Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v14n2p318

Detection of outliers with a Bayesian hierarchical model: application to the single-grain luminescence dating method

By Galharret, Philippe, Mercier

Published: 20 November 2021

Electronic Journal of Applied Statistical Analysis
Vol. 14, Issue 02, November 2021, 318-338
DOI: 10.1285/i20705948v14n2p318

Detection of outliers with a Bayesian hierarchical model: application to the single-grain luminescence dating method

Jean-Michel Galharret*[a], Anne Philippe[a], and Norbert Mercier[b]

[a]*Nantes Université, CNRS, Laboratoire de Mathématiques Jean Leray, LMJL, F-44000 Nantes, France*
[b]*Centre de Recherche en Physique Appliquée à l'Archéologie, Université Bordeaux Montaigne (IRAMAT-CRP2A- UMR 5060), F-33600 Pessac, France*

Published: 20 November 2021

The event model was proposed by Lanos and Philippe (2018) to combine measurements in the context of archaeological chronological dating. We extend this model to luminescence dating and define a new strategy to detect outliers from the hyperparameters of the event model. This procedure is applied to the combination of Gaussian measurements, data count and luminescence age estimation. We illustrate through simulations that it is preferable, in terms of accuracy and precision, to exclude detected outliers rather than use the robust estimation method (e.g. the event model).

keywords: application to luminescence dating method; event hierarchical model; outliers.

1 Introduction

There are several Bayesian approaches to addressing the issue of outlying observations. When outliers are present two approaches are possible: the outliers can either be integrated in the modeling or a robust estimation method can be built. In this paper, we propose a new approach for detecting outliers which is applied to calculate an age by the luminescence dating method. Optically Stimulated Luminescence (0SL) is used to date the latest exposure to sunlight of grains extracted from sediments. It is fairly common

*Corresponding author: jean-michel.galharret@univ-nantes.fr

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

that the sample contains poorly bleached grains, i.e. grains that were not sufficiently exposed to light to fully reset. For such a grain the OSL method will provide an older age than the true one. From a statistical point of view, the poorly bleached grains can be viewed as outliers.

Finite mixture distributions are the most studied models for taking into account different types of outliers (see for instance Box and Tiao, 1968; Verdinelli and Wasserman, 1991; Inverardi and Taufer, 2020). One of the components models the non outliers and the others correspond to outliers. The modeling of outliers can be complex due to the wide variety of sources of error and the small number of observations. This approach has been applied for computing ages in archaeology for instance by Bronk Ramsey (2009); Christen and Pérez (2009); Christophe et al. (2018). Note that Bronk Ramsey (2009) uses the mixture model to detect outliers, which are then removed from the sample before age estimation.

In a Bayesian framework, different prior distributions are introduced to obtain a robust model (see West, 1984; Peña et al., 2009; Gagnon et al., 2018). Lanos and Philippe (2017, 2018) considered a hierarchical model to estimate the date of an archaeological event from various dating techniques (radiocarbon, luminescence). This is a particular case of meta-analysis with individual effect for the second stage variance. They showed that this model (called event model) is robust for determinating the age of an event, and for constructing chronologies of archaeological sites. The drawback of robust approaches is the loss of efficiency (in the sense that the variance is not optimal in the absence of outlier).

In this paper we propose a process in two steps. Using the variance parameters in the event model we define a criterion to identify the observations with the worst agreement. Then, we proposed two strategies for estimating the parameter of interest taking into account of detected outliers.

The paper is organized as follows. In Section 2, we describe the methodology for estimating a parameter in the presence of outliers. In Section 3, we apply this method to popular hierarchical models: Normal-Normal model and Poisson counting model. Numerical results are provided to illustrate its performances. An application from the field of archaeology is given in Section 4. We propose a new approach for determinating OSL ages.

2 Methodology

We observe a sample $X_1, ...X_n$ modelled by the Bayesian model. Assume that the likelihood function of $X_1, ...X_n$ belongs to $\{p^{(n)}(\cdot \mid \theta),\ \theta \in \Theta\}$, and we denote by π the prior distribution on θ.

$$X_1, ..., X_n \sim p^{(n)}(\cdot \mid \theta) \tag{1}$$

We consider a hierarchical model parametrization, this is a classical way to include random effects in practice (see for instance Spiegelhalter et al. (2004) for meta-analysis, Congdon (2020); Gelman and Carlin (2014) for numerous models).

$$\begin{aligned} p^{(n)}(X_1,...X_n \mid \theta) &= \int f^{(n)}(X_1,...,X_n,\theta_1,...,\theta_n \mid \theta)\mathrm{d}\theta_1...\mathrm{d}\theta_n \\ &= \int f^{(n)}(X_1,...,X_n \mid \theta_1,...,\theta_n,\theta)\pi_1(\theta_1,...,\theta_n \mid \theta)\mathrm{d}\theta_1...\mathrm{d}\theta_n. \end{aligned}$$

The parameters $\theta_1,...,\theta_n$ are introduced to assist in controlling the heterogeneity of the observations $X_1,...X_n$. We will assume that for all $i \in \{1,...n\}$ the distribution of the X_i depends on θ_i. Individual parameters $\theta_1,...\theta_n$ will bring a possible variability around the central parameter θ. In the other words, the random variables $|\theta_1-\theta|,...,|\theta_n-\theta|$ measure this heterogeneity.

If no information is available to distinguish any of the parameters θ_i from any of the others, we choose a prior distribution on $(\theta_1,\ldots,\theta_n)$ satisfying the property of exchangeability (see Figure 1.A for the representation of this model by a Directed Acyclic Graphs (DAG)):

$$\pi_1(\theta_1,...,\theta_n \mid \theta) = \prod_{i=1}^{n} \pi_{\sigma^2}(\theta_i \mid \theta), \tag{2}$$

$$f^{(n)}(X_1,...,X_n \mid \theta_1,...,\theta_n,\theta) = \prod_{i=1}^{n} f(X_i \mid \theta_i) \tag{3}$$

We parametrize the common conditional distribution of θ_i given θ by a scale parameter σ^2. This unknown parameter controls the heterogeneity in the sample $(\theta_1,...,\theta_n)$, since we have

$$\sigma^2 = \mathrm{Var}(\theta_i-\theta \mid \theta) = \mathrm{Var}(\theta_i \mid \theta).$$

The choice of π_1 given in (2) is not adapted to the presence of outliers in the sample $X_1,...,X_n$. Indeed, the assumption that, conditionally to θ, $\theta_1,...,\theta_n$ are identically distributed is not justified. On the other hand, the assumption of independence can be maintained. Therefore, to take into account the possible presence of outliers, we modify the form of π_1 by including individual scale parameters $\sigma_1^2,...,\sigma_n^2$, for all $i \in \{1,...,n\}$,

$$\sigma_i^2 = Var(\theta_i \mid \theta).$$

More precisely, the prior distribution (2) is replaced by:

$$\pi_1(\theta_1,\ldots,\theta_n \mid \theta) = \prod_{i=1}^{n} \pi_{\sigma_i^2}(\theta_i \mid \theta). \tag{4}$$

The individual random effects bring robustness with respect to outliers, however this strategy usually results in a loss of precision. We propose a new strategy to detect outliers which is based on the unknown parameters $\sigma_1^2,...,\sigma_n^2$. To complete the definition of the Bayesian model, we have to choose a prior distribution for $\sigma_1^2,\ldots,\sigma_n^2$. We assume that $\sigma_1^2,...,\sigma_n^2$ are independent and identically distributed (i.i.d.) from π_s. The choice of π_s depends on the parametrical family $\{\pi_{\sigma^2}(\bullet \mid \theta), \sigma^2 \in \mathbb{R}^+\}$.

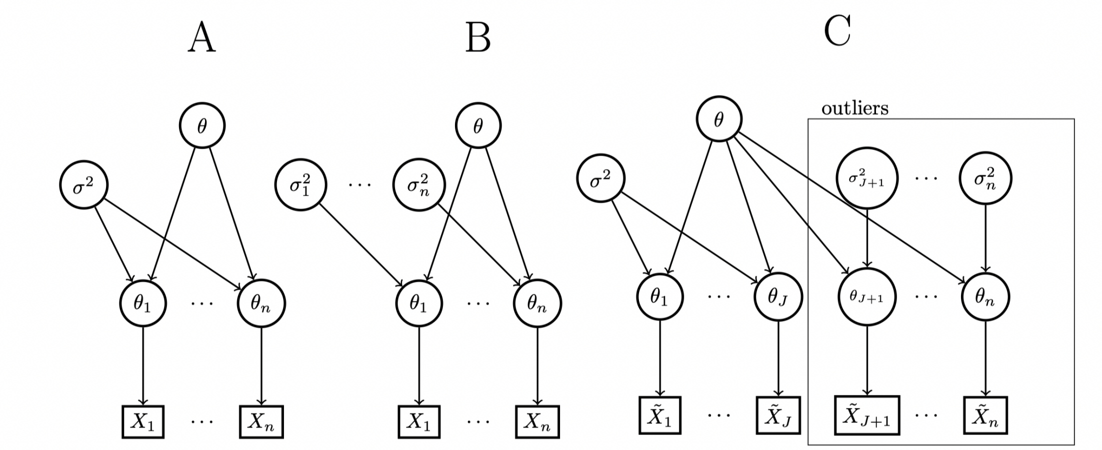

Figure 1: Summary of the models described in Section 2 . A: the DAG of the model (2)-(3) with a common random effect, B: the DAG of the model (4)-(3) with individual random effects and C: the DAG of the model (6)-(7) combining A and B after identification of the outliers. Here J is the sample size of observations non identified as outliers among the n observations.

Decision rule for detecting outliers We seek to identify the outliers among the data. If the observation X_i is an outlier then the variable σ_i tends to infinity as $\mid \theta_i - \theta \mid$ goes to infinity. Using this property we propose a decision rule, based on the variation between the prior and the posterior distributions of σ_i. Let $\alpha \in]0,1[$ and $q_{1-\alpha}$ be the quantile of order $1-\alpha$ of the prior distribution i.e. $\mathbb{P}(\sigma_i > q_{1-\alpha}) = \alpha$. The decision is the following:

$$\text{if } \mathbb{P}(\sigma_i > q_{1-\alpha} \mid X_1, ..., X_n) \geq \alpha, \text{ then the observation } X_i \text{ is an outlier.} \quad (5)$$

i.e. the posterior quantile of order $1-\alpha$ is greater than the prior quantile $q_{1-\alpha}$. According with this decision rule, we denote by $(\tilde{X}_i)_{i\in\{1,...,J\}}$ the resulting subsample of $(X_i)_{i\in\{1,...,n\}}$ whose outliers $(\tilde{X}_i)_{i\in\{J+1,...,n\}}$ have been excluded.

Remark 2.1. *The decision rule* (5) *depends on the posterior distribution of* σ_i*. This distribution will not be usually explicit. For instance, in the Normal-normal discussed in Section 3), we get in* (18) *the joint posterior density of* $\sigma_1, ..., \sigma_n$ *(up to a multiplicative constant) but the marginal densities are not. In practice, the implementation of MCMC algorithm will be required to draw sample from the posterior distribution of the parameters* $\theta, \theta_1, ..., \theta_n, \sigma_1, ..., \sigma_n$*. This simulation step can be easily done using the well-known applications JAGS, STAN, BUGS (see Plummer, 2019; Stan Development Team, 2020; Lunn et al., 2009). All the numerical results provided in Sections 3 and 4 have been performed using STAN software.*

Estimation of θ After detecting outliers, we propose two different strategies to estimate the parameter of interest θ.

[OM-1] The first strategy consists in excluding the outliers to the observations. Then we estimate the parameter of interest θ from the subsample $(\tilde{X}_i)_{i\in\{1,\ldots,J\}}$. As this subsample is assumed to be free of outliers, it is not necessary to include individual random effects. Therefore, we use the model defined in (2)-(3):

$$f^{(J)}(\tilde{X}_1,...,\tilde{X}_J \mid \theta_1,...,\theta_J) = \prod_{i=1}^{J} f(\tilde{X}_i \mid \theta_i) \quad \theta_i \in \Theta$$

$$\pi_1(\theta_1,\ldots,\theta_J \mid \theta,\sigma^2) = \prod_{i=1}^{J} \pi_e(\theta_i \mid \theta,\sigma^2) \tag{6}$$

Although the new estimation is made on a smaller sample, we hope that it will be more precise than the model with individual random effects defined in (3)-(2). This gain in precision would be more particularly important in presence of a small number of outliers.

[OM-2] A second strategy is to add in the model [OM-1] the outliers with individual random effects. Such approach has been addressed by Gumedze and Jackson (2011) in a frequentist context. Our model can be written as follows (see Figure 1.C):

$$f^{(n)}(\tilde{X}_1,...,\tilde{X}_n \mid \theta_1,...,\theta_n) = \prod_{i=1}^{n} f(\tilde{X}_i \mid \theta_i) \quad \theta_i \in \Theta \tag{7}$$

$$\pi_1(\theta_1,\ldots,\theta_n \mid \theta,\sigma_1^2,...,\sigma_J^2,\sigma^2) = \prod_{i=1}^{J} \pi_e(\theta_i \mid \theta,\sigma^2) \prod_{i=J+1}^{n} \pi_e(\theta_i \mid \theta,\sigma_i^2) \tag{8}$$

The individual random effects on the outliers bring robustness by an automatic penalty. Since the individual effects only affect the outliers, we should win in precision with respect to the model (3)-(2). This treatment less drastic of the outliers makes it possible to keep the observations that would be false positives.

Remark 2.2. *In some applications, stronger information is available on the parameters $\theta_1,...,\theta_n$: for instance $\theta_1 = ... = \theta_n = \theta$ (see Luminescence Dating in Section 4). In this case we can assume that $\sigma^2 = 0$ in both models [OM-1] and [OM-2].*

3 Practical application to hierarchical models

In this section, we show how to implement our approach on different well-known hierarchical model. We present in detail the Normal-Normal model which is widely used in applications, as for instance clinical meta-analysis, (see for instance Congdon, 2010; Spiegelhalter et al., 2004; Gelman et al., 2014). For this model, we provide simulation results and an application on real data coming from Radiocarbon dating. Note that from the description of the Gaussian case we can also apply to all probability distributions parametrized by the mean and the scale parameter, as for instance Student, lognormal distributions (see Section 4 for an application with Cauchy distribution for the determination of OSL age). We show also that our approach can be applied with other forms of distributions . We describe the example of the Poisson distribution of which one can find applications in the books previously mentioned for the Gaussian case.

3.1 Hierarchical Normal Model

This hierarchical model is defined as follows

$$\begin{aligned} X_i &= \theta_i + s_i\epsilon_i, \qquad \forall\, i = 1, ..., n \\ \theta_i &= \theta + \sigma_i\rho_i \end{aligned} \tag{9}$$

where $(\epsilon_1, ...\epsilon_n, \rho_1, ..., \rho_n)$ are independent and identically Gaussian distributed random variables with zero mean and variance 1, and where $s_1^2, \ldots, s_n^2$ are known. We denote $s_0^{-2} = \frac{1}{n}\sum_{i=1}^{n} s_i^{-2}$. This model called *event model* was applied by Lanos and Philippe (2017, 2018) to combine repeated measurements or dates in archaeology. This is an extension of the usual meta-analysis model (see for instance Spiegelhalter et al., 2004). The classical choice of prior distribution on the second stage variances $\sigma_1^2, ..., \sigma_n^2$ is the uniform shrinkage distribution, that is: for all $i \in \{1, ..., n\}$

$$\frac{s_0^2}{s_0^2 + \sigma_i^2} \sim \mathcal{U}niform[0, 1], \quad \sigma_1^2, ..., \sigma_n^2 \text{ are independent.}$$

Hereafter, we denote the uniform shrinkage distribution by $\mathcal{S}hrin(s_0^2)$. The median of this distribution is equal to s_0^2, that quantifies measurement error. Therefore, this prior choice ensures equal weight to these measurement errors and the model errors. This choice is also motivated by the fact that this prior distribution has infinite moments. This heavy tail property allows large values for σ_i^2 which can take into account the presence of outliers. The posterior distributions of σ_i^2 have also infinite moments (see Proposition 5.1 in appendix). A major drawback of this choice is that the Bayes estimate under quadratic loss is infinite. Note that our quantile-decision rule is not affected by the absence of posterior moments.

Data simulation with outliers: We illustrate the performance of our method on simulated samples contaminated with outliers. We fix τ the rate of contamination of the sample. We simulate two independent Gaussian samples:

- a sample $X_1, ..., X_{n-[n\tau]}$ of independent random variables from the true model
$$X_i \sim \mathcal{N}(\theta, s_i^2), \quad i = 1, ..., n - [n\tau] \tag{10}$$
- a sample $X_{n-[n\tau]+1}, ..., X_n$ of independent random variables corresponding to outliers
$$X_i \sim \mathcal{N}(\theta + \mu, s_i^2), \quad i = n - [n\tau] + 1, ..., n \tag{11}$$
where $\mu \neq 0$.

As the variances s_i^2 are assumed known, we simulated these positive observations from Gamma distribution with parameter $(a_s, 1)$. All the simulations are done with $a_s = 2$.

Outliers detection: Figure 2 and Table 1 give the performances of our decision rule for detecting outliers. First, we fix μ, τ and we evaluate the sensitivity and the specificity as function of the cut-off α. Recall that sensitivity (respectively specificity) is the ability to detect true outliers (respectively true non-outliers)sensitivity. As usual, we have to find a trade-off between the sensitivity and the specificity. When the cut-off α increases, we improve the detection of outliers but the number of false outliers increases. The value $\alpha = 0.01$ ensures the best trade-off (see Figure 2). However, the estimation of θ using model [OM-1] or model [OM-2], requires good performances in terms of sensitivity since these models are not robust. Therefore, we recommend the choice of $\alpha = 0.05$. Table 1 gives the properties of this choice as function of the values of μ, τ. The specificity is not sensitive to the number and the mean value of the outliers. The sensitivity does not depend of the contamination rate τ and improves as expected when μ increases. These results confirm the good performance of the choice $\alpha = 0.05$. We use this value of cut-off in the rest of the simulations.

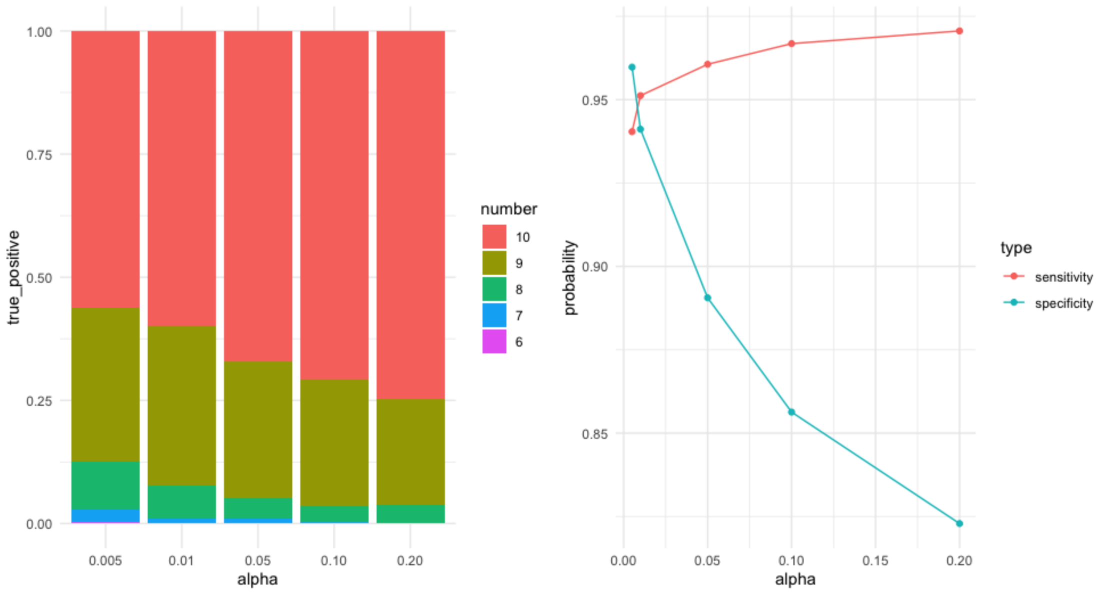

Figure 2: Performances of the decision rule as function of the cut-off α. [Left] Empirical distribution of the number of detected outliers evaluated on simulated sample containing 10 outliers. [Right] Representation of the sensitivity and specificity as function of α. The number of replications is $N = 500$.

Estimation of θ: We compare the estimation of θ obtained by the models [OM-1] and [OM-2] and the *event model* defined in (9). Note that *event model* corresponds to the Gaussian version of the model defined in (3)-(2). To evaluate the precision of the models, we estimate the standard deviation of the posterior distribution of θ and we give informations on the credible interval (length and coverage probability). On the other hand, the mean of the posterior distribution is used to assess the accuracy of the

	$\tau\backslash\mu$	15	10	5
Sensitivity	5%	0.99	0.96	0.81
	10%	0.99	0.96	0.80
	20%	0.99	0.97	0.80
Specificity	5%	0.89	0.89	0.89
	10%	0.89	0.89	0.89
	20%	0.89	0.89	0.89

Table 1: Estimation of the sensitivity and the specificity as function of contamination rate τ and the mean value μ of the distribution of the outliers. The cut-off is fixed to $\alpha = 0.05$, and the number of replications is $N = 500$.

estimations. Figure 3 and Figure 4 and Table 2 provide some properties of the posterior distributions for the three models. In absence of outlier, the performances of the *event model* confirm the loss of precision of robust model (see Figure 3). In this case, even if observations are wrongly removed, [OM-1] provides more precise estimation. Figure 4 represents the posterior mean and the posterior standard deviation of θ as function of the contamination rate τ and the mean value μ of the outlier component in (11). When the contamination rate τ increases, as expected, the precision and the accuracy of the models are deteriorated, however [OM-1] and [OM-2] are less affected by the presence of outliers . The performances of the model [OM-1] remain excellent in all configurations (see Table 2). The identification of outliers is easier for large values of μ (see Table 1, the number of identified outliers tends to τ), this property leads to better results in particular in terms of accuracy. In conclusion, despite a smaller number of observations, the estimation results of model [OM-1] are better than [OM-2] and both are much better than *event model.* Indeed, we get shorter credible intervals and their empirical frequentist coverage is close to the nominal level (see Table 2).

		empirical bias			empirical coverage			credible interval length		
μ	τ	[OM-1]	[OM-2]	event model	[OM-1]	[OM-2]	event model	[OM-1]	[OM-2]	event model
no outlier		0.01	0.02	0.05	0.98	0.98	0.99	0.36	0.36	0.49
5	5	0.01	0.06	0.10	0.98	0.96	0.97	0.47	0.38	0.51
	10	0.03	0.15	0.25	0.97	0.83	0.70	0.41	0.45	0.59
	20	0.01	0.02	0.03	0.98	0.97	1.00	0.36	0.36	0.50
10	5	0.01	0.04	0.07	0.98	0.96	0.99	0.38	0.38	0.51
	10	0.01	0.09	0.15	0.99	0.93	0.91	0.41	0.42	0.56
	20	0.00	0.00	0.00	0.99	0.98	1.00	0.35	0.35	0.48
15	5	0.00	0.01	0.02	0.98	0.99	1.00	0.36	0.36	0.49
	10	0.00	0.02	0.04	0.98	0.98	0.99	0.38	0.37	0.51
	20	0.00	0.05	0.10	0.99	0.97	0.98	0.41	0.41	0.55

Table 2: Length and frequentist coverage probability of 95%-credible intervals calculated on the same simulated samples as Figure 3 and Figure 4.

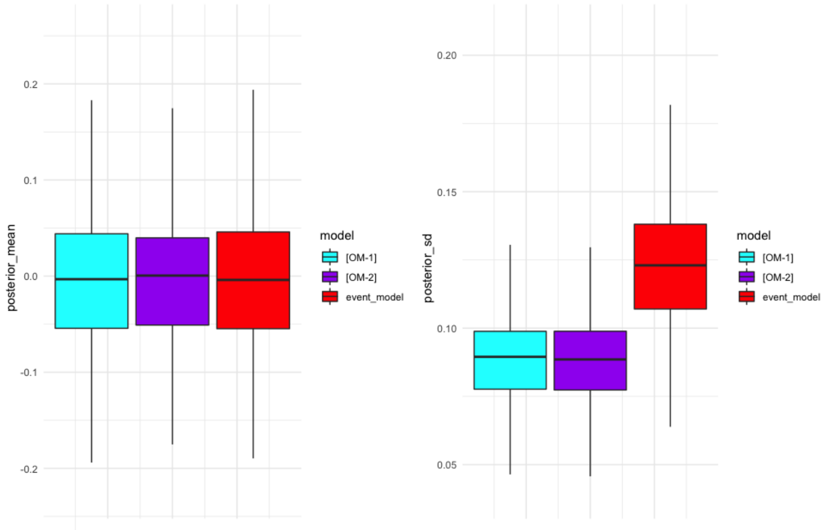

Figure 3: Comparison of the three following models : [OM-1], [OM-2] and the event model on simulated dataset without outlier ($\tau = 0$). We represent the boxplot of the mean (left) and standard deviation (Right) of the posterior distribution of θ. The number of independent replications is $N = 500$. We fix $n = 100$ (sample size) and $\alpha = 0.05$ (cut-off).

3.2 Example in Radiocarbon dating

We analyse the dataset from context X in Tell Qasile provided by Boaretto et al. (2005); Sharon et al. (2007) (see Table 3). All the archaeological samples QS1,...,QS11 are assumed to have the same age. Bronk Ramsey (2009) has already addressed the problem of outliers on this dataset using mixture model. In his approach, the outliers are modeled by a component in a mixture Gaussian model. All the observations have 5% prior probability to be an outlier. His study leads to the following conclusions: QS2 and QS6 are identified as outliers and QS3 is more likely to be an outlier than not.

We compare our methodology to these results. Figure 5 provides the posterior distributions of the scale parameters $\sigma_1, ..., \sigma_{11}$ on which our decision rule (5) is constructed. In Table 3, we provide for each date the posterior probability $\mathbb{P}(\sigma_i > q_\alpha \mid X_1, ..., X_n)$. For $\alpha = .05$, our results confirm the identification of the two outliers QS2 and QS6. Moreover, QS3 is not identified as an outlier but its posterior probability is very close to the cut-off.

Using the model [OM-1] (i.e. after removing QS2 and QS6), the 95%-credible interval of the common age θ is $[2830, 2888]$. Using the *event model* we get the credible region $[2814, 2881]$. This confirms the gain in precision when outliers are removed.

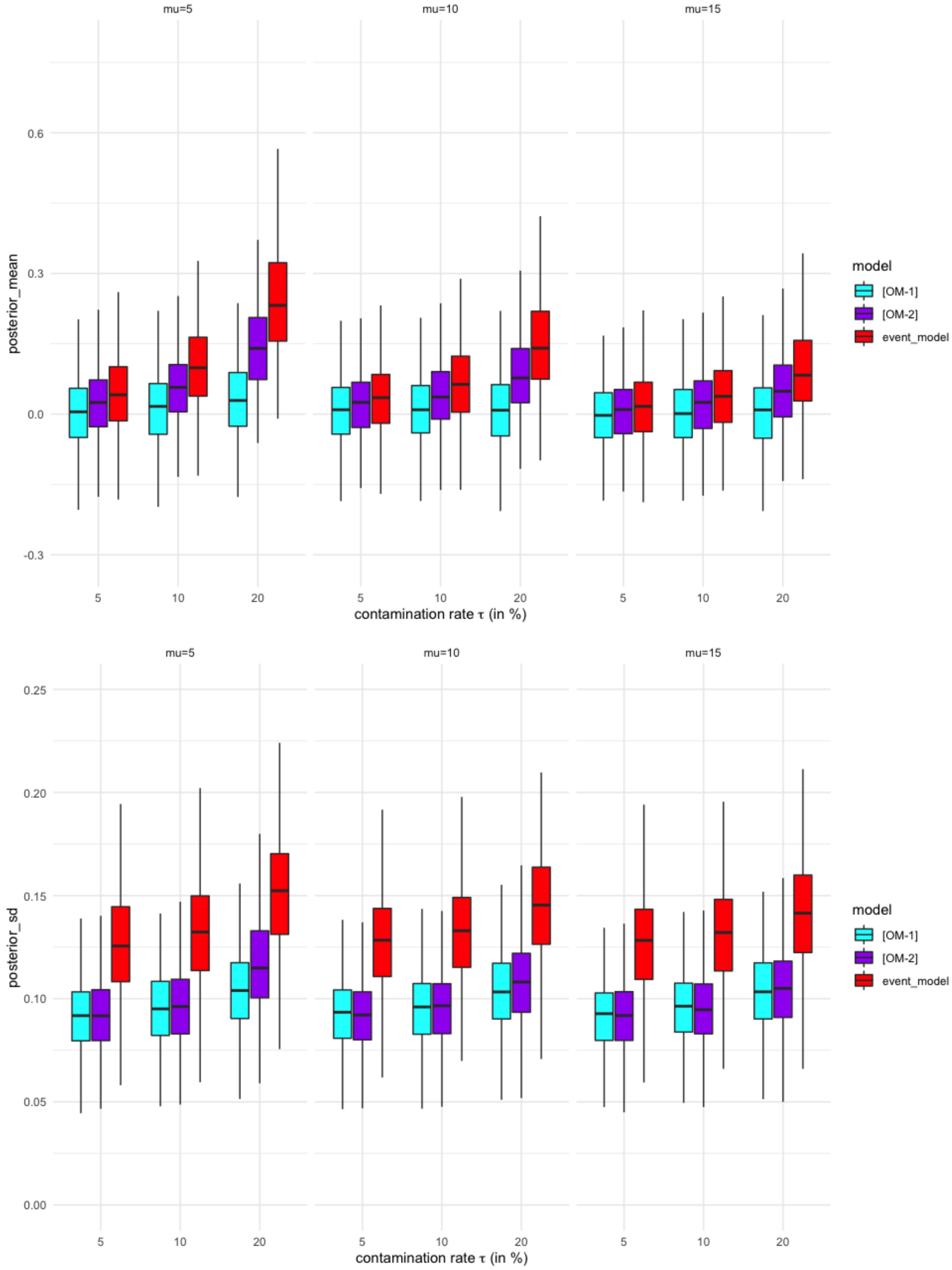

Figure 4: Comparison of the three models [OM-1], [OM-2], event model on simulated dataset with outliers for different values of contaminated rates τ and parameter μ defined in (10) and (11). We represent the boxplot of the mean (Top) and standard deviation (Bottom) of the posterior distribution of θ calculated on $N = 500$ replications. We fix $n = 100$ (sample size) and $\alpha = 0.05$ (cut-off).

method	ident	date X_i	s_i	$\mathbb{P}(\sigma_i > q_{.95} \mid X_1, ..., X_n)$
^{14}C	QS1	2818	26	0.017
^{14}C	QS2	2692	24	**0.358**
^{14}C	QS3	2911	26	0.046
^{14}C	QS4	2853	25	0.016
^{14}C	QS5	2895	25	0.030
^{14}C	QS6	2753	22	**0.128**
^{14}C	QS7	2800	25	0.030
^{14}C	QS8	2882	28	0.020
^{14}C	QS9	2864	40	0.015
^{14}C	QS10	2818	38	0.019
^{14}C	QS11	2897	44	0.023

Table 3: Dates from Tell Qasile X and outputs of the decision rule (5): bold values indicate radiocarbon dates detected as outliers.

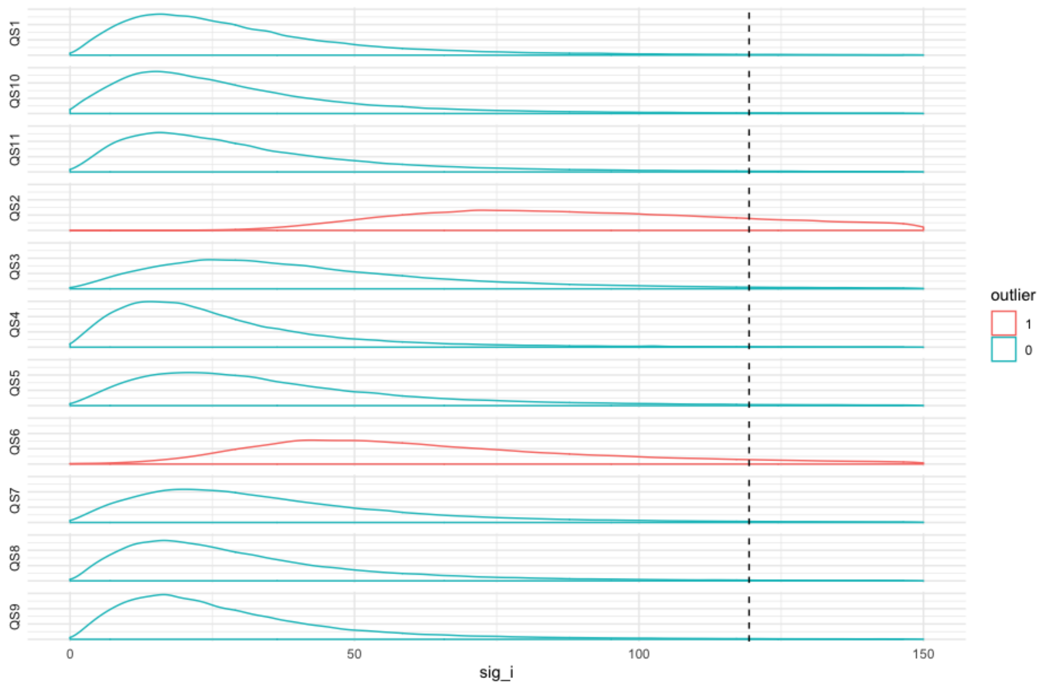

Figure 5: Posterior distribution of the individual scale parameters $\sigma_1, ..., \sigma_{11}$ of QS1,...,QS11 dates.

3.3 Hierarchical Poisson Model

Congdon (2010) suggests the following model for count data in which there is more variability than that predicted by Poisson model. We propose to parametrize this model as follows

$$X_i \sim \mathcal{P}(\theta_i), \qquad \forall\, i = 1, ..., n$$
$$\log(\theta_i) = \log(\theta) + \varepsilon_i$$

where $\mathcal{P}$ denotes the Poisson distribution. As previously, we assume that the errors terms $\varepsilon_1, ..., \varepsilon_n$ have Gaussian independent distributions, $\forall i \in \{1, ..., n\}$,

$$\log(\theta_i) \sim \mathcal{N}(\log(\theta), \sigma_i^2),$$

As in section 3.1, we assume that σ_i^2 has a shrinkage distribution, its parameter is fixed so as not to favor measurement errors over the error between θ_i and θ. In Poisson model, we can evaluate the measurement error by s_0^2 where $s_0^{-2} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i}$. As σ_i corresponds to the dispersion of $\log(\theta_i)$ around $\log(\theta)$, we fix the shrinkage parameter of the form $s_{\log}^2 = s_0^2/\theta^2$, i.e. conditionally to θ,

$$\sigma_i^2 \sim \mathcal{S}hrin(s_0^2/\theta^2), \quad \forall\, i = 1, ..., n$$

Indeed, using the approximation $\theta_i - \theta = \theta(\exp(\varepsilon_i) - 1) \simeq \theta\varepsilon_i$ (for small ε_i), as $\theta\sigma_i$ corresponds the dispersion of θ_i around θ. Thus, the conditional distribution of σ_i^2 given θ is

$$\theta^2\sigma_i^2 \sim \mathcal{S}hrin(s_0^2),$$

in the other words $\sigma_i^2 \sim \mathcal{S}hrin(s_0^2/\theta^2)$.

Remark 3.1. *Note that in the decision rule, we compare as previously* $s_0^2\frac{(1-\alpha)}{\alpha}$ *the prior quantile of* $\theta^2\sigma_i^2$ *with its posterior quantile. From the joint distribution of* (θ, σ_i) *approximated by MCMC experiments, we easily deduce the* $(1-\alpha)$*-quantile of the posterior distribution of* $\theta^2\sigma_i^2$*.*

Numerical results. We show that the decision rule, calibrated in the Gaussian case with the cut-off $\alpha = 0.05$, still provides good performances. We simulate samples contaminated with τ percent of outliers as described in (10) and (11)

- a sample $X_1, ..., X_{n-[n\tau]}$ of independent random variables from the true model

$$X_i \sim \mathcal{P}(\theta), \quad i = 1, ..., n - [n\tau]$$

- a sample $X_{n-[n\tau]+1}, ..., X_n$ of independent random variables corresponding to outliers

$$X_i \sim \mathcal{P}(\theta + \mu), \quad i = n - [n\tau] + 1, ..., n$$

  where $\mu \neq 0$.

Table 4 gives the sensitivity and the specificity as function of the values of μ, τ. As already observed in Table 1, the specificity does not dependent of the number and the mean value of the outliers. In Poisson case, the specificity is very close to 1 which means that almost all the non-outliers are correctly identified. The sensitivity does not depend of the contamination rate τ and improves as expected when μ increases. For $\mu = 5$, the loss of sensitivity is explained by the dissymmetry of the Poisson distribution. The contaminated values keep a high probability in the true model. These results confirm the good performance of the choice $\alpha = 0.05$.

	$\tau \backslash \mu$	15	10	5
Sensitivity	5%	0.98	0.90	0.63
	10%	0.99	0.90	0.66
Specificity	5%	0.99	0.99	0.99
	10%	0.99	0.99	0.99

Table 4: Estimation of the sensitivity and the specificity as function of contamination rate τ and the mean value μ of the distribution of the outliers. The cut-off is fixed to $\alpha = 0.05$, and the number of replications is $N = 500$.

4 Application to the determination of OSL age

4.1 The context

In a sediment, quartz grains are continuously irradiated by cosmic and gamma rays, and by particles (alpha and beta) produced by the decay of natural radioelements. Moreover, they behave as dosimeters in the sense that quartz grains are able to store the radiation doses they receive. The luminescence technique allows determination of the dose (hereafter named "equivalent dose") accumulated by a grain since its last exposure to daylight, before burial. In estimating the dose it received per year, or dose-rate, the time elapsed since the burial of the grain i. e. the Age- can be calculated. In practice, it is common to extract hundreds of grains from a sediment sample and individually determine their equivalent dose, which gives a distribution of discrete values, each one being affected by an experimental error. In parallel, modeling the sediment sample with a software application such as DosiVox-2D (see Martin et al., 2018; Fang et al., 2018), which allows particle-matter interactions with a Monte-Carlo, it is possible to compute the distribution of dose-rates to which each grain have been exposed during its burial. The equivalent doses and dose rates distributions should theoretically be similar (except a scaling factor) but the former distribution may contain outliers corresponding to grains that have not been exposed to daylight. As a consequence, we want to estimate the age A of a sample dated by the luminescence method by considering the distributions of the equivalent dose D and dose rate $\dot{d}$. The fundamental relationship between D and $\dot{d}$ is

$$D \overset{\mathcal{L}}{=} A\dot{d} \tag{12}$$

where $X \overset{\mathcal{L}}{=} Y$ means that both variables X and Y have the same probability distribution.

Modeling of $\dot{d}$: Numerical experiments provide a sample $(\widetilde{\dot{d}_i})_{i\in\{1,...,N\}}$ from the dose rate with a systematic relative error $\dot{\varepsilon}$: $\widetilde{\dot{d}_i} = \dot{d}_i(1+\dot{\varepsilon})$. In our context, the sample is simulated using DosiVox-2D application. On the basis of an expert opinion, we assume that the error $\dot{\epsilon}$ is Gaussian with zero mean and with standard deviation 10%. We fit a parametric model on $(\widetilde{\dot{d}_i})_{i\in\{1,...,N\}}$ defined as follows: for $i \in \{1,...,n\}$

$$\begin{aligned}
\widetilde{\dot{d}_i} &= \dot{d}_i(1+\dot{\varepsilon}) \\
\dot{d}_i &\sim \mathcal{LS}(\mu,\sigma) \\
\dot{\varepsilon} &\sim \mathcal{N}(0, 0.01)
\end{aligned}$$

where $\mathcal{LS}$ denotes a location-scale family of probability distributions. The distribution is parametrized by a location parameter μ and a scale parameter $\sigma > 0$, and the density is of the form

$$x \to \frac{1}{\sigma} f_0 \left(\frac{x-\mu}{\sigma}\right). \tag{13}$$

where f_0 is a probability density. The prior distribution on parameters (μ,σ) is the Lebesgue measure on $[0,+\infty[^2$.

After a preliminar step of the modeling to select f_0 in (13), the estimated parameters on the $(\widetilde{\dot{d}_i})_{i\in\{1,...,N\}}$ will be assumed fixed. We denote $(\dot{\mu}, \dot{\sigma}^2)$ the value of estimates.

Modeling of D: For each quartz grain, its equivalent dose D is measured with a Gaussian error with a known variance. We denoted $(\widetilde{D}_j)_{j\in\{1,...,n\}}$ the observed equivalent doses determined for n grains:

$$\widetilde{D}_j \sim \mathcal{N}(D_j, s^2_{D_j}),\ j \in \{1,...,n\} \tag{14}$$

where $s^2_{D_1}, ..., s^2_{D_n}$ are known and determined experimentally.

Each equivalent dose D_j is associated with an age A_j. According to (13), the conditional distribution of D_j given A_j is

$$D_j \sim \mathcal{LS}\left(A_j\dot{\mu}, A_j^2\dot{\sigma}^2\right),\ j \in \{1,...,n\} \tag{15}$$

The parameters of the distribution of $\dot{\varepsilon}$ are calibrated with prior information given by the laboratory.

Then we use the same hierarchical event model defined in (9) to combine the ages $A_1, \ldots, A_k$. Let A be the age of the target event, i.e. the burial time of the sediment sample. The Bayesian is defined by the DAG in Figure 6 and the following conditional distributions:

$$\begin{aligned}
A_j &\sim \mathcal{N}(A, \sigma_j^2),\ j \in \{1,...,n\} \\
\sigma_j^2 &\sim \mathcal{S}hrin(s_0^2),\ j \in \{1,...,n\} \\
A &\sim \mathcal{U}niform[\underline{A}, \overline{A}]
\end{aligned}$$

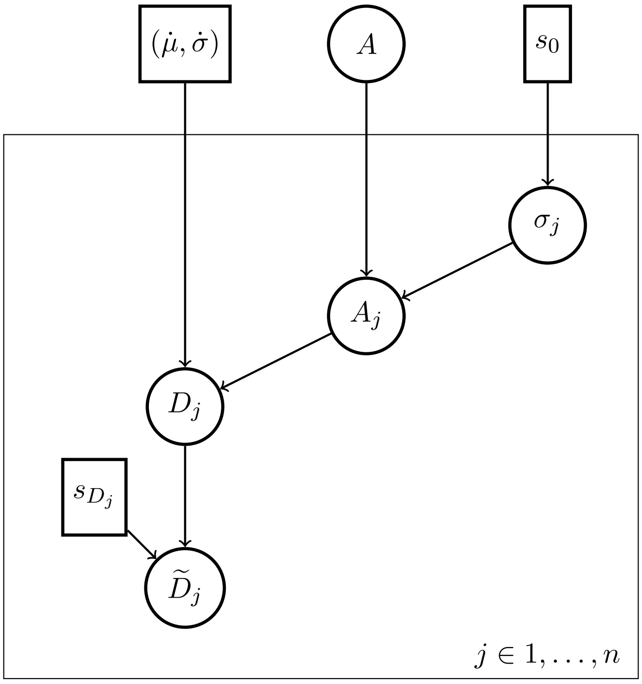

Figure 6: DAG of the model defined in (14) and (15)

where:

- The interval $[\underline{A}, \overline{A}]$ designates the study period,

- Since the event model is applied latent variables $A_1, ..., A_n$, a preliminary estimation of the A_j is required to calculate s_0^2. In such situations, Lanos and Philippe (2017, 2018) suggest to estimate $A_1, ..., A_n$ using n individual models, for $j \in \{1, ..., n\}$:

$$\begin{aligned} \widetilde{D}_j &\sim \mathcal{N}(D_j, s_{D_j}^2), \\ D_j &\sim \mathcal{LS}\left(A_j\dot{\mu}, A_j^2\dot{\sigma}^2\right), \\ A_j &\sim \mathcal{U}niform[\underline{A}, \overline{A}] \end{aligned}$$

  Then, we take s_0^2 equal to the harmonic mean of the variances $Var(A_j \mid \widetilde{D}_j)$.

As described in Section 2, we detect outliers from the posterior distribution of $\sigma_1, ..., \sigma_n$. We identify these outliers according to the decision rule (5). We re-estimate the age A from the subsample $(\widetilde{D}_j)_{j \in J}$ remaining after eliminating outliers. In this particular case, we know that all the dated sample have the same age $A_1 = ... = A_J = A$. Therefore, we adapt model [OM-1] as follows:

$$\begin{aligned}\widetilde{D}_j &\sim \mathcal{N}(D_j, s^2_{D_j}),\\ D_j &\sim \mathcal{LS}\left(A\dot{\mu}, A^2\dot{\sigma}^2\right),\\ A &\sim \mathcal{U}niform[\underline{A}, \overline{A}]\end{aligned}$$

To approximate the posterior distribution of all the parameters, an MCMC algorithm is required. In the following application, we use Stan software in the R interface (see Stan Development Team, 2020).

4.2 Data analysis:

We analyse a sample of $n = 384$ equivalent doses. To estimate the age A of this sample, we have a sample of size N values provided by the numerical modeling of the sediment and distributed according to the dose rate $\widetilde{\dot{d}}$. For the parametric modeling of $\widetilde{\dot{d}}$, represented by a parametric family $\mathcal{LS}$, we have selected the Cauchy distributions i.e. $f_0(x) = \frac{1}{\pi}\frac{1}{1+x^2}$ in (13) . The estimated parameters of the Cauchy distribution are $\dot{\mu} = 1.34$ and $\dot{\sigma} = 0.024$. The study period is $[10, 150]$. According to the method described above we obtain the value $s_0 = 4.53$.

Outliers detection: On the equivalent doses sample, we apply our outliers detection methodology. We detect 3% of outliers. Figure 7 represents the posterior mean distribution of the standard deviation σ_j. It can be seen that the standard deviation of the kept doses have the same order of magnitude.

Age estimation: After removing outliers, we estimate the parameter of interest A with the remaining data (associated to the σ_j with the green boxplot in Figure 7). Figure 8 represents the posterior distribution of the sample age. We also compare this density with the posterior distribution of A resulting from the robust method. This comparison shows the interest of our methodology in terms of precision (see also the 95%-CI in the caption of Figure 8).

Validation of the model: To check the choice of the model, we compare the cumulative distribution of $(D_j)_{j\in J}$ with $A \times (\dot{d}_i)_{i\in 1,\dots N}$. Firstly we compare the empirical distribution of $(\dot{d}_i)_{i\in 1,\dots N}$ with the adjusted Cauchy distribution (see Figure 9 [left]). This adjustment is correct in particular around the median. Secondly, we compare the cumulative distribution of the equivalent dose D and $A\dot{d}$. The empirical cumulative distribution of equivalent doses $(D_j)_{j\in J}$ is

$$F_D(t) := \frac{1}{\mid J \mid}\sum_{j\in J} \mathbb{1}_{D_j \leq t} \tag{16}$$

where $\mid J \mid$ denotes the cardinal of J. $F_D(t)$ is a function of these unknown doses: its posterior distribution can be obtained from the posterior distribution of $(D_j)_{j\in J}$. In particular, the Bayes estimate is

$$\mathbb{E}\left(F_D(t) \mid (\widetilde{D}_j)_{j\in J}\right) = \frac{1}{\mid J \mid}\sum_{j\in J} F_{D_j \mid \widetilde{D}}(t).$$

where $F_{D_j \mid \widetilde{D}}$ is the cumulative distribution function of the posterior distribution of D_j. Similarly, the Cauchy distribution of $A\dot{d}$ depends on unknown parameters $(A, \dot{\varepsilon})$. Therefore, the cumulative distribution function $F_{A\dot{d}}$ is an unknown parameter

$$F_{A\dot{d}}(t) = \dot{G}\left(\frac{t}{A}\right) \tag{17}$$

where $\dot{G}$ is the Cauchy cumulative distribution function with location parameter $\dot{\mu}$ and scale parameter $\dot{\sigma}$. Its posterior distribution can be obtained by a change of variables from the posterior distribution of A.

For fixed t, we construct MCMC sample from the posterior distribution of $F_D(t)$ and $F_{A\dot{d}}(t)$, we represent in Figure 9 [right] their 95%−credible intervals. The two intervals coincide over a wide range of values around the median. This confirms the goodness of fit between the distributions of $D, A\dot{d}$. However, a possible way to improve the model is to take into account the asymmetry of the latent variable D.

5 Appendix:

In Proposition 5.1, we show that the unform shrinkage leads to posterior distribution with infinite variance. One way to obtain posterior distribution with finite variances is to use the extension of the uniform shrinkage introduced by Gustafson et al. (2006). It is defined as follows:

$$\frac{C_0^2 + \sigma_1^2}{C_0^2} \sim \mathcal{P}(1, b)$$

where $b \geq 1$. The case $b = 1$ corresponds to the uniform case. We denote this distribution by $\sigma_1^2 \sim \mathcal{S}hrink(b, C_0)$.

The existence of the moments of the posterior distribution depends on the choice of b. Note that, for arbitrary b, we can preserve the same prior information on the median of σ_1^2 by taking:

$$C_0^2 = s_0^2/(2^{1/b} - 1).$$

Proposition 5.1. *Assume the variances σ_i^2 are independent and identically distributed according to the non uniform shrinkage with parameters (b, C_0). Their posterior distribution admit finite moment of order p if and only if $b > p$.*

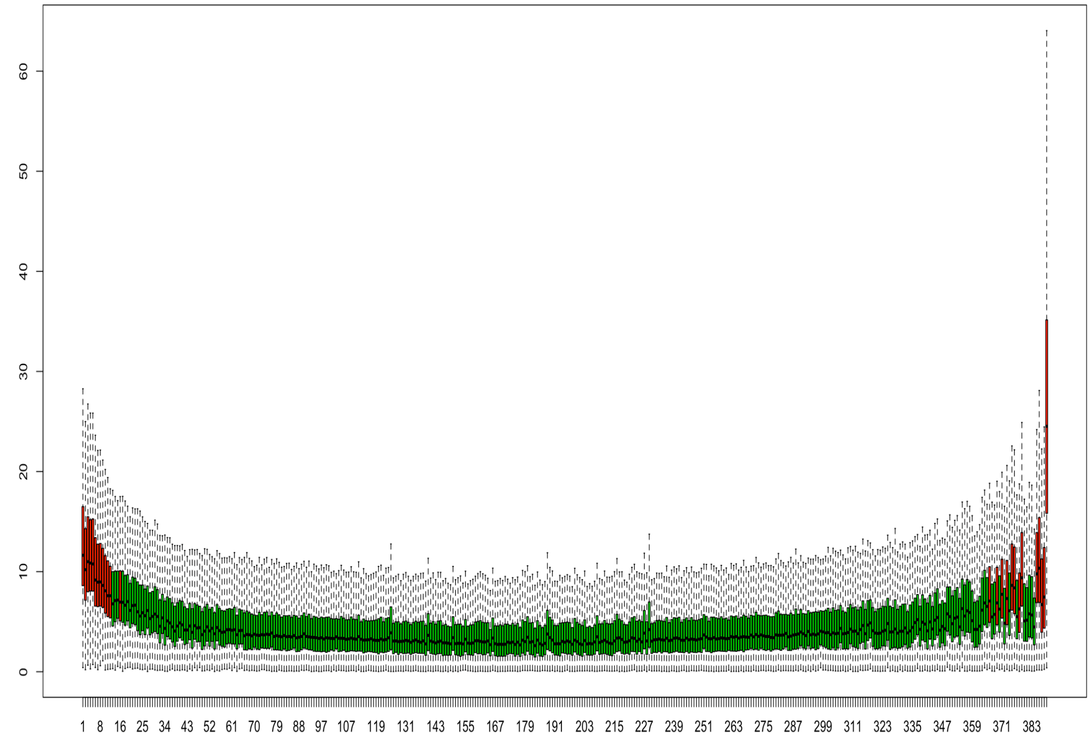

Figure 7: Boxplot of the posterior distributions of σ_j ordered by the posterior median of ages A_j. The red color indicates the detected outliers.

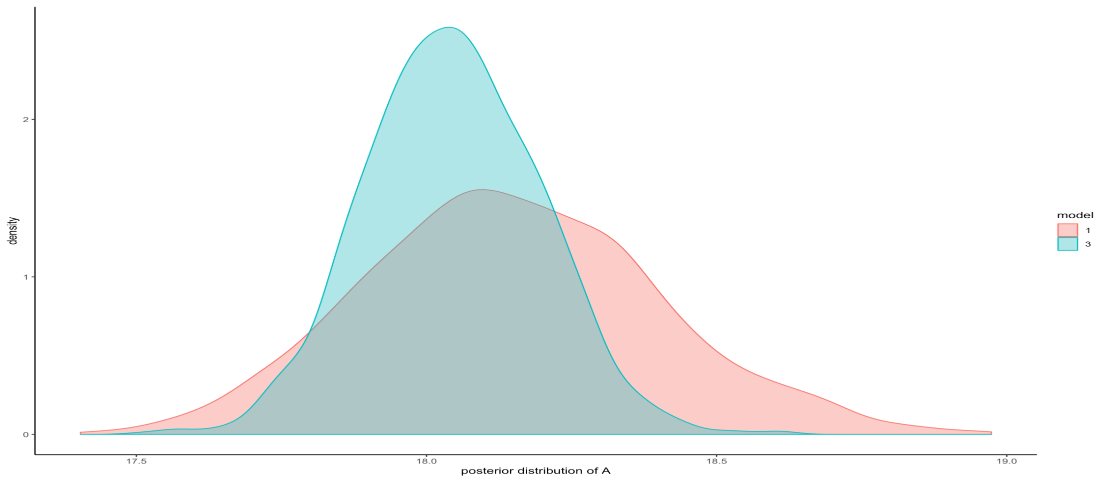

Figure 8: Comparison of the posterior distributions of A: in blue the estimation after removing outliers (posterior mean: 18.06 and 95%-CI= $[17.78, 18.33]$) and in red the robust estimation on the total sample (posterior mean: 18.15 and 95%-CI= $[17.65, 18.64]$).

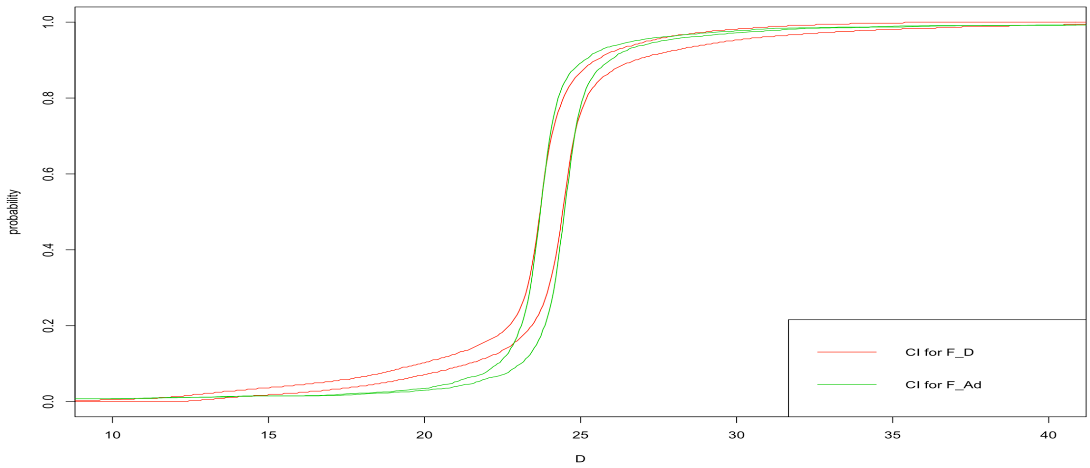

Figure 9: Representation of the 95%−credible intervals of F_D (red),$F_{A\dot{d}}$ (green) defined in (16), (17).

Proof. According to (9), the likelihood of the event model is:

$$f(\underline{X} \mid \theta, \sigma_1^2, \ldots, \sigma_n^2, s_1^2, \ldots, s_n^2) = \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2/(s_i^2+\sigma_i^2)\right)}{(2\pi)^{n/2}\prod_{i=1}^{n}\sqrt{s_i^2+\sigma_i^2}}$$

where $\underline{X} := (X_1, \ldots, X_n)$. The prior distribution is

$$\pi(\theta, \sigma_1^2, ..., \sigma_n^2) \propto \mathbb{I}_{\mathbb{R}}(\theta)\prod_{i=1}^{n}\frac{1}{(C_0^2+\sigma_i^2)^{1+b}}\mathbb{I}_{\mathbb{R}^+}(\sigma_i^2)$$

Note that this prior distribution of σ_i^2 admits moments of order p if and only if $b > p$. After integrating with respect to θ, the posterior distribution of $\sigma_1^2, ..., \sigma_n^2$ is:

$$\pi(\sigma_1^2, ..., \sigma_n^2 \mid \underline{X}) \propto \prod_{i=1}^{n}\frac{1}{(C_0^2+\sigma_i^2)^{1+b}\omega_i^{1/2}}\mathbb{I}_{\mathbb{R}^+}(\sigma_i^2) \quad \frac{1}{\sqrt{\sum_{i=1}^{n}\omega_i}}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\omega_i V(\sigma_1^2, ..., \sigma_n^2)\right) \tag{18}$$

where $\omega_i = \dfrac{1}{s_i^2+\sigma_i^2}$ and $V(\sigma_1^2, ..., \sigma_n^2)$ is the weighted empirical variance:

$$V(\sigma_1^2, ..., \sigma_n^2) = \frac{\sum_{i=1}^{n} X_i^2\omega_i}{\sum_{i=1}^{n}\omega_i} - \left(\frac{\sum_{i=1}^{n} X_i\omega_i}{\sum_{i=1}^{n}\omega_i}\right)^2$$

We have

$$\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\omega_i V(\sigma_1^2,...,\sigma_n^2)\right)=\exp\left(-\frac{1}{2}\sum_{i=1}^{n}X_i^2\omega_i+\frac{1}{2}\left(\sum_{i=1}^{n}X_i\omega_i\right)\left(\sum_{i=1}^{n}X_i\frac{\omega_i}{\sum_{i=1}^{n}\omega_i}\right)\right)$$

For fixed n and $X_1,\dots,X_n$ in $\mathbb{R}$, $\sum_{i=1}^{n}X_i\frac{\omega_i}{\sum_{i=1}^{n}\omega_i}$ is bounded uniformly in $\sigma_1^2,\dots,\sigma_n^2$. Moreover, for arbitrary h function, $\sum_{i=1}^{n}h(X_i)\omega_i$ tends to 0 as $\sigma_1^2,\dots,\sigma_n^2$ tends to infinity. Therefore, the asymptotic behavior of the posterior distribution defined in (18) is

$$\pi(\sigma_1^2,...,\sigma_n^2 \mid \underline{X}) \sim \prod_{i=1}^{n}\frac{1}{(\sigma_i^2)^{b+1}}, \quad \text{as } \sigma_1^2\to\infty,\dots,\sigma_n^2\to\infty.$$

This conclude the proof. □

References

Boaretto, E., Jull, A. J. T., Gilboa, A., and Sharon, I. (2005). Dating the Iron Age I/II Transition in Israel: First Intercomparison Results. *Radiocarbon*, 47.

Box, G. E. P. and Tiao, G. C. (1968). A Bayesian Approach to Some Outlier Problems. *Biometrika*, 55.

Bronk Ramsey, C. (2009). Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon*, 51(3):1023–1045.

Christen, J. and Pérez, S. (2009). A new robust statistical model for radiocarbon data. *Radiocarbon*, 51(3):1047–1059.

Christophe, C., Philippe, A., Guérin, G., Mercier, N., and Guibert, P. (2018). Bayesian approach to OSL dating of poorly bleached sediment samples: Mixture distribution models for dose (MD2). *Radiation Measurements*, 108:59 – 73.

Congdon, P. D. (2010). *Applied Bayesian Hierarchical Methods.* Chapman and Hall/CRC.

Congdon, P. D. (2020). *Bayesian Hierarchical Models: With Applications Using R.* Chapman & Hall/CRC Press/Taylor & Francis Group, 2nd edition edition.

Fang, F., Martin, L., Williams, I. S., Brink, F., Mercier, N., and Grün, R. (2018). 2d modelling: A Monte Carlo approach for assessing heterogeneous beta dose rates in luminescence and ESR dating: Paper II, application to igneous rocks. *Quaternary Geochronology*, 48:195 – 206.

Gagnon, P., Desgagné, A., and Bédard, M. (2018). A New Bayesian Approach to Robustness Against Outliers in Linear Regression. *Bayesian Analysis.* Advance publication.

Gelman, A. and Carlin, J. B. (2014). *Bayesian Data Analysis.* CRC, 3 edition.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis.* Chapman & Hall/CRC Texts in Statistical Science. Chapman and Hill/CRC, 3 edition.

Gumedze, F. N. and Jackson, D. (2011). A random effects variance shift model for detecting and accommodating outliers in meta-analysis. *BMC Medical Research Methodology*, 11:19.

Gustafson, P., Hossain, S., and MacNab, Y. C. (2006). Conservative prior distributions for variance parameters in hierarchical models. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 34(3):377–390.

Inverardi, P. L. N. and Taufer, E. (2020). Outlier detection through mixtures with an improper component. *Electronic Journal of Applied Statistical Analysis*, 13(1):146 – 163.

Lanos, P. and Philippe, A. (2017). Hierarchical Bayesian modeling for combining dates in archaeological context. *Journal de la Société Française de Statistique*, 158:72–88.

Lanos, P. and Philippe, A. (2018). Event date model: a robust bayesian tool for chronology building. *Communications for Statistical Applications and Methods*, 158(2):131–157.

Lunn, D., Spiegelhalter, D., Thomas, D., and Best, N. (2009). The bugs project: Evolution, critique and future directions. *Statistics in Medicine*, 28:3049–3067.

Martin, L., Fang, F., Mercier, N., Incerti, S., Grün, R., and Lefrais, Y. (2018). 2d modelling: A Monte Carlo approach for assessing heterogeneous beta dose rate in luminescence and ESR dating: Paper I, theory and verification. *Quaternary Geochronology*, 48:25 – 37.

Peña, D., Zamar, R., and Yan, G. (2009). Bayesian likelihood robustness in linear models. *Journal of Statistical Planning and Inference*, 139(7):2196 – 2207.

Plummer, M. (2019). *rjags: Bayesian Graphical Models using MCMC*. R package version 4-10.

Sharon, I., Gilboa, A., Jull, A. J. T., and Boaretto, E. (2007). Report on the First Stage of the Iron Age Dating Project in Israel: Supporting a Low Chronology. *Radiocarbon*, 49.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester.

Stan Development Team (2020). RStan: the R interface to Stan. R package version 2.21.2.

Verdinelli, I. and Wasserman, L. (1991). Bayesian analysis of outlier problems using the Gibbs sampler. *Statistics and Computing*, 1:105–117.

West, M. (1984). Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society Series B (Methodological)*, 46.