International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

Credal Model Averaging: dealing robustly with model uncertainty on small data sets.

A. Mignatti [a], **G. Corani**[b], **and A. Rizzoli**[b]

[a] *Politecnico di Milano, Via Ponzio 34/5, Milano, Italy (mignatti@elet.polimi.it)*

[b] *IDSIA Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Galleria 2, Manno, Switzerland (giorgio@idsia.ch, andrea@idsia.ch)*

Abstract: Datasets of population dynamics are typically characterized by a short temporal extension. In this condition, several alternative models typically achieve close accuracy, though returning quite different predictions (*model uncertainty*). Bayesian model averaging (BMA) addresses this issue by averaging the prediction of the different models, using as weights the posterior probability of the models. However, an open problem of BMA is the choice of the prior probability of the models, which can largely impact on the inferences, especially when data are scarce. We present Credal Model Averaging (CMA), which addresses this problem by simultaneously considering a set of prior probability distributions over the models. This allows to represent very weak prior knowledge about the appropriateness of the different models and also to easily accommodate expert judgments, considering that in many cases the expert is not willing to commit himself to a single prior probability distribution. The predictions generated by CMA are intervals whose lengths shows the sensitivity of the predictions on the choice of the prior over the models.

Keywords: Bayesian model averaging. Linear regression. Alpine ibex. Model uncertainty. Credal sets. Imprecise probability.

1 Introduction

Datasets of population dynamics are usually limited because of the difficulties in the data collection stage; this emphasizes the need for robust methodologies for data analysis. Model selection criteria such as *akaike information criterion* (AIC) or *bayesian information criterion* (BIC) aim at selecting a single model from a set of competing medels [see Burnham and Anderson, 2002, for a review]. Yet, several alternative models typically achieve close accuracy, though returning quite different predictions (*model uncertainty*). Ecological models are in particular affected by high model uncertainty [Conroy et al., 1995], and thus the choice of a single model can be inadequate. The problem is further emphasized by the small data amount which are generally available. Model averaging techniques permits to avoid this choice, producing the predictions by averaging the predictions of many models. BMA uses the posterior probability of the models as model weights. The method requires to specify two types of prior distributions: on the model structure and on the parameter values. Dealing with Bayesian models and scarce data, the choice of the prior can impact on the inferences produced by the models, leading to fragile prior-dependent conclusions.

The specification of the prior over the models is a serious open problem for Bayesian ensembles of models. We address this problem by adopting the paradigm of *imprecise probability*, namely dropping the unique prior in favor of a *set* of priors (prior *credal set*)

Walley [1991]. While a traditional non-informative priors represents a condition of *indifference* between the alternative models, a credal set describes a condition of prior *ignorance*, letting thus vary the prior probability of each model over a wide interval, instead of fixing it to a specific number. We called this novel method *Credal model averaging* (CMA), and we implement it here in the case of linear regression . We tested the method both on generated and real datasets; the real dataset concerns the population of Alpine ibex (*Capra i. ibex*), a long-lived mammal species. We found that the methodology is robust and permits to deal with model prior uncertainty, giving prediction intervals instead of prediction points.

The paper is organized as follows: in Section 2 we summarize the main concepts of BMA methodology; in Section 3 we present the theoretical development of CMA and we propose a metric to evaluate interval predictors; in Sections 4 and 5 we present the study cases and the results of the experiments; the main conclusions are reported in Section 6.

2 Bayesian model averaging (BMA)

Let us consider a simple linear regression model structure of the type $y = \beta_0 + \sum_{\mathcal{X}} \beta_j X_j + \varepsilon_t$, where $\mathcal{X}$ is the set of the covariates $\{X_1, X_2, \dots X_k\}$ and ϵ a white noise. Given k covariates, there are 2^k candidate model structures, obtained by combining in all the possible ways the presence/absence of each covariate. The model size is defined as the number of covariates included in the model.

Even using a well-established model selection criteria (e.g. AIC or BIC), the choice of the supposedly "best" model is often uncertain, because many models show a similar score. However, different models with similar scores can return quite different predictions; this is the problem of *model uncertainty*. BMA addresses model uncertainty weighting the inferences produced by the different models. Given a dataset D, the weights are constituted by the models' posterior probabilities $P(M_i|D)$, where by M_i we denote the *i*-th model. Inference about the expected value quantity of interests Δ is therefore obtained taking into consideration all the different model structures [see Clyde and George, 2004]:

$$E[\Delta|D] = \sum_{i=1}^{2^k} E[\Delta|M_i, D] P(M_i|D) \tag{1}$$

where the obtained distribution is a sum of distributions and thus has generally a multimodal shape.The posterior probabilities of the models are calculated using the Bayes formula, that requires to specify a prior probability $P(M_i)$ for each model: $P(M_i|D) = \frac{P(M_i)P(D|M_i)}{\sum_{k \in 1 \dots m} P(M_k)P(D|M_k)}$, where $P(D|M_i) = \int P(D|M_i, \boldsymbol{\beta_i}) P(\boldsymbol{\beta_i}|M_i) d\boldsymbol{\beta_i}$ is the marginal likelihood of the model M_i, and $\boldsymbol{\beta_i}$ is the vector of parameters of model M_i. We refer the reader to Raftery and Madigan [1997] for details about the computation of the marginal likelihood in case of linear regression. Moreover, corresponding to model M_i, we should specify a prior distribution on the parameters $\boldsymbol{\beta_i}$, namely $P(\boldsymbol{\beta_i}|M_i)$. In this work we adopt Zellner's *g*-prior on the regression parameters as in Fernandez and Ley [2001].

The summation of (1) is extensive over 2^k models. To keep the computation feasible, *Markov Chain Montecarlo* methods (MCMC) are generally adopted to sample the model space, without thus implementing all the 2^k models. Only for small k it is possible to exhaustively treat the model space.

BMA requires to specify a prior over the models. A simple non-informative prior is the uniform one. In this case $P(M_i) = P(M) = 2^{-k}$ and the expected model size (i.e., the expected number of included covariates) is $k/2$. A popular alternative is the binomial prior [Raftery and Madigan, 1997; Fernandez and Ley, 2001], under which the prior probability of a model is computed on the basis of θ_j, which represents the probability of inclusion of each covariate X_j. As shown in Ley and Steel [2009], if the probability of inclusion of each covariate is independent and constant ($\theta_1 = \theta_2 = \ldots = \theta_k = \theta$), then the model size will have a binomial distribution $\sum_{j=1}^{k} \gamma_j \sim Bin(k, \theta)$, with mean $\theta \cdot k$ and variance $\theta(1-\theta)k$. The prior probability of a model is calculated as

$$P(M_i) = \theta^{k_i}(1-\theta)^{k-k_i} \tag{2}$$

where k_i is the number of the covariates included by model M_i. If one wants to express prior knowledge, one should elicit from an expert the value of a single parameter, θ, or alternatively the model size, which corresponds to $\theta \cdot k$. Under this prior, all the models with the same size (k_i) have the same probability. If $\theta = 0.5$ the binomial prior coincides with the uniform prior.

The binomial model prior can be extended by adopting a hierarchical prior; treating θ as a random variable increases the model size variance and makes prior distribution over the models flatter. The use of a beta-binomial prior is widely discussed in Ley and Steel [2009].

3 Credal Model Averaging(CMA)

Here we define the CMA, a method which substitutes the single prior over the models by a set of priors; in this way, it represents a condition of prior *ignorance* about the appropriateness of the models, letting their prior probability vary within a large range This allows eliciting from the decision maker an interval instead of a single value for the expected model size a priori; an expert is indeed generally more confident in providing a range rather than a single point estimate, thus being allowed to express the uncertainty in his prior knowledge. As a result of having a set of priors, CMA produces interval predictions, rather than point predictions.

More specifically, we developed an imprecise version of the binomial model prior defined in (2). Thus, we assume that the expert specifies an upper and a lower probability of inclusion for the covariates, respectively $\underline{\theta}$ and $\overline{\theta}$. Introducing $\underline{\theta}$ and $\overline{\theta}$, the expected model size a priori ranges between $\underline{\theta} \cdot k$ and $\overline{\theta} \cdot k$. Starting from the prior interval, we find the expected interval of the prediction for every available instance. To completely define the prediction interval, it is sufficient to find the prediction bounds. Therefore we have to solve two optimization problems: finding the maximum and the minimum of the prediction (1). The minimization is formalized as

$$\begin{aligned} \min_{\theta} E[\Delta|D] = \min_{\theta} \sum_i E[\Delta|M_i, D] \frac{P(D|M_i)P(M_i)}{\sum_i P(D|M_i)P(M_i)} = \\ = \min_{\theta} \frac{\sum_i E[\Delta|M_i, D] P(D|M_i)\theta^{k_i}(1-\theta)^{k-k_i}}{\sum_i P(D|M_i)\theta^{k_i}(1-\theta)^{k-k_i}} \\ \text{subject to: } \underline{\theta} \leq \theta \leq \overline{\theta} \end{aligned} \tag{3}$$

Note that we have a single unknown, having assumed that the prior probability of inclusion is equal for all covariates. Defining the k sets $G_1 \ldots G_k$, containing respectively $\{1, 2, \ldots, k\}$ covariates, for every instance the function to minimize/maximize with respect to θ become, from (3):

$$E[\Delta|D] = h(\theta) = \frac{\sum_{j=0}^{k} \theta^j (1-\theta)^{k-j} Z_j}{\sum_{j=0}^{k} \theta^j (1-\theta)^{k-j} L_j} \tag{4}$$

where $Z_j = \sum_{v \in G_j} E[\Delta | M_v, D] P(D|M_v)$ and $L_j = \sum_{v \in G_j} P(D|M_v)$. Notice that the length of this prediction interval changes instance by instance. This interval does not have the coverage properties of a confidence interval; rather, it shows the sensitivity of the prediction to the prior over the models. Therefore, the prediction interval represents how much the BMA prediction varies when the θ used to specify the binomial prior varies between the assigned $\overline{\theta}$ and $\overline{\theta}$.

In the interval [$\underline{\theta}, \overline{\theta}$], the maximum (minimum) of the prediction $h(\theta)$ corresponds to $\theta = \underline{\theta}$, $\theta = \overline{\theta}$ or to a value of θ for which the first derivative of (4) is zero. The first derivative of (4) can be identified, setting $f(\theta) = \sum_{j=0}^{k} \theta^j (1-\theta)^{k-j} Z_j$ and $g(\theta) = \sum_{j=0}^{k} \theta^j (1-\theta)^{k-j} L_j$, by

$$\frac{dh(\theta)}{d\theta} = \frac{f'(\theta)g(\theta) - f(\theta)g'(\theta)}{g(\theta)^2} \tag{5}$$

The value of $g(\theta)$ is strictly positive because L_j is a sum of marginal likelihoods, therefore we search the solutions looking only at the numerator of $\frac{dh(\theta)}{d\theta}$, $f'(\theta)g(\theta) - f(\theta)g'(\theta)$, which is a polynomial of degree $2k$ and thus it has $2k$ solutions in the complex plain. We are interested only in the *real* solutions that lie in the interval ($\underline{\theta}, \overline{\theta}$). Such solutions, together with the boundary solutions $\theta = \overline{\theta}$ and $\theta = \underline{\theta}$, constitute the *candidate solutions set*. To find the minimum and the maximum predictions, we finally calculated the prediction values in correspondence of each *candidate solution*. If $\underline{\theta} < 0.5$ and $\overline{\theta} > 0.5$, the prediction of the uniform prior case ($\theta = 0.5$) is always included in the prediction interval, even though it does *not* generally constitute the center of the BMA interval. Notice that CMA requires additional computational complexity compared to BMA, because of the need for solving two optimizations problems for every prediction point.

3.1 Evaluating CMA performances

Evaluating performances is a challenging task because we need to compare the single-point predictions of the BMA and the interval estimates of the CMA. A possibility is to use the squared error between the measure (y) and the central point of the interval (y_c): CMA *central error* $= (y - y_c)^2$. This permits an easy comparison with the BMA squared error but does not take into account the information linked with the extension of the prediction interval of BMA. Currently, there are no well-established metrics to compare a point prediction and an interval prediction. On one hand, the metric should penalize large prediction intervals, because they are little informative; and on the other hand, it should reward the prediction interval when it contains the measured value or when its boundaries falls close to it. We propose the following Interval Error (IE) metric:

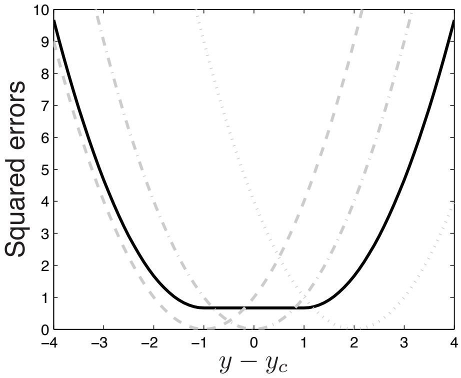


Figure 1: Shape of IE, defined in (6), as function of the distance $y - y_c$ and using the parameters' setting $\alpha = 2/3$, $\beta = r = 1$. We also plot the errors made by three different punctual predictions, positioned at $y = y_c$ (dashed line), $y = y_c - r$ (dash-dotted line) and $y = y_c + 2r$ (dotted line).

$$IE = \begin{cases} \alpha r^2 & \text{if } y_c - r < y < y_c + r \\ \alpha r^2 + \beta min[(y_c - r - y)^2, (y_c + r - y)^2] & \text{otherwise} \end{cases} \tag{6}$$

where r is defined as half of the length of the prediction interval, α and β are constants whose values have to be chosen. If y is in the prediction interval of the CMA, IE measures a constant error which is quadratic in the length r of the semi-interval. Conversely, when y is outside the interval, the metric considers both the interval length and the quadratic distance from the nearest interval bound. This preliminary metric has several limits, especially the subjective choice of α and β. In Figure 1 we report, as function of $y - y_c$, the IE error with its parameters fixed to a specific value, and the quadratic errors of three punctual predictions. In the following we consider a IE defined by $\beta = 1$ and $\alpha = 2/3$ or $\beta = 1$ and $\alpha = 1$. Notice that, with the chosen value of $\alpha = 2/3$, if the measured value is inside the prediction interval, a single-point estimator performs better than CMA only if it is within a distance of $r\sqrt{\frac{2}{3}} \sim 0.82r$ from the measured value. More generally, we can see that, if the punctual prediction is "near" the measured value, it outperforms the interval prediction.

4 Study cases

We tested CMA both on generated data and on "real" datasets, setting $\underline{\theta} = 0.05$ and $\overline{\theta} = 0.95$. In this way, we model a condition of prior ignorance. For BMA we set $\theta_{BMA} = 0.5$, namely a uniform prior on the models. Since $\underline{\theta} \leq \theta_{BMA} \leq \overline{\theta}$, the predictions of the BMA are always included in the prediction interval of CMA. For both artificially generated data or real one, we performed experiments with different size of the training set. For each size n, we performed 30 different training/test experiments, of which we report the mean.

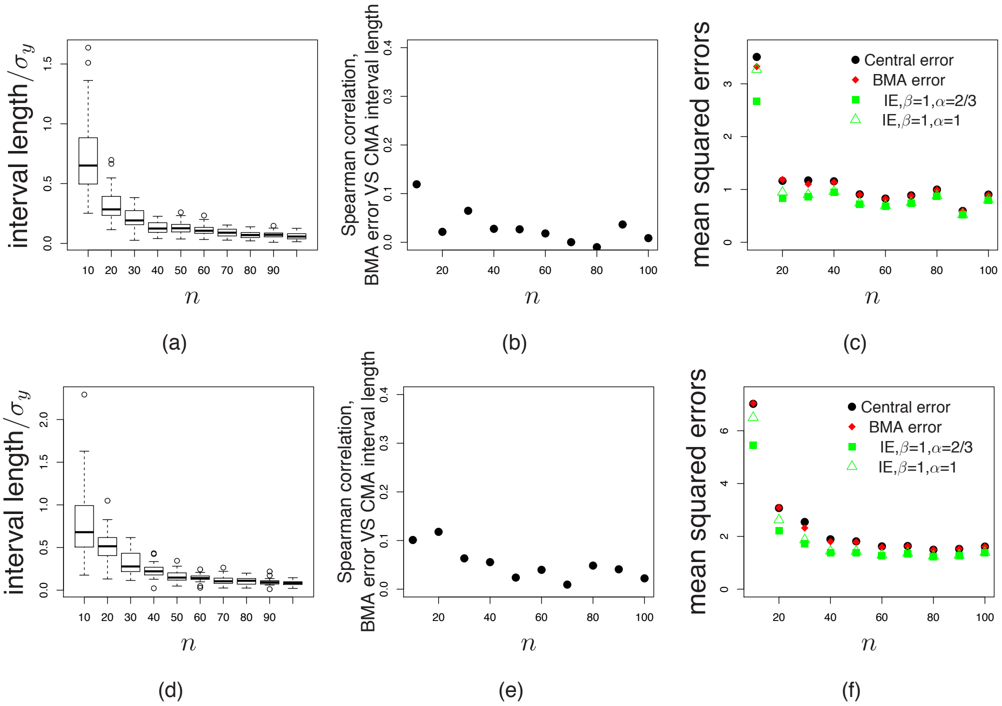
As real study case, we investigated the population dynamic of Alpine ibex of the Gran Paradiso National Park, Italian Alps ($45°$ 25' N, $7°$ 34' E). The counts, performed from 1956 until nowadays, represent the longer continuous existing data series of Alpine ungulates abundance. The set of available covariates are presented in the Appendix A of Jacobson et al. [2004]. The covariates contain the population abundance (N_t) and ten meteorological variables, such as: the average snow depth (cm), the number of days of snow depth above the mean, the number of days of snow depth above the mean plus one standard deviation, the average daily maximum temperature in winter ($°C$), the average daily minimum temperature in winter ($°C$), the average daily maximum temperature in summer ($°C$), the average daily minimum temperature in summer ($°C$), the total precipitation in spring (mm), the total precipitation in winter (mm) and the total precipitation in summer (mm). Starting from these covariates, the authors select only the two most correlated with the logarithm of the growth rate: N_t and the mean winter snow depth, S_t. Finally they select as best, using AIC, a linear model for the logarithm of the growth rate ($Log[{}^{N_{t+1}}/_{N_t}]$) that includes N_t, S_t and the interaction term $N_t S_t$. Here we adopt the BMA and the CMA methodology for the same dataset considering all the eleven available covariates for the model, avoiding the step of covariate selection. The general model structure is $Log\left({}^{N_{t+1}}/_{N_t}\right) = \beta_0 + \sum_{\mathcal{X}_i} \beta_j X_{j,t} + \varepsilon_t$; the dataset contains 38 instances.

The setting to generate a synthetic dataset is defined as $S = \{\tilde{\theta}, k, n, snr\}$, were k is the number of covariates, $\tilde{\theta}$ is the probability of inclusion of the covariates, n is the dataset length and snr is the ratio between the output noise and the input variance, as we formulate below. A dataset was generated, using a specific S, as follows: we extracted n values for each covariate from a standard Gaussian; then we randomly included each covariate in the model with probability $\tilde{\theta}$; for each covariate X_j in the set of the included ($\mathcal{X}_i$), we randomly extracted the coefficient of regression β_j from a standard Gaussian; the model output was finally calculated as $y = \sum_{\mathcal{X}_i} \beta_j X_j + \varepsilon$, where $\varepsilon \sim N\left(0, snr \sum_{\mathcal{X}_i} \beta_j^2\right)$. Eventually, we validated the model on other 100 generated points. We generated the datasets using two different values for snr (0.05 and 0.15), two different values for k (10 and 30), nine values for $\tilde{\theta}$ ($0.1, 0.2 \ldots, 0.9$) and ten values for n ($k, 2k, \ldots, 10k$). Notice that the value $\tilde{\theta} = 0.5$ is the value we set also for the BMA predictions, and that corresponds to a uniform prior above all the models.

5 Results

We use the BMA implementation from the R package BMS [Feldkircher and Zeugner, 2009], that we have extended to carry out the CMA methodology; we adopt Zellner's g prior with zero-mean priors for the coefficients.

As an example, we show the results of the following setting with generated data: $snr = 0.15$, $k = 10$, $\theta = \{0.5, 0.9\}$ and $n = \{k, 2k \ldots 10k\}$. These settings produce almost the same results, as you can see comparing the first ($\theta = 0.5$) and the second ($\theta = 0.9$) row of Figure 2. The experiments made under other experimental settings produce quite similar results. As expected, the length of the prediction interval decreases with the training set dimension (Figure 2.[a,d]). Notice that both the median and the variance of the prediction interval length decrease with the training set dimension. There is a positive correlation between the squared error of BMA and the interval length (Figure 2.[b,e]). This correlation tends to vanish with the increase of n. This means that when the CMA prediction interval increases, the BMA error tends to increase too, especially if the dataset available for the model identification is small (say $n < 4k$). This correlation is not large; still it is significant. The interval length reflects the prior uncertainty, but there are many other sources that contribute to the prediction error, such as the uncertainty linked with the parameter estimation and the noise. Therefore we could not expect a very large correlation between the length of the CMA prediction interval and the BMA error.



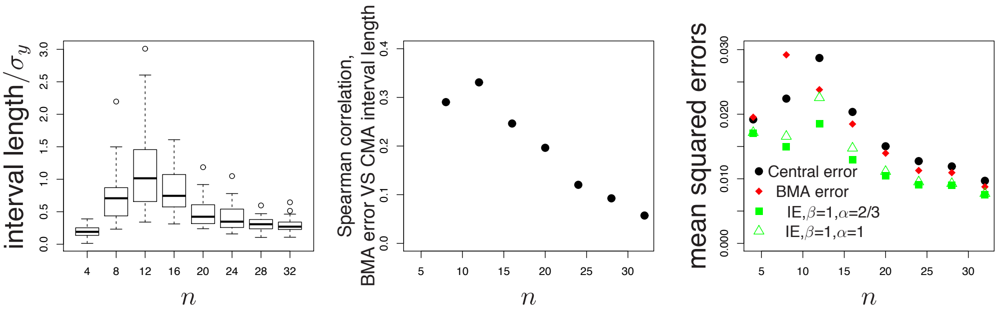
(a) (b) (c)

(d) (e) (f)

Figure 2: Results of the experiments on randomly generated datasets, as function of the training set dimension, for $k = 10$ and $snr = 0.15$. The θ used to generate data is 0.5 for the first row of plots and 0.9 for the second. Sub-figures (a) and (d) show the boxplots of the mean interval length of the CMA predictions, divided by the standard deviation of the output. Sub-figures (b) and (e) show the mean Spearman correlation between the CMA interval length and the BMA prediction error. Finally, Sub-figures (c) and (f) show the squared predictions errors of the BMA and CMA, using both the CMA central error and the IE defined in (6). The black dots represent the mean CMA central errors, the red diamonds the mean BMA squared errors, the green squares the IE with $\beta = 1$ and $\alpha = 2/3$ and the green triangles the IE with $\beta = 1$ and $\alpha = 1$.

The square loss of BMA has similar value, across the various experiments, to the square loss obtained by using the central point of the CMA interval as prediction (CMA central error), as shown in Figure 2[c, f]. In fact, the BMA prediction position seems to fluctuate uniformly at random in the prediction interval of the CMA. The mean distance between the CMA central prediction and the BMA prediction is indeed ~ 0 and its absolute value is about a quarter of the CMA interval length ($\frac{1}{2}r$). The IE defined in (6) with $\beta = 1$ and $2/3 \leq \alpha \leq 1$ is always smaller than the CMA central prediction error and than the BMA prediction error (see Figure 2.[c, f]). These metrics indicate an encouraging performance for CMA, although the IE is only a preliminary metric for scoring interval predictions.

In the Alpine ibex case, BMA identifies the following covariates as those with bigger posterior probability of inclusion (sum of the posterior probability of the models that contain the specific covariate): N_t (0.988), S_t (0.871) and the total precipitation in summer (0.673), while the other covariates have a probability of inclusion that is less than 0.23. Notice that, while N_t and S_t are included in the model of Jacobson et al. [2004], the summer precipitation is excluded by the authors but has a large probability of inclusion using BMA. As in the generated datasets, both the median and the variance of the interval length tend to decrease with the training set dimension (see Figure 3.[a]). As partial exception, we can see that small training sets are characterized by small interval lengths. This behavior is due to the fact that, when the training set dimension n is less then k, the method implemented in the BMS package takes by default only models with no more than $n - 3$ covariates. The correlations of the squared error made by the BMA prediction with the CMA interval length are significant and positive (see Figure 3.[b]). Moreover they are larger in this real case than in the generated datasets. The errors made by the BMA and the CMA central errors have similar values for $n >> k$ (see Figure 3.[c]), with a slightly bigger error for the CMA central point prediction. For smaller n, BMA error is smaller or larger than CMA central error depending on the study case, without showing a clear pattern. Even in this case the BMA prediction position can be considered randomly and uniformly distributed in the prediction interval of the CMA. Using IE defined in (6) with $\beta = 1$ and $2/3 \leq \alpha \leq 1$, CMA always outperforms BMA to make predictions.



(a) (b) (c)

Figure 3: Results of the experiments made on the Alpine ibex dataset, as a function of the training set dimension. The sub-figures have the same meaning of their corresponding sub-figures of Figure 2.

6 Conclusions

In the BMA methodology, the prior model probability has to be given as a precise value. Often, the prior knowledge on the system is instead not so precise. Our method, CMA, expands the BMA methodology considering the uncertainty in the model prior probability. It in fact allows to specify a set of priors, instead of a single value, for the probability of inclusion of the covariates in a linear regression problem. Given a set for the probability of inclusion of the covariates, we found a solution that analytically generates prediction intervals. The length of the interval represents the effects of the prior of the models and, therefore, decreases with the dimension of the training set. Moreover, the length of the interval is significantly correlated with the error made by BMA; namely, larger intervals of CMA tend to correspond to larger errors for BMA. This is noteworthy, since BMA has no natural tool for evaluating the sensitivity of the prediction on the specification of the prior. The obtained results are consistent for both real and generated datasets. A new metric called *Internal Error* (IE), has been defined to assess the goodness of the prediction interval in comparison with the punctual prediction. Using this metric, the performances of CMA seems encouraging. Nevertheless, currently the IE can only be considered a preliminary metric to evaluate an interval prediction against a punctual one.

As future extensions, to make the methodology more elastic, we will allow the prior probability of inclusion θ_i to be different for each covariate. This allows the expert to precisely express his/her prior knowledge on the probability of inclusion for each covariate. In this case an analytical solution cannot be found, and solve the optimization problem becomes much more complicated.

The performance of CMA with small training set dimensions makes the method particularly suitable to address ecological problems, especially population dynamics problems. These are, in fact, usually characterized by small datasets and by many environmental characteristics that can be defined as covariates, as in the Alpine ibex study case presented here.

Acknowledgments

Research partially supported by Fondazione Lombardia per l'Ambiente (SHARE-Stelvio project), the Swiss NSF grants no. 200020-132252 and by the Hasler foundation grant n. 10030.

References

Burnham, K. and D. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. - Springer Verlag, May 2002.

Clyde, M. and E. I. George. Model Uncertainty. *Statistical Science*, 19(1):81–94, February 2004.

Conroy, M., Y. Cohen, F. James, and Y. Matsinos. Parameter estimation, reliability, and model improvement for spatially explicit models of animal populations. *Ecological*, 1995.

Feldkircher, M. and S. Zeugner. Benchmark priors revisited: On adaptive shrinkage and the supermodel effect in bayesian model averaging. *IMF Working Papers*, 2009.

Fernandez, C. and E. Ley. Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 2001.

Jacobson, A., A. Provenzale, A. Von Hardenberg, B. Bassano, and M. Festa-Bianchet. Climate forcing and density dependence in a mountain ungulate population. *Ecology*, 85(6):1598–1610, 2004.

Ley, E. and M. Steel. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *Journal of Applied Econometrics*, 24(4):651–674, 2009.

Raftery, A. and D. Madigan. Bayesian model averaging for linear regression models. *Journal of the American Statistical*, 92(437):179– 191, 1997.

Walley, P. *Statistical reasoning with imprecise probabilities*. Chapman and Hall London, 1991.