim Conserv* 12:62–70
- Perrin C, Berre D (1993) Socio-spatial organization and activity distribution of the Alpine Marmot *Marmota marmota*: preliminary results. *Ethology* 93:21–30
- Peterson AT (2003) Predicting the geography of species' invasions via ecological niche modeling. *Q Rev Biol* 78:419–433
- Raftery AE (1995) Bayesian model selection in social research. *Sociol Methodol* 25:111–164
- Riley SJ, DeGloria S, Elliot R (1999) A terrain ruggedness index that quantifies topographic heterogeneity. *Intermt J Sci* 5:23–27
- Schweiger AKA, Nopp-Mayr U, Zohmann M (2012) Small-scale habitat use of black grouse (*Tetrao tetrix* L.) and rock ptarmigan (*Lagopus muta helvetica* Thienemann) in the Austrian Alps. *Eur J Wildl Res* 58:35–45

- St-Louis V, Clayton MK, Pidgeon AM, Radeloff VC (2012) An evaluation of prior influence on the predictive ability of Bayesian model averaging. *Oecologia* 168:719–726
- Thomson JR, Mac Nally R, Fleishman E, Horrocks G (2007) Predicting bird species distributions in reconstructed landscapes. *Conserv Biol* 21:752–766
- Walley P (1991) *Statistical reasoning with imprecise probabilities*. Chapman and Hall London, London
- Walley P (1996) Inferences from multinomial data: learning about a bag of marbles. *J Roy Stat Soc B* 58:3–57
- Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biol Conserv* 122:99–112
- Wintle B, McCarthy M, Volinsky C, Kavanagh R (2003) The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conserv Biol* 17:1579–1590

Giorgio Corani obtained the M.Sc. Degree (1999) and the Ph.D. in Information Technology (2005) from Politecnico di Milano. He is currently researcher at IDSIA (www.idsia.ch), where he is part of the Imprecise Probability Group (<http://ipg.idsia.ch>), which focuses on probabilistic modelling and data mining. His research interests include probabilistic graphical models, data mining, imprecise probability, applied statistics, ecological modelling. He is author of more than 50 publications. For more information, see www.idsia.ch/~giorgio.

Andrea Mignatti obtained the M.Sc. Degree (2010) and the Ph.D. in Information Technology (2013) from Politecnico di Milano. His research interests include ecological modelling and probabilistic models.