# Should industrial uncertainty analysis be Bayesian?

Keller Merlin
*EDF R&D, Industrial risk management department*
*6, quai Watier*
*78401 Chatou, France*
*merlin.keller@edf.fr*

Parent Eric
*AgroParisTech, Statistics department*
*16, rue Claude Bernard*
*F-75231 Paris, France*
*eric.parent@agroparistech.fr*

Pasanisi Alberto
*EDF R&D, Industrial risk management departmentalberto.pasanisi@edf.fr*

**Abstract.** *Quantitative simulation is probably the major tool of industrial R&D studies. Computer codes are widely used to predict the behavior and the reliability of a complex system in given operating conditions or, in the design phase, to ensure that it will fulfill given required performances. Whatever their complexity, quantitative models are essentially physics-based and, consequently, deterministic. On the other hand, their inputs and/or the code itself may be affected by uncertainties of various nature which affect the final results and must be properly taken into account. Uncertainty analysis has gained more and more importance in industrial practice in the last years and is at the heart of several working groups and funded projects involving industrial and academic researchers.*

*We present hereby how the common industrial approach to uncertainty analysis can be formulated in a full Bayesian and decisional setting. First, we will introduce the methodological approach to industrial uncertainty analysis. On the other hand, we will recall some features of the Bayesian setting which prove useful in the industrial practice. We will particularly insist on the role of decision theory and we will show some usual misunderstandings and shortcuts when decisional issues are not formally stated.*

*These concepts will be shown by detailing the stakes and the results of some industrial examples.*

*Key words and phrases: Uncertainty analysis; Bayesian inference; Decision theory; heuristic predictive estimation.*

## Introduction

More and more applied industrial studies involve nowadays explicit uncertainty assessment [12]. Indeed, in industrial practice, engineers coping with quantitative predictions using computer models must face numerous and quite different sources of uncertainty, that affect the results of their studies. In most practical problems, the objective is to study the probability distribution of the output of a deterministic model, the inputs of which are random variables. Note that according to this framework, sometimes called *Uncertainty Analysis* [8], we do not tackle some other sources of uncertainties as model inadequacy and the model is assumed to provide exact (deterministic) outputs for given inputs.

Within this paper, we will voluntarily stay into this rather common framework [5]. The ingredients of the problem are then :

- a preexisting physical or industrial system, that is represented by a deterministic model $G(\cdot)$,
- inputs of the physical model, affected by various sources of uncertainty, and modeled jointly as

a random variable,

- outputs of the physical model, *i.e.*, deterministic functions of the inputs,
- an amount of data and expertise available to assess inputs uncertainty,
- industrial stakes that motivate the uncertainty assessment more or less explicitly (safety and security, environmental control, resources optimization etc.).

The industrial stakes suggest the choice of one or more particular *quantities of interest* summarizing the main results of the uncertainty analysis, e.g. the mean and/or a given quantile of the model output.

## Decision theory and Bayes procedures

### General problem formulation

In the following, the model inputs will be modeled jointly as a random variable X, distributed according to a certain probability distribution function (pdf) $f(\cdot|\theta)$, specified by a certain (unknown) parameter vector $\theta$. The model output is noted $\mathrm{Y} = G(\mathrm{X})$. We also assume, as is in general the case, that the data are a sample $D = (x_1, \ldots, x_n)$ of input values, that is, independent realizations of the variable X.

Based on this model, our goal is twofold:

*i.* Infer on the basis of data $D$ the law of X, *i.e.* in our case estimate $\theta$;

*ii.* Deduce an estimate of the pdf of Y, noted $p(\cdot|\theta)$ in the following, and of a certain interest quantity (mean, quantile, etc.) of this law, that is, a certain function $\phi = \phi(\theta)$ of the model parameters.

### Cost function

Because $\phi$ is not known exactly, we must derive from the observation vector $D$ a certain decision, which we note $\delta = \delta(D)$, that we wish to be as close as possible to $\phi$ in a certain sense. This is formalized in the context of decision theory by the concept of *loss* or *cost function*, that is, a certain function $C(\phi, \delta)$ that measures the cost resulting from decision $\delta$ when the interest quantity is $\phi$.

**Usual cost functions** A popular choice of cost function is the quadratic loss, defined by

$$(1) \quad C(\phi, \delta) = (\phi - \delta)^2.$$

Its use implies that the cost is quadratic in the error on $\phi$, and does not depend on its sign. Often a default choice, this loss may be inappropriate when the cost may depend on whether $\phi$ is over or under-estimated. In this case an asymmetric cost function might be preferable, such as the weighted absolute loss:

$$(2) \quad C(\phi, \delta) = |\phi - \delta| \times \{C_1 \mathbf{1}_{\delta<\phi} + C_2 \mathbf{1}_{\delta>\phi}\}.$$

### Bayesian approach

Minimizing the loss $C(\phi, \delta)$ is impossible in practice because $\phi$ depends on $\theta$, which is unknown (otherwise, we would have the trivial solution $\delta = \phi$). Instead, the Bayesian approach consists in describing the uncertainty on the unknown parameter $\theta$ by a prior probability density $\pi(\theta)$, specifying the distribution of all possible values of $\theta$. This prior information is updated using the data to yield the posterior distribution $\pi(\theta|D)$, which is derived according to Bayes' theorem:

$$(3) \quad \pi(\theta|D) \quad \propto \quad f(D|\theta)\pi(\theta).$$

The optimal decision in this setting is obtained by minimizing the *expected posterior loss,*

$$(4) \quad \mathbb{E}[C(\phi, \delta)|D] \quad = \quad \int_{\theta} C(\phi(\theta), \delta)\pi(\theta|D)d\theta.$$

As stressed in [6], this procedure can be interpreted in terms of an *integrated sensitivity analysis*, where each possible cost resulting from decision $\delta$ is weighted according to the probability of such a cost, evaluated conditionally on the available data. The decision that minimizes this quantity is called the Bayes estimator, relative to the prior $\pi$ and the loss $C(\phi, \delta)$. In the following, we will note this quantity: $\widehat{\phi}_{\text{BAY}} = \widehat{\phi}_{\text{BAY}}(D)$.

Besides being intuitive, Bayes procedures are also known to have remarkable frequentist properties, such constituting, together with their limits, the set of all *admissible*, or *non dominated*, decision rules [7]. This means that the choice of an optimal estimator, in the framework of decision theory, is uniquely determined by the problem specification (the cost function), and the amount of prior information available on the uncertain parameter.

### Bayes estimates for simple cost functions

- **Quadratic loss.** It is straightforward to see that the Bayes estimate for the quadratic loss (1) is the posterior mean:

$$\begin{aligned} \widehat{\phi}_{\text{BAY}} &= \mathbb{E}[\phi|D] \\ &= \int_{\theta} \phi\pi(\theta|D)d\theta. \end{aligned} \quad (5)$$

- **Weighted absolute loss.** Under the weighted absolute loss (2), calculations are more involved. In [6], it is shown that the resulting Bayes estimate $\widehat{\phi}_{\text{BAY}}$ is given by:

$$(6) \quad \mathbb{P}[\phi < \widehat{\phi}_{\text{BAY}}|D] \quad = \quad \frac{C_1}{C_1 + C_2}.$$

  In other terms, if $\frac{C_1}{C_1+C_2} = 95\%$ for instance, then $\widehat{\phi}_{\text{BAY}}$ is the 95-th percentile of the posterior density of the interest quantity $\phi$. In the special case where $C_1 = C_2$, the optimal decision is seen to be the posterior median.

### MLE approach

Maximum likelihood estimation (MLE) approach is by far the most popular method of estimation currently used. The MLE of model parameters $\theta$, noted $\widehat{\theta}_{\text{MLE}}$ in the following, is simply the value from wich the data is the most likely to have arisen:

$$(7) \quad \widehat{\theta}_{\text{MLE}} \quad := \quad \arg\max_{\theta} f(D|\theta).$$

This quantity is then substituted to the unknown value of $\theta$ in the definition of $\phi$, yielding the so-called 'plug-in' estimate:

$$(8) \quad \widehat{\phi}_{\text{MLE}} \quad := \quad \phi(\widehat{\theta}_{\text{MLE}}).$$

### Asymptotic properties

This method of estimation is engagingly simple compared to the Bayes procedures described in the previous Sections, as it requires no specification of a cost function or a prior distribution. Furthermore, its use can be justified by its good asymptotic behaviour (see [14] for instance). First, the MLE is known to be *consistent* under very mild assumptions. Secondly, under regularity conditions of the likelihood function $f(D|\theta)$, the MLE is asymptotically normal, for $n$ 'large enough' (in practice, $n = 40$ is considered sufficient in most applications).

However, these good large sample poperties do not constitute a sufficient reason to choose MLE over Bayes estimates, since the latter share this good asymptotic behaviour [14]. Indeed, Doob's theorem ensures that, as soon as the prior density is strictly positive almost everywhere (a.e.), then the posterior distribution concentrates around the true parameter value. Hence, Bayes estimates are generally consistent. Furthermore, under essentially the same regularity assumptions as for the MLE, the posterior distribution of $\phi$ is asymptotically normal. Note that in this large-sample limit, the posterior distribution of $\theta$ becomes independent of the prior distribution, whose influence becomes negligible in regard of the information provided by the data.

### Small sample comparison

The results we have recalled above suggest that, when many observations are available, MLE and Bayes estimates become indistinguishable. Hence, it may be reasonable in this case to use the simpler MLE approach and dispense from the effort of specifying a cost function and a prior distribution.

However, in industrial studies it is often the case that in fact very few observations are at our disposal. This especially the case when predicting the occurence of some extreme event of which by definition very few observations have been made. In such cases the MLE looses all theoretical justification, and may perform very badly for a given loss function when compared to the Bayes estimate, which is allways optimal in the sense described in Section .

In fact, it may happen that the MLE is *inadmissible*, meaning that one can find an estimator that performs always better in terms of risk. The most famous example of MLE inadmissibilility is Stein's paradox [3], and corresponds to a special case where $n \geq 3$ normal means are estimated from $n$ obervations (one per mean) with known variance. More recently, the inadmissibility of the MLE for the shape parameter in the Weibull was proven [10] for the quadratic loss. Indeed, it can be replaced by the Bayesian estimate with the Jeffreys prior, which is uniformly better.

### Example: quantile estimation in the exponential model

As a simple illustration of the concepts described in the previous sections, we consider an exponential lifetime model and assume we wish to estimate the $\alpha$-th quantile $q_\alpha$, such that $P[X > q_\alpha|\theta] = \alpha$. Based on the observed lifetimes $x_1, \ldots, x_n$, this can be done for instance by selecting the empirical quantile $\widehat{q}_\alpha^{\text{EL}}$ of order $\alpha$, that is, the $\lceil n\alpha \rceil$-th largest observation,

$$(9) \quad \widehat{q}_\alpha^{\text{EL}} \quad := \quad x_{(\lceil n\alpha \rceil)},$$

where $x_{(1)} \geq \ldots \geq x_{(1)}$ are the observations, sorted in decreasing order, and $\lceil \cdot \rceil$ is the ceiling function (smallest larger integer).

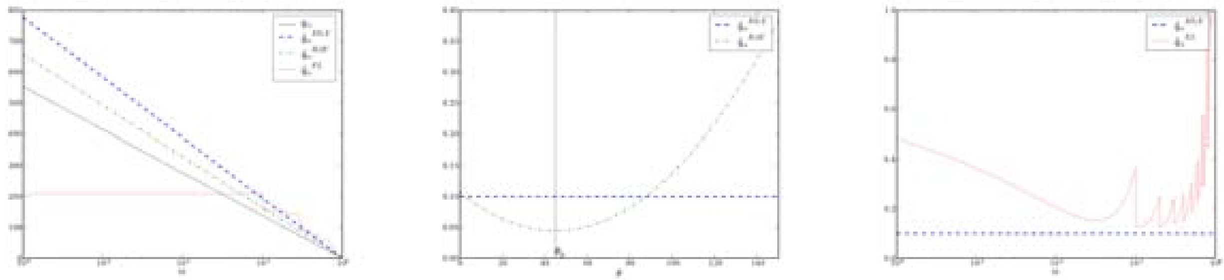

Figure 1: (left:) Estimates of a quantile in the exponential model, as a function of the quantile order. $n = 10$ data where simulated from the $\mathcal{E}(\theta)$ distribution, with $\theta = 60, n_0 = 5, \theta_0 = 45$. Compared relative quadratic risks $\mathcal{R}(\widehat{q}_\alpha, q_\alpha)/q_\alpha^2$ of quantile estimates in the exponential model, for $n = 10$ observations. (center:) MLE vs. Bayes (right:) MLE vs. EL. The prior hyperparameters for the Bayes estimate in this example are $n_0 = 5, \theta_0 = 45$.

Though $\widehat{q}_\alpha^{\text{EL}}$ may seem reasonable, we now show that it is in fact inadmissible for the quadratic loss (1).

To do so, we first consider the MLE. One can easily show that $q_\alpha = \theta \ln \frac{1}{\alpha}$, where $\theta$ is the mean time to failure. The MLE of $\theta$ is simply the empirical mean $\hat{\theta}_{\text{MLE}} = \sum_i x_i/n = \bar{x}$, so:

$$\text{(10)} \quad \widehat{q}_\alpha^{\text{MLE}} = \bar{x} \ln(1/\alpha).$$

Our third candidate is the Bayes estimate, relative to the quadratic loss and to the usual inverse-Gamma prior on $\theta$, that is: $\pi(\theta) = \mathcal{IG}(\theta; n_0 + 1, n_0\theta_0)$. The Bayes estimate relative to the quadratic loss is the posterior mean:

$$\text{(11)} \quad \widehat{q}_\alpha^{\text{BAY}} = \frac{n_0\theta_0 + n\bar{x}}{n_0 + n} \ln(1/\alpha).$$

Here $\theta_0$ can be interpreted as a 'prior guess' on $\theta$, such as given by an expert on the industrial process under study, and $n_0$ as the confidence in the expert, expressed as the number of observations providing an equivalent information. It can also be observed that, in the limiting case $n_0 = 0$, that is, when the expert information is neglected, the Bayes estimate reduces to the MLE.

These estimators are illustrated in Figure 1, on a simulated dataset. We can see that the MLE and Bayes estimators have a similar behavior. In this case the Bayes estimator is slightly closer to the true quantile because it profits from the added prior information. In contrast, the empirical estimator behaves erratically; in particular, it cannot estimate a quantile that is largest than the largest observation, which is why it performs very poorly for small orders $\alpha$.

**Quadratic risks** The risks $\mathcal{R}(\widehat{q}_\alpha, q_\alpha)$ of the above estimates, can be computed explicitly:

$$\text{(12)} \quad \mathcal{R}(\widehat{q}_\alpha^{\text{MLE}}, q_\alpha) = \frac{q_\alpha^2}{n};$$

$$\text{(13)} \quad \mathcal{R}(\widehat{q}_\alpha^{\text{BAY}}, q_\alpha) = q_\alpha^2 \frac{n}{(n+n_0)^2} + q_\alpha^2 \frac{n_0^2(\theta - \theta_0)^2/\theta^2}{(n+n_0)^2};$$

$$\text{(14)} \quad \mathcal{R}(\widehat{q}_\alpha^{\text{EL}}, q_\alpha) = q_\alpha^2 \frac{\sum_{i=\lceil n\alpha \rceil}^n 1/i^2 + (\sum_{i=\lceil n\alpha \rceil}^n 1/i - \ln(1/\alpha))^2}{\ln^2(1/\alpha)}.$$

The quadratic risk of the Bayes estimate is seen to be the sum of two terms, the estimate's variance (systematically lower than that of the MLE, due to the addition of prior information, which counts as

$n_0$ new observations), and the *relative prior bias* (which measures the accuracy of the prior guess $\theta_0$.) Thus, at low sample sizes and if $\theta_0$ is not too far away from the true value $\theta$, the Bayes estimator has a lower risk than the MLE because of the stabilizing effect of the prior, also known as the *shrinking effect*. This is illustrated in Figure 1. When the sample size $n$ becomes very large, one can easily check that the risk (13) of the Bayes estimate becomes equivalent to that (12) of the MLE, illustrating the equivalence of MLE and Bayes approaches for large sample sizes (see Section ).

Contrary to the Bayes estimate, the difference between the risks of the empirical and MLE estimates does not depend on $\theta$, but on the order $\alpha$ of the quantile. This difference is shown in Figure 1, right, for $n = 10$. It is always strictly positive, hence the empirical estimate is inadmissible. This example thus illustrates how seemingly reasonable heuristics can lead to faulty estimation procedures.

## Predictive approach

### The heuristic

In order to account for the uncertainty on the model parameter $\theta$, it is common to replace the unknown pdf $p(y|\theta)$ by the *posterior predictive density:*

$$(15)\quad p(y|D) \quad := \quad \int_\theta p(y|\theta)\pi(\theta|D)d\theta,$$

that is, to average the pdf over all possible values of $\theta$, weighted by their probability given the information provided by the data $D$. Here the output Y is interpreted as a future observation we wish to predict.

A seemingly reasonable use of the predictive density is to estimate any characteristic quantity of the (unknown) density $p(y|\theta)$, such as expected values, tail probabilities, quantiles, etc., by the corresponding characteristic quantity of $p(y|D)$ [5, 12]. Indeed, this provides a generic procedure to estimate any quantity of interest $\phi$. Furthermore, this procedure has the advantage of accounting for the uncertainty on $\theta$, having weighted each of its possible values according to their probability given the data. Variants of the posterior predictive $p(y|D)$ are also commonly used, representing the uncertainty on $\theta$ by other densities than its posterior. These include:

– The prior density $\pi(\theta)$ (see [9] for instance). This yields the *prior predictive density*:

$$(16)\quad p(y) \quad := \quad \int_\theta p(y|\theta)\pi(\theta)d\theta;$$

– The normal $p(\theta) = \mathcal{N}\left(\widehat{\theta}_{\text{MLE}}, \frac{1}{n}\mathcal{I}^{-1}(\widehat{\theta}_{\text{MLE}})\right)$, which is in fact the MLE estimate of the asymptotic approximation to the posterior distribution $\pi(\theta|D)$. Thus it may be used as a convenient substitute when a large number of observations are available. However, we have already stressed that is often not the case in real-life studies. For small samples, this approximation looses all justification, and may even not be well defined.

### Interpreting the predictive density

A first difficulty of the predictive density $p(y|D)$ is that it is very easily misinterpreted. Indeed, it is tempting to view it as the density of a future output Y, computed in a way that somehow deals with the uncertainty on the unknown parameter $\theta$. Though convenient, and suggested by the very name of *predictive density*, this interpretation is in fact misleading. Indeed, the predictive density mixes the *aleatory* uncertainty associated to Y with the *epistemic* uncertainty on the parameter value. As such, it looses all phenomenologic interpretation.

Rather, the predictive density represents one's current state of knowledge concerning the possible values of Y. After all, as stated in [11], 'Bayesian theory only describes how an individual's uncertainties are formed and modified by evidence'.

Thus, when using $p(y|D)$ as a basis for inference, one should refrain from considering it as a substitute to the density $p(y|\theta)$, or from interpreting the results in phenomenologic terms, such as 'given the data $D$, Y has .04 probability of exceeding threshold $t$', since these may lead to erroneous conclusions.

### Estimation of an expected value

We now apply the predictive approach to the estimation of the expected value of the output Y, that is, the integral of Y with respect to its density, so that the interest quantity is:

$$(17) \quad \phi = \mathbb{E}[\mathrm{Y}|\theta] = \int_y y\, p(y|\theta)dy.$$

This is estimated by the *predictive mean*:

$$(18) \quad \widehat{\phi}_{\text{PRED}} = \mathbb{E}[\mathrm{Y}|D] = \int_y y\, p(y|D)dy.$$

However, recalling that $p(y|D)$ is itself defined in (15) as an integral, and exchanging the order of integration,we can re-write the predictive mean as:

$$\begin{aligned} \mathbb{E}[\mathrm{Y}|D] &= \int_\theta \left\{ \int_y y\, p(y|\theta)dy \right\} \pi(\theta|D)d\theta = \int_\theta \phi(\theta)\pi(\theta|D)d\theta \\ (19) \quad &= \mathbb{E}[\phi|D]. \end{aligned}$$

In other terms, in this case $\phi$ is simply estimated by its posterior mean! Furthermore, as recalled earlier, the posterior mean is the Bayes estimate associated to the quadratic loss function. Thus, applying the predictive approach to $\mathbb{E}[\mathrm{Y}|\theta]$, we have in fact performed a Bayesian estimation, implicitly assuming a quadratic loss, which may seem reasonable in this case.

The same reasoning applies in fact to any interest function $\phi$ that can be expressed as the expectation of a function of Y. This includes the expectation of any power of Y, or the probability that Y exceeds a certain value $t$, which can be written as the expectation of $\mathbf{1}_{\{\mathrm{Y}>t\}}$ [4].

This result raises some concerns, since it suggests that the predictive approach heuristic is in fact a Bayesian estimation procedure in disguise. Moreover, this heuristic implicitly forces the choice of a particular cost function, which seems to depend on the expression of the interest quantity, rather than on decisional aspects. For instance, as mentioned earlier , when estimating the probability that an industrial component's lifetime exceeds a certain length, under and over estimations may have very different consequences, so we may want to use a dissymmetric cost function in this case rather than the quadratic loss.

### Estimation of a quantile

Is the predictive approach always equivalent to Bayes estimation under the quadratic loss? Well, the answer is no. Applied to quantiles, it does output a Bayes estimate, but for a radically different cost function. Indeed, if $\phi$ is taken to be the quantile $q_\alpha = q_\alpha(\theta)$ of order $\alpha \in (0, 1)$ of the density of Y, defined by:

$$(20) \quad \mathbb{P}[\mathrm{Y} > q_\alpha|\theta] = \int_{y=q_\alpha}^{\infty} p(y|\theta)dy = \alpha,$$

then it is estimated by the corresponding *predictive quantile* $\widehat{q}_\alpha^{\text{PRED}}$, such that:

$$(21)\quad \mathbb{P}[\mathrm{Y} > \widehat{q}_\alpha^{\text{PRED}}|D] = \int_{y=\widehat{q}_\alpha^{\text{PRED}}}^{\infty} p(y|\theta)dy \;\;=\;\; \alpha.$$

$\widehat{q}_\alpha^{\text{PRED}}$ is seen to be the Bayes estimate of Y relative to the weighted absolute loss:

$$(22)\quad c(y,\delta) \;\;=\;\; |y-\delta| \times \{C_1\mathbf{1}_{\delta<y} + C_2\mathbf{1}_{\delta>y}\},$$

where $C_1, C_2$ are any positive quantities such that $\frac{C_1}{C_1+C_2} = \alpha$. Since this loss is expressed in terms of the output Y rather than $q_\alpha$, we need to integrate it with respect to Y to obtain a proper cost function. This yields the following result:

**Theorem 1 Predictive estimate of a quantile.** *The predictive estimate of the $\alpha$-th quantile $q_\alpha$ of a future observation* Y *is the $\alpha$-th quantile of its predictive density, that is, the Bayes estimate associated with the expected absolute weighted loss:*

$$\begin{aligned} C(q_\alpha,\delta) &= \mathbb{E}[c(y,\delta)|\theta] \\ (23)\qquad &= \int_y c(y,\delta)p(y|\theta)dy. \end{aligned}$$

In contrast to the predictive estimate of an expected value, $\widehat{q}_\alpha^{\text{PRED}}$ cannot be expressed simply in terms of the posterior distribution of $q_\alpha$; hence it is difficult to interpret it directly as a Bayesian estimate of $q_\alpha$.

However, (23) may be reasonable when the future observation Y is central to the decision process. Such a situation is described for instance in [6], in the context of dam conception. Here Y represents the yearly maximal water level of a river, and $\delta$ the height of a dam to be constructed next to the same river.

### Example: Dike reliability estimation

We now compare the MLE, HPE and Bayes estimates on a case-study, concerning the safety evaluation of a flood protection dike. The variable of interest $Y$ is here the maximal water level, noted $Z_c$ in the following, of the river at the location of the dike. Following [1], we assume that $Z_c$ can be computed given a number of input variables, following the analytical formula:

$$(24)\quad Z_c \;\;=\;\; Z_v + C \cdot Q^{3/5},$$

where:

- $Q$ is the yearly maximal water discharge ($\text{m}^3$/s);
- $Z_v$ is the riverbed level (m asl) at the downstream part of the river section under investigation;
- $C$ is a certain constant, depending on the Strickler friction coefficient, as well as the slope, width and length of the river section.

Additionally, we note $d$ the height of the protection dike, and consider the problem of estimating the probability of a flood, that is, the probability that the maximal water level $Z_c$ exceeds $d$. The case presented here is of course an over-simplified toy example, meant for illustrative purposes only, and is not representative of the models used to assess hydrological risks in real-life industrial studies.

### Observation model and prior

Further simplifying the case-study in [1], we assume that $Z_v$ and $C$ are known, fixed quantities. The variable $Q$ on the other hand is intrinsically random, and we suppose that a sample of $n = 50$ annual maximal values $D = (q_1, \ldots, q_n)$ is available, and can be used to assess the uncertainty on $Q$. As is a common approach in extreme value analysis, we chose to model the annual maxima of the river discharge according to the Weibull distribution $\mathcal{W}(\eta, \beta)$, with scale parameter $\eta$ and shape parameter $\beta$. This means that the probability that $Q$ exceeds a certain level $t$ is given by:

$$(25)\quad \mathbb{P}[Q > t|\eta, \beta] \;=\; \exp\left\{-(t/\eta)^\beta\right\}.$$

Given the simplified hydrological model (24) considered here, this means that the flood probability $p$ we are interested is related simply to the parameters $(\eta, \beta)$, following:

$$(26)\quad p = \mathbb{P}[Z_c > d|\eta, \beta] = \mathbb{P}[Z_v + C \cdot Q^{3/5} > d|\eta, \beta] = \exp\left\{-\left(\frac{((d - Z_v)/C)^{5/3}}{\eta}\right)^\beta\right\}.$$

Following [2], we adopt a hierarchical form for the prior density, by defining $\pi(\eta|\beta)$ as a generalized inverse gamma (GIG) distribution, meaning that $\mu = \eta^{-1/\beta}$ follows a gamma distribution conditional on $\beta$. We also choose a Gamma prior for $\beta$, adding a lower bound to ensure existence of the posterior moments of $\eta$ [13], so that:

$$(27)\quad \pi(\mu|\beta) \;=\; \mathcal{G}\left(\mu; m, b(m, \beta)\right)$$

$$(28)\quad \pi(\beta) \;\propto\; \mathcal{G}\left(\beta; m, m/\beta_0\right) \mathbf{1}_{\beta > \beta_l}.$$

Here $b(m, \beta) = \frac{t_e^\beta}{2^{1/m} - 1}$, $t_e$ is a prior guess on the median lifetime, $\beta_0$ is a prior guess on $\beta$, and $m$ is a virtual data size which measures the confidence in the prior information.

**Methods compared.**

We then considered the four following estimates for the tail probability $p$ :

1. The MLE $\hat{p}_{\text{MLE}}$, obtained by substituting the most likely parameter values $(\widehat{\eta}_{\text{MLE}}, \widehat{\beta}_{\text{MLE}})$ to their uknown value in (26);
2. The HPE $\hat{p}_{\text{HPE}}$, which can be seen to be simply the posterior mean $\mathbb{E}[p|D]$ ;
3. We are less interested in the precise value of $p$ than by its order of magnitude, that is, $-\log_{10} p$. Thus, we computed the Bayes estimate $\hat{p}_{\text{BAY}}$ relative to the log-quadratic loss;
4. The above Bayes estimate equally penalizes over and under estimation of the log-tail probability, even though the risks associated to each type of error are very different. Thus we also considered the Bayes estimate $\hat{p}_{\text{BAY},2}$ relative the log-weighted absolute loss, meaning that $-\log_{10} \hat{p}_{\text{BAY},2}$ is the Bayes estimate relative to the weighted absolute loss (2). We chose here to consider that under-estimating the flood probability was 10 times as costly as over-estimating it.

**Results.**

Figure 2 shows the values of the above estimators, when applied to a simulated dataset, the $n = 50$ maximal yearly water discharges $q_1, \ldots, q_n$ being generated from the Weibull distribution with shape parameter $\beta = 2$ and scale parameter $\eta = 1\,000$ (Figure 2, left). We chose a dike height of $d = 53.6$m, so that the true flood probability was $p = .001$.

For this particular dataset, the HPE and the Bayes estimate $\hat{p}_{\text{BAY},2}$ relative to the log-weighted absolute loss over-estimated $p$, *i.e.*, behaved conservatively, whereas the MLE and the Bayes estimate

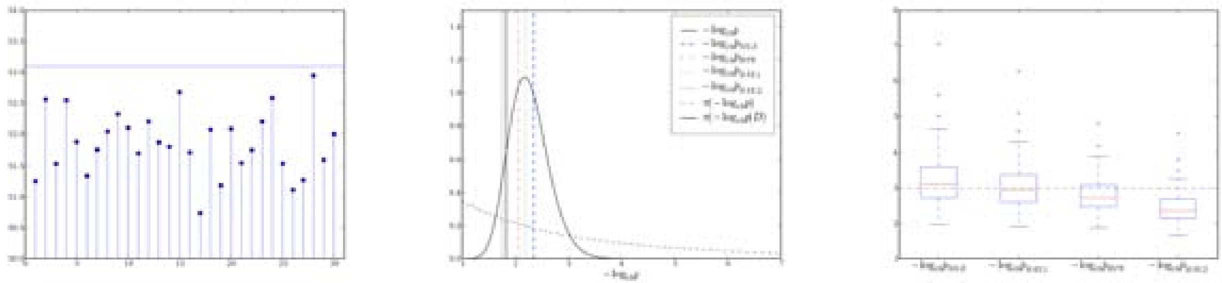

Figure 2: (left:) Simulated annual maxima of river discharges. (center:) Different estimates of a flood probability, on a logarithmic scale. (right:) Different estimates of a flood probability, on a logarithmic scale, over 100 simulated datasets.

$\hat{p}_{\text{BAY},1}$ relative to the log-quadratic loss were anti-conservative. Note that the HPE is in fact always smaller on the logarithmic scale than $\hat{p}_{\text{BAY},1}$. Indeed, due to the concavity of the logarithm, applying Jensen's inequality:

$$(29)\quad -\log_{10} p_{\text{HPE}} = -\log_{10} \mathbb{E}[p|D] \leq \mathbb{E}[-\log_{10} p|D] = -\log_{10} p_{\text{BAY},1}.$$

Otherwise, the above conclusions may depend on the particular datset considered. To check the reproducibility of these results, we computed the above estimators for 100 datasets, all comprising a sample of 50 draws from the same Weibull distribution $\mathcal{W}(2, 1\,000)$. As shown in Figure 2, the MLE and $\hat{p}_{\text{BAY},1}$ frequently under-estimate the flood probability, whereas the HPE and $\hat{p}_{\text{BAY},2}$ are in general conservative, the most conservative estimate being $\hat{p}_{\text{BAY},2}$. This is hardly surprising, since it is based on a loss function that explicitly favours overestimation of the quantity of interest. Thus in a real-life problem, it would appear that the safest estimate to adopt is $\hat{p}_{\text{BAY},2}$, since it gives the best guarantees that the possibility of the dike being flooded will not be under-assessed.

**Discussion**

We have compared the main paradigms used to perform statistical inference in industrial studies, namely Bayesian inference, maximum likelihood estimation (MLE) and predictive inference (PI). This comparison was conducted from the point of view of decision theory.

MLE is by far the most popular approach due to its simplicity, however it is justified only when a large number of observations are available. Its main weakness is that ignores the uncertainty on the unknown parameter, making it very unstable when only a few observations are available.

PI represents a definite improvement over the MLE in that it explicitly accounts for parameter uncertainty by averaging the output's density over all possible values of the parameter, weighted by their posterior probabilities. However, we have shown that, for a certain number of interest quantities, PI is in fact equivalent to Bayes estimation for a certain cost function, depending on the expression of the interest quantity.

This raises some concern. Indeed, cost functions should be chosen based on an evaluation of the risks associated with potential estimation errors, rather than the particular expression of the quantity under study. In other terms, the major drawback of PI is that it robs the user of the freedom of specifying exactly what estimation problem he wishes to solve, by implicitly choosing for him the cost function to be minimized. Furthermore, we have seen that interpreting the results of PI is delicate, since the predictive density has no phenomenological interpretation.

## References

[1] Alberto Pasanisi, Eric Parent, Nicolas Bousquet. Some useful features of the Bayesian setting while dealing with uncertainties in industrial practice. In *Proceedings of ESREL 2009*, 2009.

[2] N. Bousquet. Calibrating Weibull priors in Bayesian reliability and risk assessment. *submitted*, 2005.

[3] Charles Stein. Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution. In *Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 197 – 206, 1956.

[4] Ronald Christensen and Michael D. Huffman. Bayesian Point Estimation Using the Predictive Distribution. *The American Statistician*, 39(4):319 – 321, 1985.

[5] S. Tarantola E. de Rocquigny, N. Devictor. *Uncertainty in Industrial Practice*. Wiley, 2008.

[6] Jacques Bernier Eric Parent. *Le raisonnement Bayésien* (in French). Springer, 2007.

[7] Lurdes Inoue Giovanni Parmigiani. *Decision Theory*. Wiley, 2009.

[8] Marc C. Kennedy and Anthony O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B, Methodological*, 63:425–464, 2001.

[9] Roman Krzysztofowicz. Bayesian forecasting via deterministic model. *Risk Analysis*, 19(4):739–749, 1999.

[10] Merlin Keller, Alberto Pasanisi, Eric Parent, Nicolas Bousquet. Bayesian and frequentist estimation in the Weibull Model. In *Proceedings of the Valencia 9 meetings and ISBA 2010*, 2010.

[11] Michael Goldstein. External Bayesian computer analysis. In *Bayesian Statistics*, volume 9, pages 1 – 17, 2010.

[12] Elisabeth Paté-cornell. Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering and System Safety*, 54:95 – 111, 1996.

[13] D. Sun and P.L. Speckman. A note on existence of posterior moments. *Canadian Journal of Statistics*, 3, 2005.

[14] A. W. van der Vaart. *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press, June 2000.