Bayesian Approach using Experts' knowledge: The Impact of Rainfall on the Production of Périgord Black Truffles

Paul-Marie Grollemund [1], Christophe Abraham [2], Meïli Baragatti [2]

[1] Department of Mathematics, Imperial College London, London, United Kingdom

[2] MISTEA UMR 729, Montpellier SupAgro, INRA, CNRS, Univ Montpellier, Montpellier, France

---

Address for correspondence: Paul-Marie Grollemund, 6M30 Huxley Building South Kensington Campus, Department of Mathematics, Imperial College London, London, SW7 2AZ, United Kingdom.

E-mail: `pgrollem@ic.ac.uk`.

Phone: .

Fax: .

Abstract: An important part of Bayesian statistical modelling is to specify the prior distribution of model parameters. When prior knowledge from subject-matter experts is available, it is possible to build a model which includes prior information. The collection of this prior knowledge is a delicate task because the statistician has to state the experts' knowledge in probabilistic terms. In this paper, we present two approaches to elicit experts' knowledge about the Bliss model which is a parcimonious Bayesian Functional Linear Regression model. The proposed methodologies will be illustrated on synthetic data and will be applied to estimate the influence of the rainfall on the production of Périgord black truffles.

Key words: Bayesian statistics; Elicitation; Expert knowledge; Functional linear regression ; Informative prior

1 Introduction

The choice of a prior is a central issue in Bayesian modeling. A standard approach is to determine a noninformative prior in order to perform an Objective Bayesian inference. Nevertheless, a noninformative prior may not be an appropriate choice to examine small samples, for instance when the data collecting process is delicate or expensive. In this case, data may not provide enough information to inform a complex model with relevant statistical inference. One option, among others, is to perform Subjective Bayesian inference by modeling prior knowledge. The use of Subjective or Objective Bayesian inference raises conceptual and philosophical discussions, see Hoffmann (2017) for an overview. In this paper we employ a Subjective approach as a tool which enables us to complement sparse data using expert knowledge.

Expert knowledge has to be collected in order to model an informative prior and Elicitation denotes such a prior information collecting process. Many reviews on elicitation are available, see among others Garthwaite et al. (2005a); Ouchi (2004); Kynn (2005); Jenkinson (2005). Literature on elicitation methods covers various scientific fields. For instance, some papers concern psychological issues, such as biais of elicitation when collecting information from several experts due to human behavior and human interaction, see Kadane and Wolfson (1998) and Kynn (2008). Alternatively, other papers can be found in application fields, for instance in economy, ecology, genomics, military intelligence or nuclear energy (see O'Hagan et al., 2006 for various examples for which an elicitation is required). Some important literature concerning elicitation mainly deals with the elicitation process and modeling of prior information in probabilistic terms. In other words, the question is how to interact with experts for collecting their knowledge on the studied subject in order to efficiently inform prior probabilities about a model parameter θ. The difficulty of such work is twofold. Firstly the method of obtaining information must be adapted to experts who are often unfamiliar with probability concepts. Hence, it is usual to develop an elicitation process which is as simple as possible for experts. For this purpose, protocols describe how to carry out successful elicitation by following important steps (see among others Winkler et al., 1992; Cooke and Goosens, 2000; Low-Choy et al., 2009). The second main difficulty is to include such elicited knowledge in a probabilistic model. For instance, an important issue is the aggregation of knowledge from several experts and there are mainly two manners to aggregate, which are described in Ouchi (2004). On the one hand, the behavioral approach aims at reaching a consensus, like for example the Delphi method (see Dalkey and Helmer, 1963; Chu and Hwang, 2008) or the Nominal group method (Delbecq and Van de Ven, 1971). On the other hand, the mathematical approach aims at gathering knowledge from experts by pooling it (Burgman et al., 2011) or by averaging it (Genest and Zidek, 1986). Of course, other approaches exist, see for example Hunns and Daniels (1981) and Cooke (1991) for more details. Another way to examine the literature on elicitation is to consider the separation between the direct and indirect elicitation processes. In some cases, experts can be viewed as statisticians in such a way that they may tune prior distribution or hyperparameter values in accordance with their knowledge of the studied subject (Zellner, 1972). Then, for a direct elicitation, experts do not need assistance for expressing their knowledge in probabilistic terms (see for example Winkler, 1967; Fleishman et al., 2001; Kadane et al., 1980; O'Leary et al., 2008). However, experts have generally insufficient understanding of statistical modeling and it may not be effecient to interview them about theoretical quantities (Garthwaite et al., 2005b). In this case, it is necessary to perform an indirect elicitation process by interviewing experts about observable quantities which are familiar for them (see for example O'Hagan et al., 2006; Albert et al., 2012). As the interaction with experts is not directly about model parameters, an effort must be made to model their elicited knowledge with a prior.

In the context of the Linear Regression model, many elicitation processes have been developed for dealing with specific problems or for handling a certain type of expert information. For example, elicitation processes are introduced in order to achieve variable selection (Garthwaite and Dickey, 1992) or for quantifying knowledge about the error variance (Garthwaite and Dickey, 1991). For the Linear Regression model, James et al. (2010) stress the difficulty of asking experts about regression coefficients. As the design is usually not orthogonal, slope coefficients do not have a meaningful interpretation and some authors deal with the elicitation of features of the regression coefficients, as for example the sign of the regression coefficients for which an interpretation is reliable (see for instance Kuhnert et al., 2005; Martin et al., 2005; O'Leary et al., 2008). Alternatively, it is relevant to perform indirect elicitation by asking experts about observable quantities for collecting expert knowledge about regression coefficients (see among others Crowder, 1992).

In this paper, we address the problem of assessing a prior distribution informed by expert knowledge for the Bliss model which is a particular case of the Functional Linear Regression model (Grollemund et al., 2018). This model aims to provide a simple estimate of the coefficient function in order to ease interpretation. Suppose we observe an outcome variable y_i related to a functional covariate $x_i(\cdot)$ depending on the time $t \in [0,1]$, for $i = 1, \ldots, n$. The Functional Linear Regression model (FLR) is given by

$$y_i|\mu, \beta(\cdot), \sigma^2 \overset{\text{ind}}{\sim} \mathcal{N}\left(\mu + \int_0^1 x_i(t)\beta(t)\mathrm{d}t \ , \ \sigma^2\right), \text{ for } i = 1, \ldots, n. \tag{1.1}$$

We refer the reader to Reiss et al. (2016) for a recent comprehensive survey of methods for fitting (1.1). The Bliss approach is based on an adaptive decomposition of β on a set of K step functions:

$$\beta(t) = \sum_{k=1}^{K} b_k \frac{1}{|\mathcal{I}_k|} \mathbf{1}_{\mathcal{I}_k}(t),$$

where each b_k is a number and each $\mathcal{I}_k$ is an interval. The interest of the above decomposition is that the union of the intervals $\mathcal{I}_k$ is not equal to the domain $[0,1]$ and provides an interpretable exression of β (see Grollemund et al., 2018 for details). By considering the previous constraint, The Bliss model is derived from (1.1) as follows:

$$y_i|\mu, \boldsymbol{b}, \sigma^2, \boldsymbol{\mathcal{I}} \overset{\text{ind}}{\sim} \mathcal{N}\left(\mu + x_i(\boldsymbol{\mathcal{I}})^T \boldsymbol{b} \ , \ \sigma^2\right), \text{ for } i = 1, \ldots, n \tag{1.2}$$

where $x_i(\boldsymbol{\mathcal{I}})$ is a vector with k^{th} entry $\frac{1}{|\mathcal{I}_k|}\int_{\mathcal{I}_k} x_i(t)\mathrm{d}t$ and $\boldsymbol{b} = (b_1, \ldots, b_K)^T$. When each interval $\mathcal{I}_k$ is set as

$$\mathcal{I}_k = [m_k - \ell_k, m_k + \ell_k] \cap [0,1] \text{ and } \boldsymbol{\mathcal{I}} = (\mathcal{I}_1, \ldots, \mathcal{I}_K),$$

the prior distribution is

$$\begin{aligned} \mu|\sigma^2 &\sim \mathcal{N}\left(0, v_0\sigma^2\right), \\ \boldsymbol{b}|\sigma^2, \boldsymbol{\mathcal{I}} &\sim \mathcal{N}_K\left(0, \sigma^2\Sigma(\boldsymbol{\mathcal{I}})\right), \\ \pi(\sigma^2) &\propto 1/\sigma^2 \\ m_k &\overset{i.i.d.}{\sim} \mathrm{Unif}\left(\mathcal{T}\right), \quad k = 1, \ldots, K, \\ \ell_k &\overset{i.i.d.}{\sim} \mathcal{E}(a), \quad k = 1, \ldots, K, \end{aligned} \tag{1.3}$$

where $\Sigma(\boldsymbol{\mathcal{I}})$ (depending on the intervals $\boldsymbol{\mathcal{I}}$) is the Ridge Zellner prior covariance. The authors used this model to study the impact of rainfall on the production of the Périgord black truffles (data set provided by J. Demerson). For this study, only limite data was available so the coefficient function is complicated to estimate (13 observed years and 11 parameters). Therefore, it appears important to rely on supplementary information in order to reinforce the statisical inference. Furthermore, scientists studying truffles and truffle farmers have relevant knowledge regarding the growth or the complex reproduction mechanisms of truffles. Such important information is overlooked if the inference is only based on observed data.

In this paper, we aim to fit the Bliss model by taking expert knowledge into account, so we propose an elicitation process for obtaining information by targeting two underlying goals. Firstly, we aim to develop a method which is as simple as possible for the experts. Secondly, we aim to drive the impact of expert knowledge on the posterior distribution. Below, two approaches are developed to achieve these goals. In the first approach, we ask the experts to provide pseudo data, in other words we expect a set of covariate functions and associated response values which reflect their knowledge. Furthermore, we ask each expert to state their certainty concerning theirs pseudo data by specifying values between 0 and 1. We propose a prior distribution based on the elicited pseudo data and we introduce weights to tune the impact of pseudo data on the posterior. For the second approach, we elicit features of the coefficient function which are interpretable for them. A prior distribution is derived from (1.3) by introducing a penalization term to promote parameter values in accordance with the experts' knowledge and a tuning parameter which drives the strength of the penalization.

The remainder of the paper is structured as follows. In Section 2, we describe the elicitation process for obtaining pseudo data and the prior modeling of pseudo data. In Section 3, we propose to collect knowledge about features of the coefficient function and a different prior distribution is introduced to incorporate this knowledge into the model. Section 4 is a brief description of the implementation required for sampling the posterior distributions of each proposed method. In Section 5, the proposed approaches are applied to simulated datasets and we present an agronomic study which aims to identify the impact of climatic conditions on the production of the Périgord black truffles. The elicitation process is applied to scientists and truffle farmers and we present the results obtained by using the two approaches. Concluding remarks are given in Section 6.

2 Prior distribution based on pseudo data

In the context of The Functional Linear Regression model, elicitation of the prior knowledge relative to the coefficient function is a laborious task. Indeed, the interpretation of the coefficient function is complicated, so it is difficult to propose a value of the function at a fixed time point or for a time period. Hence, a direct elicitation of the regression coefficients is not easy. Nevertheless, experts may have a valid opinion about the impact of rainfall on the truffle production from their experience, and we aim to capture the maximum amount of information about their knowledge. Therefore, we propose to elicit expert knowledge by using an indirect method since it is more convenient in this context to ask the experts about observable quantities. In the regression framework, we identify that it is appropriate to ask the experts to provide likely data, which we name pseudo data (see for instance Neal, 2001; Wolfson and Bousquet, 2014; Bousquet and Keller, 2018). We claim not only that this method is simple for the experts but also that such elicited data reflects their knowledge about the relationship between covariates and outcomes. Moreover, we collect it without specifying any probabilistic model.

Below, we denote the observed data by y_i and $x_i(\cdot)$, for $i = 1, \ldots, n$, and the pseudo data from the expert e by y_i^e and $x_i^e(\cdot)$, for $i = 1, \ldots, n_e$ and $e = 1, \ldots, E$. Moreover, we denote by $\boldsymbol{y}$ the observed data vector, y^e the pseudo data vector $(y_1^e, \ldots, y_{n_e}^e)$ and $\boldsymbol{x^e}(\cdot)$ the vector $(x_1^e(\cdot), \ldots, x_{n_e}^e(\cdot))$. We keep in mind that the pseudo data is different by nature, being from a variety of sources. Therefore, we may consider that a pseudo outcome y_i^e is more uncertain than an observed outcome y_i. Indeed, it seems difficult for a person, even an expert, to accurately provide a real number. Alternatively, one may consider that expert knowledge is learned from many observations and thus pseudo data corresponds to a sort of valid prediction which is less uncertain than observed data. In this paper, we deem with regard to the aforementioned application that it is much too complicated for experts to accurately provide numbers y_i^e and especially to provide curves x_i^e. However, slight modifications of the following approach may improve the accuracy of prior knowledge. Below, we propose a new prior distribution which allows to take the uncertainty of the pseudo data into account and to drive the impact of the experts' prior knowledge on the Bayesian inference.

2.1 Model

Below we derive a new prior depending on pseudo data from an initial prior which we assume to be noninformative or weakly informative. As we aim a prior reflecting expert prior knowledge, we define it as a kind of posterior, given pseudo data. Hence, the proposed prior is what we learn from the experts' opinion. Let us assume for a moment that we obtain this new prior from pseudo data and an *initial prior* by simply applying the Bayes rule. If we apply again the Bayes rule with data and the new prior, we actually perform a sequential Bayes rule which is equivalent to applying Bayes rule with the merged data (observed data and pseudo data) and the *initial prior*. In this case, observed data and pseudo data equally contribute to the inference although we see the pseudo data as more uncertain so their influence should be weaker than the contribution of observed data. Therefore, rather than applying the classical Bayes rule, we prefer to use a different rule.

Different variations of the Bayes rule have been considered, like for instance the fractional posterior (O'Hagan, 1995 and see Bhattacharya et al., 2016 for a comprehensive recent review) or the generalized Bayesian posterior (Grünwald, 2012). Both approaches define an alternative likelihood, called the fractional likelihood, as follows:

$$L_\alpha\left(\boldsymbol{D};\theta\right) = p_\theta(D)^\alpha \tag{2.1}$$

where θ is a parameter, $p_\theta(\cdot)$ is the density of the Bliss model (1.2), $\boldsymbol{D} = (z_1, \ldots, z_n)$ is the observed dataset, and α is a tuning parameter belonging to $(0, 1)$. The fractional posterior deriving from the (2.1) is given by

$$\pi_\alpha(\theta|\boldsymbol{D}) \propto p_\theta(\boldsymbol{D})^\alpha \times \pi(\theta) \tag{2.2}$$

for a prior π. In this setting, α controls the proportion of information in the dataset used to learn the posterior from the data (and the prior). Below, we propose to use the fractional posterior of O'Hagan (1995) to learn from pseudo data.

In a different framework, the Power Prior method (Ibrahim and Chen, 2000) proposes to define the prior π as a fractional posterior given historical data $\boldsymbol{D_0}$ and for an *initial prior* π_0,

$$\pi(\theta; \boldsymbol{D_0}, \alpha) \propto p_\theta(\boldsymbol{D_0})^\alpha \times \pi_0(\theta). \tag{2.3}$$

Then, the posterior given a dataset $\boldsymbol{D}$ is

$$\pi(\theta|\boldsymbol{D}, \boldsymbol{D_0}, \alpha) \propto p_\theta(\boldsymbol{D}) \times p_\theta(\boldsymbol{D_0})^\alpha \times \pi_0(\theta). \tag{2.4}$$

In this case, we can easily understand the contribution of the historical dataset and the contribution of the observed dataset.

In this paper, the first approach we propose balances between the Power Prior (2.3) and the Weighted Likelihood (Hu and Zidek, 1995), which gives weights for each datum:

$$L_w(\theta) = \prod_{i=1}^{n} p_{\theta,i}(z_i)^{w_i}, \tag{2.5}$$

where $p_{\theta,i}(\cdot)$ is the density of the Bliss model (1.2) for the i^{th} datum and w_i is the weight of the i^{th} datum z_i.

The prior we propose is then:

$$\pi(\theta|\boldsymbol{D_1}, \dots, \boldsymbol{D_E}; w) \propto \prod_{e=1}^{E} \prod_{i=1}^{n_e} p_{\theta,i}(z_i^e)^{w_i^e} \times \pi_0(\theta), \tag{2.6}$$

where $\boldsymbol{D_e} = (z_e^1, \dots, z_e^{n_e})$ is the pseudo data set provided by the expert e, for $e = 1, \dots, E$, and n_e are the pseudo data set sizes.

Note that the proposed prior is based on the likelihood $p_\theta(\cdot)$ of the Bliss model for the sake of the posterior tractability. Thus in some sense, we model pseudo data by using (1.2), as for observed data. A potential comment is how might the Bliss equation (1.2) model the pseudo data ? As suggested in Grollemund et al. (2018), the Bliss model is a straightforward description of the functional covariates' impact on the outcomes. Indeed, the real outcomes (y_i) are only determined by average values of functional covariates (x_i) on some undetermined periods. We claim that it is also a suitable model to describe the pseudo data. Moreover it corresponds to the experts' way of thinking, since experts imagine pseudo truffle productions by considering the rainfall events during the assumed main periods. They surely do not investigate all the intricate details of the rainfall evolution. Therefore, we assume that the following is a reasonable approximation, for $\boldsymbol{\theta} = (\mu, b, \sigma^2, \mathcal{I})$:

$$y_i^e | \boldsymbol{\theta} \overset{\text{ind}}{\sim} \mathcal{N}\left(\mu + \boldsymbol{x_i^e}(\boldsymbol{\mathcal{I}})^T \boldsymbol{b}\ ,\ \sigma^2\right), \text{ for } i = 1, \dots, n_e. \tag{2.7}$$

Then, the posterior distribution given observed data and for the proposed prior (2.6) is given by

$$\begin{aligned} \pi(\boldsymbol{\theta} | \boldsymbol{y}, \boldsymbol{y^1}, \dots, \boldsymbol{y^E}; w) &\propto \prod_{i=1}^{n} p_{\boldsymbol{\theta},i}(y_i) \times \prod_{e=1}^{E} \prod_{i=1}^{n_e} p_{\boldsymbol{\theta},i}(y_i^e)^{w_i^e} \times \pi_0(\boldsymbol{\theta}) \qquad (2.8) \\ &\propto \left(\sigma^2\right)^{-\frac{1}{2}(n + \sum_{e=1}^{E} n_e w^e + K + 1) - 1} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}\left[\text{RSS} + \sum_{e=1}^{E} \text{RSS}_e + \mu^2 v_0^{-1} + \boldsymbol{b}^T \boldsymbol{\Sigma}(\boldsymbol{\mathcal{I}})^{-1} \boldsymbol{b}\right]\right\} \pi_0(\ell) \end{aligned}$$

where $\text{RSS} = \sum_{i=1}^{n} (y_i - \mu - \boldsymbol{x_i}(\boldsymbol{\mathcal{I}})^{\boldsymbol{T}} \boldsymbol{b})^2$, $\text{RSS}_e = \sum_{i=1}^{n_e} w_i^e (y_i^e - \mu - \boldsymbol{x_i^e}(\boldsymbol{\mathcal{I}})^T \boldsymbol{b})^2$ and w^e is the average of the weights $w_1^e, \dots, w_{n_e}^e$. The expression of $\text{RSS}_e / 2\sigma^2$ suggests an interpretation of σ^2 / w_i^e as the error variance for the i^{th} pseudo data of the e^{th} expert.

Thus, the higher the variance with respect to σ^2, the lower the impact of the pseudo data.

Property 1 *From the posterior (2.8), we note that the sum of the weights* $\sum_{e=1}^{E} n_e w^e$ *is similar to the sample size* n. *Thus, it corresponds to an effective sample size of the pseudo data set. We observe that if* $w_i^e = 1$, *then the* i^{th} *pseudo data of the* e^{th} *expert matters like an observed data. On the contrary, if the weights are all null, the posterior does not depend on the pseudo data.*

Without loss of generality, it is convenient to assume that for a given expert e the pseudo data vectors $\boldsymbol{y^e}$ and $\boldsymbol{x^e}(\cdot)$ are centered with regards to the relative weights, then the posterior expectation of $\boldsymbol{b}$ given the intervals $\boldsymbol{\mathcal{I}}$ can be written as

$$\mathbb{E}(\boldsymbol{b}|\boldsymbol{y}, \boldsymbol{y^1}, \ldots, \boldsymbol{y^E}, \boldsymbol{\mathcal{I}}) = \hat{\boldsymbol{b}}_{\boldsymbol{1}} + \sum_{e=1}^{E} \hat{\boldsymbol{b}}_{\boldsymbol{2,e}} \tag{2.9}$$

where

$$\begin{aligned}
\hat{\boldsymbol{b}}_{\boldsymbol{1}} &= \boldsymbol{M_w}^{-1} \boldsymbol{x}(\boldsymbol{\mathcal{I}})^T \boldsymbol{y} \\
\hat{\boldsymbol{b}}_{\boldsymbol{2,e}} &= \boldsymbol{M_w}^{-1} \boldsymbol{x^e}(\boldsymbol{\mathcal{I}})^T \boldsymbol{W^e} \boldsymbol{y^e} \\
\boldsymbol{M_w} &= \boldsymbol{\Sigma}(\boldsymbol{\mathcal{I}})^{-1} + \boldsymbol{x}(\boldsymbol{\mathcal{I}})^T \boldsymbol{x}(\boldsymbol{\mathcal{I}}) + \sum_{e=1}^{E} \boldsymbol{x^e}(\boldsymbol{\mathcal{I}})^T \boldsymbol{W^e} \boldsymbol{x^e}(\boldsymbol{\mathcal{I}}).
\end{aligned}$$

and $\boldsymbol{W^e}$ is the diagonal matrix of the weights $w_1^e, \ldots, w_{n_e}^e$, see Appendix **??** for some details.

Property 2 *The posterior expectation of* $\boldsymbol{b}$, *given the intervals* $\boldsymbol{\mathcal{I}}$, *splits into a part* $\hat{\boldsymbol{b}}_{\boldsymbol{1}}$ *relative to observed data and parts* $\hat{\boldsymbol{b}}_{\boldsymbol{2,1}}, \ldots, \hat{\boldsymbol{b}}_{\boldsymbol{2,E}}$ *relative to the pseudo data. Moreover, we note that if the weights are null, the matrices* $\boldsymbol{W^e}$ *are null and the posterior expectation matches the standard Bayesian estimator of* $\boldsymbol{b}$, *without prior knowledge.*

We notice from the Properties 1 and 2 that the weights largely drive the posterior distribution, so they are important hyperparameters to tune. Concerning the Power Prior model, the hyperparameter α is similar to the weights we deal with and Ibrahim et al. (2015) sum up several approaches to tune it. For instance, one may put a beta prior on α (Ibrahim and Chen, 2000) or α may be fixed from a theoretical point of view (Ibrahim et al., 2003; Chen et al., 2006; De Santis, 2006). Concerning the Weighted Likelihood approach, it is possible to calibrate the weights to sufficiently downweight the observations that are inconsistent with the assumed model in order to introduce an efficient and robust estimator (see among others Markatou et al., 1998). As for these frameworks, different valid considerations may be investigated to calibrate the weights w_i^e for the proposed prior. Below, we propose an approach that takes into account the special nature of the pseudo data. Indeed, pseudo data cannot be considered as data or historical data.

2.2 Building the Weights

A major feature of the proposed methodology is that there is a weight for each pseudo datum of each expert. As an interesting effect, an expert may provide a nuanced opinion. Indeed, if he is pretty sure that, for a covariate $x_i^e(\cdot)$, the associated outcome should be y_i^e, then he should deem that the weight of the pseudo datum must be close to 1. On the contrary, he also may feel free to provide pseudo data for which he is not entirely sure, by setting the associated weight close to 0. Therefore, we consider below that the weights have to depend on the confidence of the experts about theirs pseudo data. Hence, the experts must provide pseudo data $(y_i^e, x_i^e(\cdot))$ and certainty c_i^e assumed to be in $[0, 1]$. The following rule may help the expert to provide certainty. If $c_i^e = 1$, it means for the expert e that his i^{th} pseudo datum is as *realistic* as an observed datum. On the contrary, if his certainty is close to 0, it means that he is doubtful about his belief. Below, we derive the weights w_i^e's from the experts' certainty and two additional considerations.

A first approach would be to define the weights as $w_i^e = c_i^e$. However, in this case a problem can be the potential redundancy of the prior knowledge of different experts. For example, if experts work in the same research team, they probably collaborate to study the subject matter and *in fine* they may provide similar pseudo data. Thus, it appears important to take the experts dependence structure into account. For this purpose, we define $r_{e,f} \in [0,1]$ the dependency coefficient between the experts e and f and we introduce the following factor to downweight pseudo data related to dependent experts. For each expert e, the associated weights $w_1^e, \dots, w_{n_e}^e$ are reduced by using the factor:

$$\frac{1}{1 + \sum_{f \neq e} r_{e,f}}. \tag{2.10}$$

If the experts are all independent, then $r_{e,f} = 0$, for all e and f, and (2.10) equals 1. On the contrary, if we elicit E completely dependent experts, each weight is divided by E. In this paper, we propose to fix the dependency coefficients $r_{e,f}$ from known interactions between experts. Alternatively one may determine or estimate these coefficients, for instance with regard to the distance between pseudo data from a different expert.

The second consideration is about the total experts' weight which can be higher than the observed data weight. Indeed, if experts provide enough pseudo data, $\sum_{e=1}^{E} n_e w^e$ might be greater than n. In this case, pseudo data are more influent than observed data in the Bayesian inference. We consider in this paper that observed data must prevail in the learning process with regard to prior knowledge. Therefore, we force pseudo data weight ($\sum_{e=1}^{E} n_e w^e$) to be lower than the observed data weight (n) by introducing the following extra factor:

$$\frac{n}{n_e \times E}. \tag{2.11}$$

Finally, considering (2.10) and (2.11) we set the weights as

$$w_i^e = \frac{c_i^e}{1 + \sum_{f \neq e} r_{e,f}} \frac{n}{n_e \times E}. \tag{2.12}$$

In this setting, the weight of an expert ($n_e w^e = \sum_{i=1}^{n_e} w_i^e$) is bounded above by n/E and the maximum weight of the pool of experts ($\sum_{e=1}^{E} n_e w^e$) is bounded above by n. Therefore, we defined the w_i^e's by taking into account the number of experts, the number of pseudo data and the potential dependence between the experts.

3 Building a Prior Distribution using Experts' Knowledge about the Coefficient Function

In this section, we propose another approach to taking experts' prior knowledge into account by directly eliciting expert knowledge about the coefficient function. As described in Section 1, the coefficient function is difficult to elicit because it is not clearly interpretable, even for a statistician. However, we claim that experts may think about some of its simple features that we aim to elicit. For instance, they may have knowledge about 1) intervals for which the covariate may impact (or does not) on the response variable and 2) if the impact is positive or negative. In other words, the experts' knowledge is about the sign of the correlation between y and $x(t)$ for t in a fixed interval, conditionally on the values $x(t)$ on the other intervals.

In mathematical terms, this kind of knowledge corresponds to a sign function $s^\beta(\cdot)$: $s^\beta(t) = \mathbf{1}\{\beta(t) > 0\} - \mathbf{1}\{\beta(t) < 0\}$. Note that $s^\beta(t)$ is simply the sign of $\beta(t)$. Next, consider that we elicit from each expert e, a sign function which we denote by $s_e^\beta(\cdot)$. Furthermore we ask the expert to provide their certainty function $g_e(\cdot)$ in such a way that $g_e(t) \in [0,1]$, since experts may have a specific certainty for each interval. Hence, the collected sign functions and certainty functions represent the experts' knowledge that we propose to incorporate in the Bayesian inference by modifying the Bliss prior distribution of $(\boldsymbol{b}, \boldsymbol{\mathcal{I}})$. Note from (1.3) that the prior on $(\boldsymbol{b}, \boldsymbol{\mathcal{I}})$ is given by:

$$\pi_0(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\sigma^2) \propto |\boldsymbol{\Sigma}(\boldsymbol{\mathcal{I}})|^{-1/2} \exp\left\{ \frac{1}{2\sigma^2} \boldsymbol{b}^T boldsymbolSigmab - a \sum_{k=1}^{K} \ell_k \right\}.$$

Note that hyperparameters $\boldsymbol{b}$ and $\boldsymbol{\mathcal{I}}$ are related to the coefficient function and also related to the sign function. We propose the new prior given by:

$$\pi(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\sigma^2; \tau) \propto \pi_0(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\sigma^2) \times \prod_{e=1}^{E} \exp\left\{ -\tau \times \mathrm{dist}^2(s^\beta, s_e^\beta; g_e) \right\}, \tag{3.1}$$

where τ is a tuning hyperparameter and dist is a distance between sign functions which is weighted by the function g_e. By appending an exponantial kernel, the prior distribution is constrained in such a way that the hyperparameters $(\boldsymbol{b}, \boldsymbol{\mathcal{I}})$ which are 'close' to the experts' sign functions have higher prior densities. Such an idea is used and studied in Duan et al. (2018), but the exponantial kernel in (3.1) can be alternatively viewed as a penalization term and it follows that τ drives the intensity of the penalization.

The proposed prior leads to a new posterior distribution depending on the experts' knowledge and it is given by:

$$\pi(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\boldsymbol{y}, \mu, \sigma^2; \tau) \propto \exp\left\{ -\frac{1}{2\sigma^2}\mathrm{RSS} - \tau \times \sum_{e=1}^{E} \mathrm{dist}^2(s^\beta, s_e^\beta; g_e) \right\} \pi_0(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\sigma^2). \tag{3.2}$$

In Equation (3.2), the initial prior $\pi_0(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\sigma^2)$ is built to be weakly informative. Hence, the posterior distribution of $(\boldsymbol{b}, \boldsymbol{\mathcal{I}})$ is mainly determined through the residual sum of squares and the penalization term $\tau \times \sum_{e=1}^{E} \mathrm{dist}^2(s^\beta, s_e^\beta; g_e)$. Thus, the mass of the posterior distribution of $\beta(\cdot)$ is shrunk around the set of $\beta(\cdot)$ which are in accordance with the experts' sign functions $\bar{s}_E^\beta$.

3.1 The choice of the distance

Many choices are possible for the distance and they lead to different inferences. For simplicity, we consider the following weighted L^2 distance which eases some computations:

$$\mathrm{dist}^2(s^\beta, s_e^\beta; g_e) = \int_0^1 \left(s^\beta(t) - s_e^\beta(t)\right)^2 \times g_e(t)\,\mathrm{d}t.$$

By considering this choice, the sum of the distances in (3.1) can be rewritten as the distance between β_s and an average sign function $\bar{s}_E^\beta$ where

$$\bar{s}_E^\beta(t) = \sum_{e=1}^{E} s_e^\beta(t)\frac{g_e(t)}{\bar{g}_E(t)} \quad \text{and} \quad \bar{g}_E(t) = \sum_{e=1}^{E} g_e(t).$$

One of the consequences is a more concise expression of the posterior (3.2):

$$\pi(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\boldsymbol{y}, \mu, \sigma^2; \tau) \propto \exp\left\{-\frac{1}{2\sigma^2}\mathrm{RSS} - \tau \times \mathrm{dist}^2(s^\beta, \bar{s}_E^\beta; \bar{g}_E)\right\} \times \pi_0(\boldsymbol{b}, \boldsymbol{\mathcal{I}}|\sigma^2). \quad (3.3)$$

Furthermore, this choice induces a simplification in the implementation and reduces the computational time. Indeed, one computation of the distance is required in (3.3) instead of E computations of distance in (3.2).

3.2 Tuning the hyperparameter τ

Concerning the regularization hyperparameter τ, we do not have in practice any prior information concerning the regularization hyperparameter τ. Moreover there is not a canonical manner to fix it. Below, we propose an approach to tackle the tuning of τ. We consider τ as a non-random quantity and we use a Bayesian Cross-Validation procedure to fix it.

Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_K)$ be a grid of given values which each leads to a different model $\mathcal{M}_k$, for $k = 1, \ldots, K$. Bayesian Cross-Validation procedure selects the τ_k (or the model $\mathcal{M}_k$) with respect to the approximated utility of a model $\mathcal{M}$ given data $\boldsymbol{D}$:

$$\bar{u}_{\text{IS-LOO}}(\tau|\boldsymbol{D}) = -\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{1}{T}\sum_{t=1}^{T} p_{\boldsymbol{\theta_t}}(y_i|x_i)^{-1}\right), \tag{3.4}$$

for an Importance Sample $\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_T}$ (see Vehtari and Ojanen, 2012 for more details). In our context, we obtain the Importance Sample by using a Sampler algorithm described in Grollemund et al. (2018) and we select τ among $\boldsymbol{\tau}$ maximizing the utility $\bar{u}_{\text{IS-LOO}}(\cdot|\boldsymbol{D})$. More details are presented in Section 5.

4 Implementation

A prior based on pseudo data We use a minor variation of the Gibbs Sampler described in Grollemund et al. (2018) to compute a posterior sample. Indeed, the structure of the posterior distribution is not changed by taking into account pseudo data and thus, only the hyperparameter of the full conditional distributions change. See Appendix **??** for more details and for the expression of the full conditional distributions.

A prior based on expert knowledge about the coefficient function In Section 3, a new prior is introduced with in particular an unusual prior distribution put on $\boldsymbol{b}$, which leads to unusual full conditional distributions. Thus, we do not use a Gibbs Sampler for sampling from the posterior distribution and we propose to rely on a Metropolis-Within-Gibbs algorithm. We change the step of updating $\boldsymbol{b}$ by using the Gaussian random walk proposal: $\boldsymbol{b'} = \boldsymbol{b} + \boldsymbol{\varepsilon}$ where $\boldsymbol{\varepsilon} \sim \mathcal{N}_K(0, \rho \mathbf{Id}_{\boldsymbol{K}})$. The proposal $\boldsymbol{b'}$ is accepted or rejected according to the usual Metropolis acceptance rate. A major difference with the initial Gibbs Sampler is the necessity to tune the random walk scale ρ. In pratice, we run a short version of the Metropolis-Within-Gibbs algorithm a few times with different value of ρ in order to determine which value leads to an average acceptance rate between 0.2 and 0.5. In a different framework, Roberts and Rosenthal (2001) determine that the optimal acceptance rate is 0.234 but according to the literature, the interval $[0.2, 0.5]$ is reasonnable for other frameworks. More details on the proposed algorithm are given in Appendix **??**.

5 Numerical results

5.1 Simulaton Study

In this section, we apply both methodologies on a simulated data set and we study the sensitivity of the proposed methods to the experts' certainty. We simulate a dataset using the FLR model (1.1) for which

- $\mu = 1$,
- $x_i(\cdot)$ is simulated from a Gaussian Process for $i = 1, \ldots, 100$,
- β is a given step function (the dotted line of Plot (d) Figure 1) and

- σ^2 is fixed in such a way that the signal-to-noise ratio is 5.

Thus, we obtain a data set $\boldsymbol{D_0} = (y_i, x_i(\cdot))_{i=1,\dots,100}$. Below, we consider $\boldsymbol{D_0}$ as the observed data set. We refer to Grollemund et al. (2018), Section 2.3, for more details about the simulation prodecure.

Approach 1: a prior based on pseudo data

We simulate pseudo data from two uncorrelated experts. We assume that each expert generates pseudo data using his own model which we suppose to be (1.1) but with a different coefficient function used to generate the simulated observed data set $\boldsymbol{D_0}$. Hence, we use the same model to generate a data set $\boldsymbol{D_1}$ (resp. $\boldsymbol{D_2}$) which we consider as the pseudo data set from expert 1 (resp. 2). For each expert we generate 100 pseudo data and in the first instance we consider that each expert has a certainty 1 to each pseudo datum. We apply the methodology described in Section 2 and Figure 1 presents the results.

Representation of the prior and posterior distributions First, we show with Plot (a) (resp. (b) and (c)) in Figure 1, the posterior distribution of the coefficient function given the pseudo data sets. Even if these plots represent posterior distributions, they can be viewed as prior distributions from experts concerning the coefficient function. Next, we show with Plot (d) in Figure 1 the posterior distribution of the coefficient function given the observed data $\boldsymbol{D_0}$, which corresponds to the posterior distribution without any expert knowledge. Finally, Plot (e) (resp. (f) and (g)) in Figure 1 shows the posterior distribution given the observed data set and the pseudo data sets. As the certainty of the experts are 1 and the experts are fixed uncorrelated, the weight of all the pseudo data equals the weight of the observed data. Thus, for Plot (e) (resp. (f)), the posterior distribution illustrates that expert 1 (resp. expert 2) contributes as much as the obseverd data. For Plot (g), the weight of each expert is half the weight of the observed data.

Note that the posterior expectation (solid line) looks like an average between the expectation given the pseudo data and the expectation given the observed data, which is an illustration of Equation (2.9) and Property 2.

Sensitivity to the certainty. Next, we vary the certainty of the experts from 1 to 0. Figure 2 presents the posterior distributions for different certainty values. Results are illustrations of Property 1 that weights drive the tradeoff between observed data and pseudo data. Indeed, when the weights are null, the posterior expectation given the observed data and the pseudo data, coincides with the posterior expectation given the observed data. Next, when the certainty increases, the posterior expectation progressively tends to the posterior expectation given the pseudo data.

Approach 2: a prior based on the expert knowledge about the coefficient function

We apply the approach described in Section 3 on the simulated data set $\boldsymbol{D_0}$. We simulate the knowledge of two experts by specifying theirs functions: $s_e^{\beta}(\cdot)$ and $g_e(\cdot)$ which correspond to the prior information of the two experts used for the approach 1 in Section 5.1. Figure 3 summarizes the simulated experts' knowledge by showing the related average functions: $\bar{s}_E^{\beta}$ and $\bar{g}$.

Prior knowledge. As Figure 3 illustrates, the prior knowledge is split in three parts: with high certainty the coefficient function is

PK1. positive for $t \in [0.1, 0.35]$,

PK2. null for $t \in [0.35, 0.50] \cup [0.65, 0.8]$ and

PK3. negative for $t \in [0.8, 0.95]$.

Notice that there is no prior knowledge concerning the interval $[0.50, 0.65]$.

Bayesian Cross-Validation. We use the Bayesian Cross-Validation to select a value of τ among a set of values $\boldsymbol{\tau} = (\tau_0, \ldots, \tau_N)$. The lowest value of $\boldsymbol{\tau}$ is fixed at 0 which corresponds to an inference without taking expert knowledge into account. The largest value τ_N is more difficult to determine. As suggested by Equation 3.3, we fix it so that the posterior expectation of $\frac{1}{\sigma^2}$ RSS (model fitting with respect to observed data) and the posterior expectation of $\tau \times \mathrm{dist}^2(s^\beta, \bar{s}^\beta_E; \bar{g}_E)$ (model fitting with respect to pseudo data) are of the same order of magnitude. In this case, we fix $\tau_N = 282.19$.

For each $\tau \in \boldsymbol{\tau}$, we compute the utility $\bar{u}_{\text{IS-LOO}}(\tau)$ by using the approximation (3.4) and we fix the Importance Sample size T to 10000. We select the value $\tau = 84.66$ and Plot (c) in Figure 4 shows the utility for each $\tau \in \boldsymbol{\tau}$.

Results. Plot (a) (resp. (b)) in Figure 4 presents the posterior distribution of the coefficient function when τ is selected by Cross-Validation (resp. when $\tau = 0$). By comparing Plots (a) and (b), we notice the impacts of taking prior knowledge into account. First, we notice from Plot (a) that the posterior distribution is sharper with prior knowledge. Next, the estimate (solid black line) is in accordance with the prior knowledge PK1 to PK3. In particular, taking PK2 into account allows to detect more efficiently the intervals for which the coefficient function is exactly equal to 0.

Moreover, we notice that the estimate is equal to 0 for $t \in [0.45, 0.50] \cup [0.60, 0.80]$ (in accordance to prior knowledge) while the true coefficient function is positive. It is counterbalanced by an overestimation on the interval $[0.50, 0.60]$ for which there is no prior knowledge.

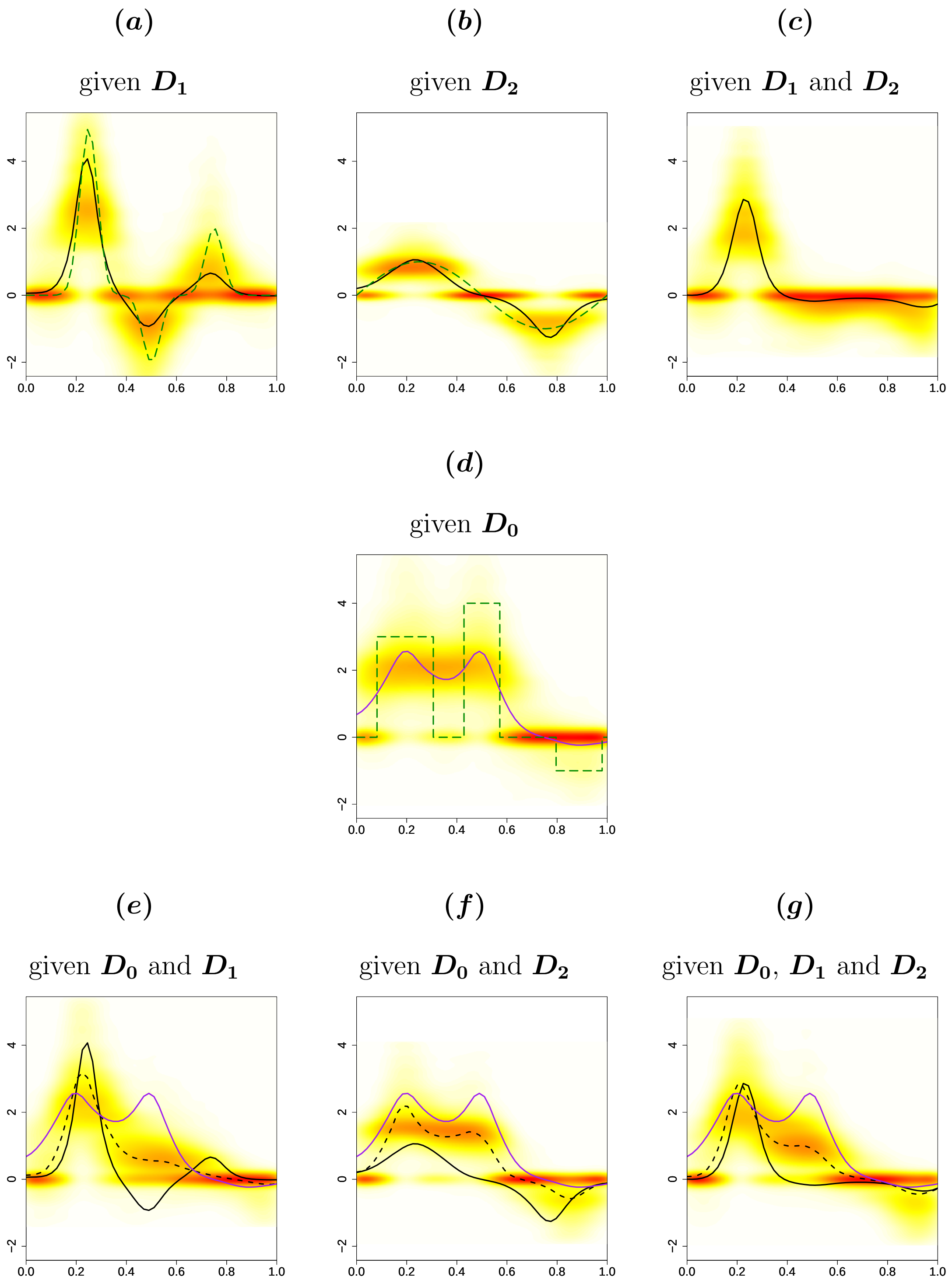

Figure 1: Posterior distributions of the coefficient function given observed data and/or pseudo data. *Plots* (a), (b) *and* (c) *present the posterior distribution given the pseudo data* $\boldsymbol{D_1}$, $\boldsymbol{D_2}$ *and* $(\boldsymbol{D_1}, \boldsymbol{D_2})$. *For these plots, the solid black line is the posterior expectation. Plot* (d) *presents the posterior distribution given the observed data* $\boldsymbol{D_0}$. *For this plot, the solid purple line is the posterior expectation. For Plot* (a) *(resp.* (b) *and* (d)*), the dashed green line is the coefficient function used for simulating the dataset* $\boldsymbol{D_1}$ *(resp.* $\boldsymbol{D_2}$ *and* $\boldsymbol{D_0}$*). Plots* (e), (f) *and* (g) *present the posterior distributions given observed data and pseudo data.*

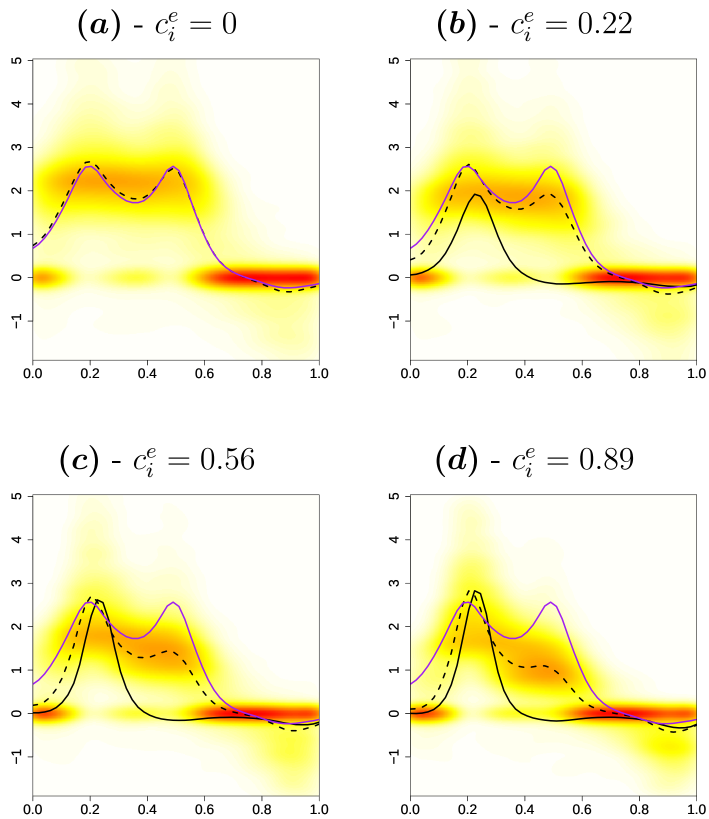

Figure 2: The posterior distribution of the coefficient function for different levels of certainty. *Each plot represents the posterior distribution of the coefficient function given* $\boldsymbol{D_0}$ *(observed data) and* $\boldsymbol{D_1}$, $\boldsymbol{D_2}$ *(pseudo data). The solid purple lines present the posterior expectation given only the observed data* $\boldsymbol{D_0}$*, like in Plot* (d) *Figure* 1*. The solid black lines present the posterior expectation given only the pseudo data* $(\boldsymbol{D_1}, \boldsymbol{D_2})$ *for different values of* c_i^e*. The dashed lines are the posterior expectation given observed data and pseudo data for different values of* c_i^e*.*

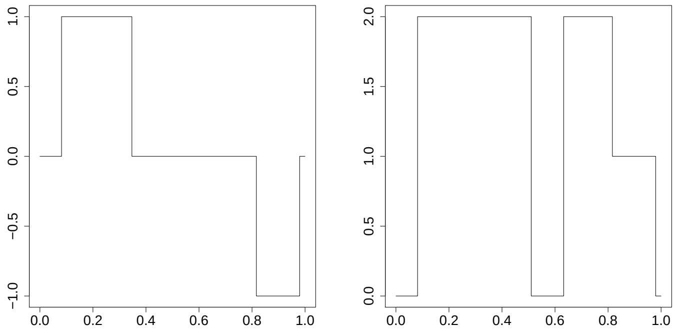

Figure 3: Simulated expert knowledge and certainty by using the elicitation method described in Section 3. *The left (resp. right) plot shows the average sign function* $\bar{s}_E^{\beta}$ *(resp. the average certainty function* $\bar{g}_E$*).*

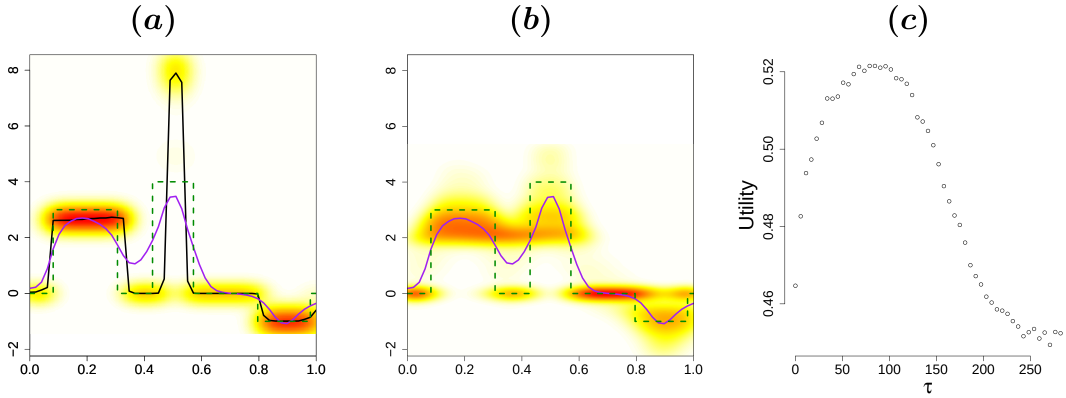

Figure 4: Tuning τ with Bayesian Cross-Validation: Posterior distributions and utilities of τ. *Plot* (a) *(resp.* (b)*) is the posterior distribution of the coefficient function when* τ *is selected by Bayesian Cross-Validation (resp.* $\tau = 0$*). The dashed green line is the step function used for generating the dataset* $\boldsymbol{D_0}$*. The solid purple line is the posterior expectation when* $\tau = 0$*. The solid black line is the posterior expectation when* τ *is selected by Bayesian Cross-Validation. Plot* (c) *presents the approximated utilities* $u_{IS\text{-}LOO}(\tau)$ *with respect to* τ*.*

5.2 Application to the Périgord black truffles dataset

We apply the proposed methologies to investigate the impact of the rainfall on the production of Périgord black truffle from observed data and experts' prior knowledge. During the 20$^{\text{th}}$ century, the production of this valuable mushroom decreased in France from one thousand tons per season to a few dozen tons (Le Tacon et al., 2014). Several reasons have been considered to explain the decrease in the production, such as sociological factors, management of truffle orchards or climatic variations. An important goal is to determine the impact of rainfall on the truffles production in order to guide the truffle farmer and to understand the effects of climate change. Measurements of truffle production are difficult to obtain and in practice very little data is available. In this Section, we study data of an orchard in Pernes-Les-Fontaines (Vaucluse, France) from 1925 to 1949 provided by J. Gravier (see Le Tacon et al., 2017 for a description of the data). The data covers 25 years of production and for each harvest, the monthly rainfall during the truffle life cycle (from January to March of the second year). Below, we address the lack of data information by including expert knowledge in the inference. Truffle experts have studied the life cycle of the truffle and have mainly identified two important periods. First, the truffles born in late spring grow until they become mature in Novembre. During this period, experts think that a hydric deficit is damaging, especially during the summer months. Hence, the first period identified by the experts is the summer months. Secondly, it has been found that water availibity is important to support the initiation of sexual reproduction in the late winter (Le Tacon et al., 2016). Such knowledge will be elicited from a collaboration with F. Le Tacon, C. Murat, P. Montpied (scientists of INRA, Nancy, France), Joël Gravier, Pierre Cunty (truffle farmers) and Michel Tournayre (president of the French Federation of Truffle farmers).

Approach 1: a prior based on pseudo data

Elicitation of pseudo data. As a first step, we need to collect pseudo data from experts. We asked the experts to provide pseudo data. The average number of collected pseudo data is about 3. The elicited pseudo data generally correspond to classical scenarios for them. As the experts' knowledge is surely more complex than 3 data, we pursued to help them to provide more data. Hence, we asked them to give likely response values for 20 given rainfall curves. These curves was taken from different historical data sets and we did not inform them if these curves were observed or created. For each pseudo datum provided, we asked the expert to provide a certainty level, which we denoted by c_i^e and which should be between 0 and 1, for $e = 1, \ldots, 6$ and $i = 1, \ldots, n_e$. We informed the experts that a certainty close to 0 means that the expert doubts his pseudo datum. On the contrary, a certainty close to 1 means that the expert is absolutely sure.

Weight tuning. As described in Section 2.2, we derive weight w_i^e from the experts' certainty and two extra factors. Concerning the first factor, we consider a dependence structure between three experts who belong to the same research team (INRA in Nancy). From discussions with the experts, the similarity between judgements appears. Therefore, we consider that these experts have similar knowledge about the impact of rainfall on truffle production. Hence, we fix a high correlation $r_{e,f} = 0.8$, where e and f are relative to these three experts, and we consider that the other correlations are null. Table 1 shows a sketch of the obtained weights in this way.

Table 1: An example of 5 certainty values and relative weights for each expert. *The weights derive from the certainty by using* (2.12) *and we consider that the correlation between F. Le Tacon and C. Murat and P. Montpied is* 0.8.

F. Le Tacon		C. Murat		J. Gravier		P. Montpied		P. Cunty		M. Tournayre	
$n_1 = 23$		$n_2 = 23$		$n_3 = 26$		$n_4 = 22$		$n_5 = 23$		$n_6 = 22$	
c_i^1	w_i^1	c_i^2	w_i^2	c_i^3	w_i^3	c_i^4	w_i^4	c_i^5	w_i^5	c_i^6	w_i^6
0.8	0.064	0.5	0.04	1	0.16	0.05	0.005	0	0	1	0.189
0.7	0.056	0.5	0.04	1	0.16	0.05	0.005	0.5	0.095	0.8	0.152
0.5	0.04	1	0.079	0.8	0.128	0.05	0.005	0.3	0.057	0.8	0.152
0.3	0.024	0.5	0.04	1	0.16	0.05	0.005	0.5	0.095	0.8	0.152
0.6	0.048	0.5	0.04	1	0.16	0.05	0.005	0.3	0.057	0.7	0.133

Results. Figure 5 shows the results when we take experts' pseudo data into account for investigating the impact of rainfall on truffle production. Plot (b) shows that the coefficient function given the pseudo data, is positive for the late spring and the summer, and it is negative for the late autumn and the winter. During the meeting, the experts largely expressed that the rainfall should have a positive impact on truffle production during summer and should have a negative impact during winter. Hence, we notice that Plot (b) is an accurate representation of the experts' knowledge and that the proposed elicitation is a suitable way to collect it. Concerning the posterior expectation of the coefficient function given the observed data and the pseudo data, we notice that the impact of the summer is again positive. Moreover, given the observed data only, the posterior expectaction from Decembre to March is too close to 0 while the posterior expectation given the observed data and the pseudo data is negative. Thus, taking experts' knowledge into account mainly highlights the impact of rainfall in winter.

Approach 2: a prior based on expert knowledge about the coefficient function

Elicitation of support and sign. During the meeting, we asked each expert e to provide several types of information: a sign function s_e^β (periods and impact of rainfall during these periods) and a certainty function g_e (their certainty for each period). Figure 6 shows the average sign function $\bar{s}_E^\beta$ and we notice that for the experts, summer rainfall events should have a positive impact and should have a negative impact during autumn and winter. In some sense, the average sign function $\bar{s}_E^\beta$ is in accordance with what we obtain from the experts by eliciting pseudo data, see Plot (b) in Figure 5.

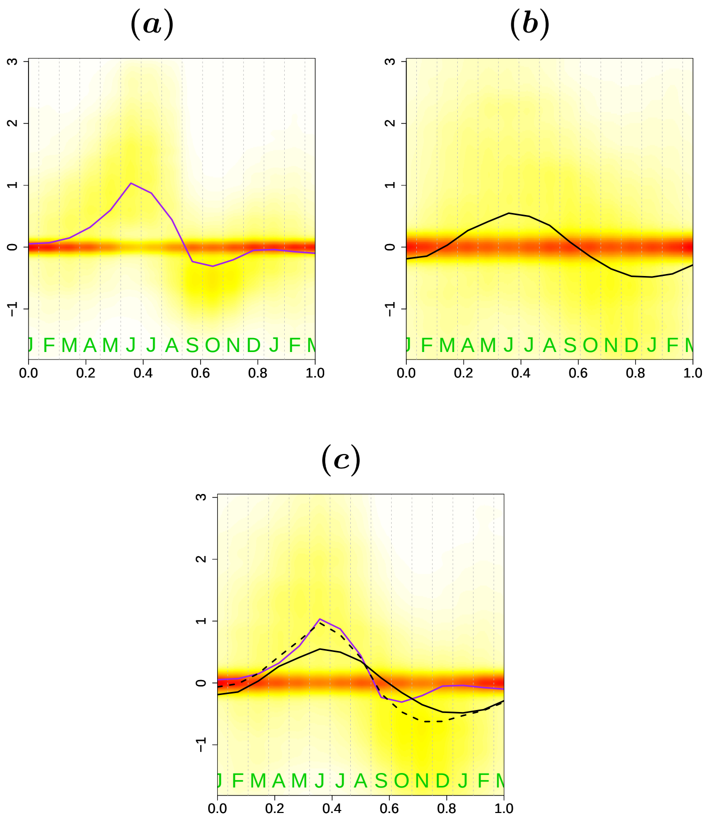

Figure 5: Results obtained with the truffle data and pseudo data. *Plot* (a) *(resp.* (b)*) represents the posterior distribution of the coefficient function given the observed data (resp. the pseudo data). Plot* (c) *shows the posterior distribution given the observed data and the pseudo data. The solid purple (resp. black) line is the posterior expectation given the observed data (resp. pseudo data). The dotted black line is the posterior expectation given the observed data and the pseudo data.*

Results. We apply the approach described in Section 3 to the truffle data set with the previous elicited functions. We consider that τ is not random and we aim to select a value of τ in a grid from 0 to 13.647. The Bayesian Cross-Validation procedure selects $\tau = 8.355$ and Figure 6 shows the posterior distribution of $\beta(\cdot)$ for this value. We notice that the main change due to the experts' knowledge is that the posterior expectation of the coefficient function (dark line) is negative for the winter months

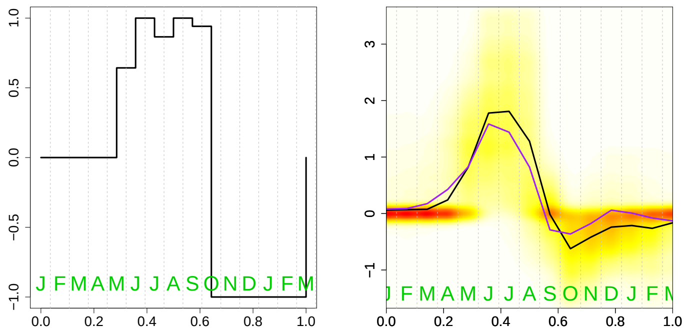

Figure 6: Numerical results for the truffle data set when τ is determined using a Bayesian Cross-Validation procedure. *The left plot shows the average sign function* $\bar{s}^{\beta}_{E}$. *The right plot presents the posterior distribution of the coefficient function. The purple line is the posterior expectation when* τ *is fixed to be 0 and the black line is the posterior expectation when* τ *is selected by Bayesian Cross-Validation.*

while it is close to 0 without expert knowledge (purple line). Moreover, for winter and spring, the posterior density around 0 is higher than the posterior density without the experts' knowledge.

6 Discussion

In this paper, we present two approaches to elicit expert knowledge about Bliss model parameters and to build an informative prior in order to perform a Subjective Inference. Properties of the approaches are studied on simulated datasets, especially for identifying the contribution of experts' prior knowledge on the posterior distribution. Then, we used the proposed methods to study the impact of rainfall on truffle production from observed data and the knowledge of scientists and truffle farmers.

The first approach is based on the elicitation of pseudo data from experts. For experts, the method is quite simple and flexible, allowing experts to provide pseudo data by qualifying their certainty. Hence, it is possible to elicit typical scenarios and other scenarios for which experts are not sure. Elicitation of these different scenarios enables us to collect complete expert knowledge. A major contribution is the proposed prior corresponding to a fractional posterior given pseudo data. Moreover, we observe in practice that prior distribution of the coefficient function accurately corresponds to the experts' knowledge that was verbally expressed. Furthermore, the comparison between the experts' prior distribution, posterior given observed data only and posterior given observed data and pseudo data, enables to understand the impact of the experts' prior knowledge, see an example with Figure 1. The main contribution is the possiblity to drive the impact of the experts' prior knowledge according to the importance of observed data by tuning pseudo data weights. In this paper, we explore tuning weight by taking application considerations into account. For this study, we choose to fix coefficients relative to the dependence structure between experts. One may want to differently fix these coefficients or to tune the weight in a different way. We consider that various successful possibilities may be explored, it depends on the application background and what it is important to take into account. For instance, it would be interesting to estimate the dependence structure by interacting with experts during several meetings or by considering a distance between elicited pseudo data. Moreover, it would be possible to tune weights according to the very different considerations that we explain below. In this paper we consider that the impact of the expert's pseudo data must not exceed the impact of the observed data in the inference but one may consider that experts create or predict pseudo data based on a large range of past experiences. Therefore, in a conceptual sense a pseudo datum is an aggregation of a lot of experiences, and a pseudo datum weight could be greater than an observed datum weight. However, if one wants to change the way to tune weights, it does not change the remainder of the proposed method. As a consequence, this approach seems to be a generic method to build an informative prior distribution from expert knowledge. Therefore, it would be interesting to develop an adaptation of this method to different frameworks and models.

The second approach aims to build a prior from expert knowledge on some interpretable features of the coefficient function. Elicitation is quite simple for experts in this case because support and sign of the coefficient function are the quantities they usually think about when they consider the relationship between rainfall and truffle production. In other words, the method is tailored to match their way of thinking. We built a prior by appending an exponantial term to the inital Bliss prior in order to penalize parameter values which are not in accordance with the elicited supports and signs. As the supplementary term is interpreted as a penalization term, the intensity parameter τ can be handled by using usual methods like a Bayesian Cross-Validation. However the use of this method requires an important computational time in particular to obtain accurate approximations of the utility. Other approaches are possible, such as putting a hyperprior on τ. However, we observed in practice that it was difficult to infer in this way the hyperparameter τ. Concerning the differences with the previous approach, we can see that the impact of a priori knowledge can be different on the posteriori distribution. Moreover, it is complex to determine the contribution of prior knowledge on inference unlike the previous approach.

Acknowledgement

We are very grateful to François Le Tacon, Claude Murat, Joël Gravier, Pierre Montpied, Jean-Luc Dupouey, Pierre Cunty and Michel Tournayre for their valuable contribution.

References

Albert, I., Donnet, S., Guihenneuc-Jouyaux, C., Low-Choy, S., Mengersen, K., and Rousseau, J. (2012). Combining Expert Opinions in Prior Elicitation. *Bayesian Analysis*, **7**.

Bhattacharya, A., Pati, D., and Yang, Y. (2016). Bayesian fractional posteriors. *arXiv preprint arXiv:1611.01125*.

Bousquet, N. and Keller, M. (2018). Bayesian prior elicitation and selection for extreme values. *arXiv preprint arXiv:1712.00685*.

Burgman, M., McBride, M., Ashton, R., Speirs-Bridge, A., Flander, L., Wintle, B., Fidler, F., Rumpff, L., and Twardy, C. (2011). Expert status and performance. *Statistical PLoS ONE*, **6**.

Chen, M.-H., Ibrahim, J. G., et al. (2006). The relationship between the power prior and hierarchical models. *Bayesian Analysis*, **1**(3), 551–574.

Chu, H.-C. and Hwang, G.-J. (2008). A delphi-based approach to developing expert systems with the cooperation of multiple experts. *Expert Systems with Applications*, **34**.

Cooke, R. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science.*

Cooke, R. and Goosens, L. (2000). Procedures guide for structured expert judgment in accident consequence modelling. *Radiation Protection Dosimetry*, **90**.

Crowder, M. (1992). Bayesian priors based on a parameter transformation using the distribution function. *Annals of the Institute of Statistical Mathematics*, **44**(3), 405–416. doi: 10.1007/BF00050695.

Dalkey, N. and Helmer, O. (1963). An experimental application of the delphi method to the use of experts. *Management Science*, **9**.

De Santis, F. (2006). Power priors and their use in clinical trials. *The American Statistician*, **60**(2), 122–129.

Delbecq, A. and Van de Ven, A. (1971). A group process model for problem identification and program planning. *Journal of Applied Behavioral Science*, **7**.

Duan, L. L., Young, A. L., Nishimura, A., and Dunson, D. B. (2018). Bayesian constraint relaxation. *arXiv preprint arXiv:1801.01525.*

Fleishman, E., Nally, R., Fay, J., and Murphy, D. (2001). Modeling and Predicting Species Occurrence Using Broad-Scale Environmental Variables: An Example with Butterflies of the Great Basin. *Conservation Biology*, **15**.

Garthwaite, P., Kadane, J., and O'Hagan, A. (2005a). Statistical methods for eliciting probability distributions.

Garthwaite, P. H. and Dickey, J. M. (1991). An elicitation method for multiple linear regression models. *Journal of Behavioral Decision Making*, **4**(1), 17–31. ISSN 1099-0771. doi: 10.1002/bdm.3960040103.

Garthwaite, P. H. and Dickey, J. M. (1992). Elicitation of prior distributions for variable-selection problems in regression. *Ann. Statist.*, **20**, 1697–1719. doi: 10.1214/aos/1176348886. URL http://dx.doi.org/10.1214/aos/1176348886.

Garthwaite, P. H., Kadane, J. B., and O'Hagan, A. (2005b). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, **100**(470), 680–701.

Genest, C. and Zidek, J. (1986). Combining probability distributions. a critique and annoted bibliography. *Statistical Science*, **1**.

Grollemund, P.-M., Abraham, C., Baragatti, M., and Pudlo, P. (2018). Bayesian functional linear regression with sparse step functions. *Statistics, Methodology arXiv:1604.08403.*

Grünwald, P. (2012). The safe bayesian-learning the learning rate via the mixability gap. In *ALT*, pages 169–183. Springer.

Hoffmann, S. (2017). *Approche hiérarchique bayésienne pour la prise en compte d'erreurs de mesure d'exposition chronique et à faibles doses aux rayonnements ionisants dans l'estimation du risque de cancers radio-induits. Application à une cohorte de mineurs d'uranium.* PhD thesis, Université Paris-Saclay. URL http://www.irsn.fr/FR/Larecherche/Formation_recherche/Theses/Theses-soutenues/PRP-HOM/Documents/2017-These-Hoffmann.pdf.

Hu, F. and Zidek, J. (1995). Incorporating relevant sample information using the likelihood. Vancouver, BC.

Hunns, D. and Daniels, B. (1981). Paired comparisons and estimates of failure likelihood. *Design Studies*, **2**.

Ibrahim, J. G. and Chen, M.-H. (2000). Power prior distributions for regression models. *Statist. Sci.*, **15**(1), 46–60.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2003). On optimality properties of the power prior. *Journal of the American Statistical Association*, **98**(461), 204–213.

Ibrahim, J., Chen, M., Gwon, Y., and Chen, F. (2015). The power prior: Theory and applications. *Statistics in Medicine*, **34**(28), 3724–3749. ISSN 0277-6715. doi: 10.1002/sim.6728.

James, A., Low-Choy, S., and Mengersen, K. (2010). Elicitator: An expert elicitation tool for regression in ecology. *Environmental Modelling & Software*, **25**.

Jenkinson, D. (2005). The elicitation of probabilities - A review of the statistical literature. *Technical Report*.

Kadane, J. and Wolfson, L. (1998). Experiences in elicitation. *The statistician*, **47**.

Kadane, J., Dickey, J., Winkler, R., Smith, W., and Peters, S. (1980). Interactive Elicitation of Opinion for a Normal Linear Model. *Journal of the American Statistical Association*, **75**(372), 845–854.

Kuhnert, P., Martin, T., Mengersen, K., and Possingham, H. (2005). Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. *Environmetrics*.

Kynn, M. (2008). The "heuristics and biaises" bias in expert elicitation. *J. R. Statist. Soc. A*, **171**.

Kynn, M. (2005). *Eliciting Expert Knowledge for Bayesian Logistic Regression in Species Habitat Modelling*. PhD thesis, Queensland University of Technology. URL http://eprints.qut.edu.au/16041/.

Le Tacon, F., Marçais, B., Courvoisier, M., Murat, C., Montpied, P., and Becker, M. (2014). Climatic variations explain annual fluctuations in French Périgord black truffle wholesale markets but do not explain the decrease in black truffle production over the last 48 years. *Mycorrhiza*, **24**, S115–S125.

Le Tacon, F., Rubini, A., Murat, C., Riccioni, C., Robin, C., Belfiori, B., Zeller, B., De La Varga, H., Akroume, E., Deveau, A., Martin, F., and Paolocci, F. (2016). Certainties and uncertainties about the life cycle of the Périgord black Truffle (Tuber melanosporum Vittad.). *Annals of Forest Science*, **73**(1), 105–117.

Le Tacon, F., Murat, C., Gravier, J., Montpied, P., Dupouey, J.-L., Grollemund, P.-M., and Baragatti, M. (2017). Evolution of the prigord black truffle (tuber melanosporum vittad.) production in the vaucluse department (france) from 1903 to 1988. influence of annual climatic variations and possible effects of climate changes or sociological factors.

Low-Choy, S., James, A., and Mengersen, K. (2009). Expert elicitation and its interface with technology: a review with a view to designing Elicitator.

Markatou, M., Basu, A., and Lindsay, B. G. (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, **93** (442), 740–750.

Martin, T., Kuhnert, P., Mengersen, P., and Possingham, H. (2005). The power of expert opinion in ecological models: a bayesian approach examining the impact of livestock grazing on birds. *Ecological Applications*, **15**.

Neal, R. M. (2001). Transferring prior information between models using imaginary data. *Technical Report 0108*.

O'Hagan, A. (1995). Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 99–138.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T. (2006). *Published Examples of the Formal Elicitation of Expert Opinion*, pages 193–216. John Wiley & Sons, Ltd. ISBN 9780470033319. doi: 10.1002/0470033312.ch10. URL http://dx.doi.org/10.1002/0470033312.ch10.

O'Leary, R., Mengersen, K., and Low-Choy, S. (2008). A mixture model approach to representing simple expert information as priors in logistic regression. *Technical report, School of Mathematical Sciences, Queensland University of Technology.*

Ouchi, F. (2004). A Literature Review on the Use of Expert Opinion in Probabilistic Risk Analysis.

Reiss, P., Goldsmith, J., Shang, H., and Ogden, T. R. (2016). Methods for scalar-on-function regression. *International Statistical Review*.

Roberts, G. O. and Rosenthal, J. S. (2001). Optimal scaling for various metropolis-hastings algorithms. *Statistical science*, **16**(4), 351–367.

Vehtari, A. and Ojanen, J. (2012). A survey of bayesian predictive methods for model assessment, selection and comparison. *Statist. Surv.*, **6**, 142–228. doi: 10.1214/12-SS102. URL http://dx.doi.org/10.1214/12-SS102.

Winkler, R. (1967). The Assessment of Prior Distributions in Bayesian Analysis. *Journal of the American Statistical Association*, **62**.

Winkler, R., Hora, S., and Baca, R. (1992). The quality of experts' probabilities obtained through formal elicitation techniques. *Center for Nuclear Waste Regulatory Analyses*.

Wolfson, L. J. and Bousquet, N. (2014). Elicitation. *Wiley StatsRef: Statistics Reference Online*, pages 1–11.

Zellner, A. (1972). On assessing informative prior distributions for regression coefficients.