Risk analysis, uncertainty and robust decision-making: an attempted introduction

Nicolas Bousquet

EDF R&D

November 21, 2022

Works and discussions shared with many people : B. Iooss, J.-M. Loubes, S. Da Veiga, F. Gamboa, F. Ruggeri, A. Raftery, E. Parent, L.-P. Rivest, V. Chabridon, M. Il Idrissi, S. Ancelet, M. Blazère, etc.

Key words for this workshop on *statistical approaches to safety and decommissioning* (of industrial facilities) (especially nuclear ones)

- Uncertainties
- Risk and reliability
- Contamination, radionuclide quantification, radiations
- Bayesian approaches
- Metrology (measurement process, GUM, etc.)
- Geostatistics and metamodeling/surrogates (Gaussian processes, neural networks) under form constraints (e.g., monotonicity)
- Sensitivity analysis
- Lunches & coffees & Apéro (and gala dinner of course)

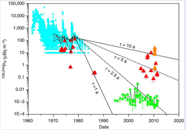
A personal view of thematics [1]

What characterizes (from my point of view) important statistical problems related to safety and decommissioning

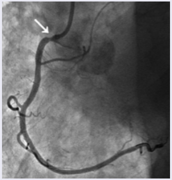
We face many uncertainties :

1. **On data information**, related to uncertainties from measurement and reconstruction processes ⇒ **How selecting good quality data ?** (small samples analyses)

- e.g., cracking, radionuclides, radiation-induced diseases ...
- talks by M. Désenfant, C. Norman *et al.*, S. Ancelet, poster by J. Baccou,...



(Atmosph. radionucl.) [2]



(RI heart disease) [25]

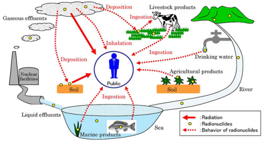
A personal view of thematics [1]

What characterizes (from my point of view) important statistical problems related to safety and decommissioning

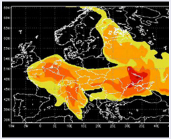
We face many uncertainties :

1. **On data information**, related to uncertainties from measurement and reconstruction processes ⇒ **How selecting good quality data ?** (small samples analyses)
2. **On the predictive behavior of components and systems**, through models or surrogates ⇒ **Calibrating / inverting input parameters, sensitivity analysis, controlling (meta)model errors or avoid using models**

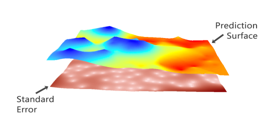
- talks by A. Clément, T. Rovary, S. Ancelet, posters by J. Baccou, C. de Fouquet, M. Wieskotten,...
- Prohibitive comput. time ⇒ *learning from simulations* (metamodeling / surrogates)
- talks by Y. Desnoyers, F. Bachoc, M. Ducoffe, posters by C. Gauchy, R. Perillat, ...

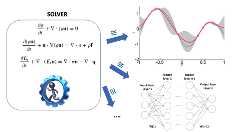


JAEA + gisgeography.com



IRSN



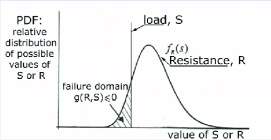


What characterizes (from my point of view) important statistical problems related to safety and decommissioning

We face many uncertainties :

1. **On data information**, related to uncertainties from measurement and reconstruction processes ⇒ **How selecting good quality data ?** (small samples analyses)
2. **On the predictive behavior of components and systems**, through models or surrogates ⇒ **Calibrating / inverting input parameters, sensitivity analysis, controlling (meta)model errors or avoid using models**
3. **On studying extreme situations** ⇒ **Computing risk indicators**, accounting for (meta)model errors

- (e.g., probabilities, quantiles) with strong guarantees (conservative)
- Guide ASN n°28 *Qualification of scientific computing tools for nuclear safety demonstration*
- talks by E. Borgonovo, F. Bachoc, A. Marrel, poster by V. Chabridon, ...



[43]

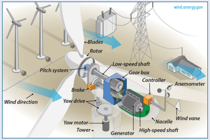
TWP

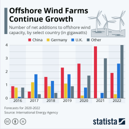
What characterizes (from my point of view) important statistical problems related to safety and decommissioning

We face many uncertainties :

1. **On data information**, related to uncertainties from measurement and reconstruction processes ⇒ **How selecting good quality data ?** (small samples analyses)
2. **On the predictive behavior of components and systems**, through models or surrogates ⇒ **Calibrating / inverting input parameters, sensitivity analysis,controlling (meta)model errors or avoid using models**
3. **On studying extreme situations** ⇒ **Computing risk indicators**, accounting for (meta)model errors
4. **On the capacity of alternative energies to efficiently complement those produced by nuclear power plants** (decommissioned one day), and preparing their future decommissioning ⇒ **Data assimilation, forward simulation, optimization, etc.**

- e.g., reliability of wind power generation
- talk by M. Fouladirad





Everyone probably knows the consensual aleatoric part of uncertainties (related to intrinsic variability of magnitudes)

Key role of epistemic uncertainty (IRSN also uses the terms "imprecision")

- due to imprecise knowledge or lack of knowledge
- affects choices tainted with some subjectivity (e.g., working hypothesis)

Could be interpreted as resulting from an **accumulation of potentially reducible errors (e.g., modeling errors)** [29], which can significantly affect critical decisions

---

Remark. *Difficulty to discriminate between uncertainties related to strong technical limits (e.g., measurement / computing limits)*

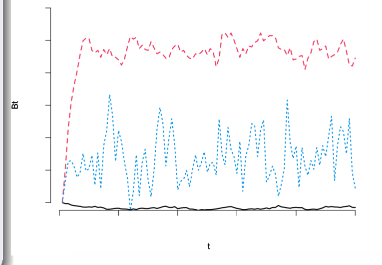
Illustrating the influence of model errors on a decision through a simple (but realistic) example

Model of resource evolution (logistic / Gray-Verhulst)

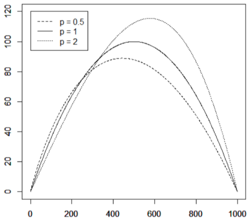
$$\frac{dB_t}{dt} \simeq B_t - B_{t-1} = g_\theta(B_t) - \phi_t B_t$$

- B_t = resource at time t
- g_θ = renewal function
  with $\theta \supseteq \begin{cases} \text{growth rate } r \\ \text{saturating resource } K \end{cases}$
  (ex : $g_\theta(B_t) = rB_t(1 - (B_t/K)^p)$)
- ϕ_t = extraction rate

Évolution with constant ϕ_t

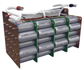


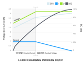
Saturating renewal function



Industry

- B_t = Lifetime of lithium-ion batteries, ϕ_t = wear rate [61]
- B_t = load forecasting of electrical systems, ϕ_t = wear rate [53]
- B_t = C02 emissions, ϕ_t = absorption rate [56]



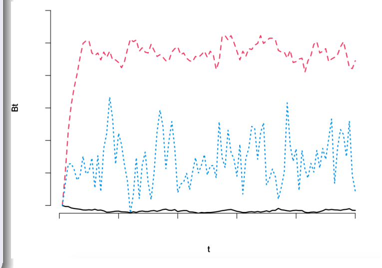


Model of resource evolution (logistic / Gray-Verhulst)

$$\frac{dB_t}{dt} \simeq B_t - B_{t-1} = g_\theta(B_t) - \phi_t B_t$$

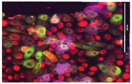
- B_t = resource at time t
- g_θ = renewal function
  with $\theta \supseteq \begin{cases} \text{growth rate } r \\ \text{saturating resource } K \end{cases}$
  (ex : $g_\theta(B_t) = rB_t(1-(B_t/K)^p)$)
- ϕ_t = extraction rate

Évolution with constant ϕ_t



Health

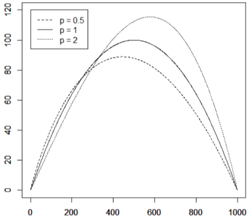
- B_t = nb of cancer cells, ϕ_t = chemother. injection [60]



Quantitative ecology

- B_t = biomass, ϕ_t = anthropic impact [27]
- e.g., effects of ionising radiations on species [46, 58]

Saturating renewal function



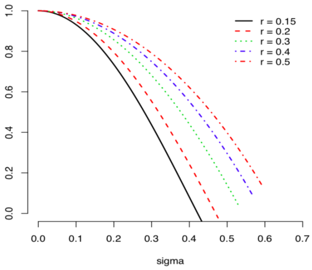
Which maximal extraction for resource sustainable equilibrium ?

Historical approach $B_{t+1} = B_t$

(Hyp : $\phi_t = \phi$, $p = 1$)

$$\phi_{opt} = \frac{r}{2} \Rightarrow \text{optimal extraction} = \frac{rK}{4}$$

Ratio of extractions (new / hist.)



Stochastic approach (B. et al. [15]) $B_{t+1} \sim B_t$

Let us introduce a **model error** ε_t

$$B_t = \{B_{t-1} + g_\theta(B_t) - \phi_t B_t\}\, \varepsilon_t$$

with $\mathbb{E}[\varepsilon_t] = 1$ and $\mathbb{V}[\varepsilon_t] = \sigma^2$. If $\sigma^2 < \sigma_0^2$ (non-extinction condition) then

$$\phi_{\text{opt}} = \frac{r}{2} - \frac{2(2-r)}{(4-r)^2}\sigma^2 + o(\sigma^3)$$

and the optimal extraction is

$$\frac{rK}{4}\left(1 - \frac{\sigma^2}{r(1-r/4)} + \frac{4\sigma^4}{r^2(4-r)^4}(1 + o(\sigma^4))\right)$$

Under the stationarity assumption, we can simplify $B_{t+1} \sim B_t$ in (for instance)

$$E[B_{t+1}|B_t] = B_t \quad \text{(martingality)}$$

and the decision will be something like "the optimal extraction is

$$B^* = \arg\min_{x \geq 0} \phi_{\text{opt}}(\theta) \int \ell(x, B_t) dP(B_t)\text{"}$$

where

- θ is the set of parameter (that need to be estimated)
- ℓ is a choice of cost function

With ℓ chosen as quadratic

$$B^* = \frac{rK}{4}\left(1 - \frac{\sigma^2}{r(1 - r/4)} + \frac{4\sigma^4}{r^2(4 - r)^4}(1 + o(\sigma^4))\right)$$

Obtaining / selecting good quality data

To apprehend a critical feature of a system :

- Defining good measurements
- Selecting representative *(prototypes)* subsets of experimental designs [14, 26, 23]

Selecting a good measurement is a decision that might be formalized as follows

Let $Y = y_i$ be an (indirect) measurement of a quantity $X = x_i$, understood as

$$y_i = g_\Sigma(x_i, \varepsilon_i)$$

where

- g_Σ is an operator modeling a measurement process Σ
- $\varepsilon \sim P(\varepsilon)$ is a random "noise" summarizing the influence of external factors

For a same (hidden) source $X = x_i$, several values of y_i due to ε_i

Example : lung cancer screening by thoracic scanner

Source : [41]

- X = tumor features
- Y = table of pixels
- ε = patient position + setting chosen by the operator

From repeated observations $\boldsymbol{Y}(x)$, assess the quality of a measurement Y by estimating (for instance) the conditional variance

$$\begin{aligned}\mathrm{Var}[Y|X=x] &= \int \ell(g_\Sigma(x,\varepsilon))dP(\varepsilon) \quad \text{with } \ell(u(\varepsilon)) = \mathrm{E}_\varepsilon[u^2(\varepsilon)] - \mathrm{E}^2_\varepsilon[u(\varepsilon)] \\ &= \textit{indicator of measurement uncertainty in } X = x\end{aligned}$$

Assuming $X \sim P_X$, a global indicator of quality for Σ could legitimately be

$$Q_\Sigma = \mathrm{E}_X\left[\mathrm{Var}[Y|X]\right]$$

(note that is can be estimated only with a sample $\boldsymbol{Y} = \{\boldsymbol{Y}_{ij}(x_i)\}_{i,j}$ without knowing the real x_i)

Now, having two competing measurement processes Σ_1 and Σ_2, may we compare the Q_{Σ_i} to check if "Σ_1 is better than Σ_2" ?

From repeated observations $\boldsymbol{Y}(x)$, assess the quality of a measurement Y by estimating (for instance) the conditional variance

$$\begin{aligned}\mathrm{Var}[Y|X=x] &= \int \ell(g_\Sigma(x,\varepsilon))dP(\varepsilon) \quad \text{with } \ell(u(\varepsilon)) = \mathrm{E}_\varepsilon[u^2(\varepsilon)] - \mathrm{E}^2_\varepsilon[u(\varepsilon)] \\ &= \textit{indicator of measurement uncertainty in } X=x\end{aligned}$$

Assuming $X \sim P_X$, a global indicator of quality for Σ could legitimately be

$$Q_\Sigma = \mathrm{E}_X\left[\mathrm{Var}[Y|X]\right]$$

(note that is can be estimated only with a sample $\boldsymbol{Y} = \{\boldsymbol{Y}_{ij}(x_i)\}_{i,j}$ without knowing the real x_i)

Now, having two competing measurement processes Σ_1 and Σ_2, may we compare the Q_{Σ_i} to check if "Σ_1 is better than Σ_2" ?

What we want from using each Σ_i is to reconstruct X, or rather P_X (in a concern of generality), using $\boldsymbol{Y}_{\Sigma_i}$ (*stochastic inversion*)

Classical approach.

1. Assume $X \sim P_X(.|\theta)$ parameterized by θ (e.g., a multivariate Gaussian)
2. Estimate θ from $\boldsymbol{Y}_{\Sigma_i}$ (e.g., using missing data, EM-type algorithms [20, 8])

$$\theta \quad \Rightarrow \quad \hat{\theta}(\boldsymbol{Y}_{\Sigma_i})$$

Then

$$Q_{\Sigma_i} \quad = \quad Q_{\Sigma}\left(\hat{\theta}(\boldsymbol{Y}_{\Sigma_i})\right)$$

but we cannot be sure to have a total order between the Q_{Σ_i} [52]
$\Leftrightarrow$ we cannot correctly compare Σ_1 and Σ_2

What we want from using each Σ_i is to reconstruct X, or rather P_X (in a concern of generality), using $\boldsymbol{Y}_{\Sigma_i}$ (*stochastic inversion*)

Bayesian approach.

1. Note that θ is a summary of the features of $X \sim P_X$, endowed with epistemic uncertainty
2. Model this uncertainty by defining technically θ as a random variable with **prior measure**
$$\theta \sim \pi(\theta)$$
3. Estimate the **posterior** $\pi(\theta|\boldsymbol{Y}_{\Sigma_i})$ (e.g., using Monte Carlo-type algorithms [31, 32])
$$\pi(\theta) \quad \Rightarrow \quad \pi(\theta|\boldsymbol{Y}_{\Sigma_i}) \qquad \textit{(Bayesian updating)}$$

Then

$$Q_{\Sigma_i} \quad = \quad \mathrm{E}_\theta \left[\mathrm{E}_X \left[\mathrm{Var}[Y|X] | \theta \right] | \boldsymbol{Y}_{\Sigma_i}) \right]$$

It is a **Bayes estimator** then we are sure to get a total order between the Q_{Σ_i}
$\Leftrightarrow$ we can compare Σ_1 and Σ_2

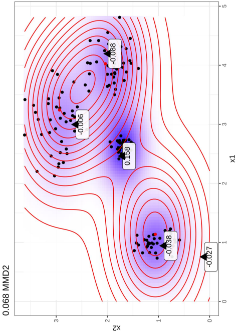
More generally, designing / selecting good (informative) experiments

- so-called *support points* or *representative points*
- *prototypes* from a database (in machine learning)

Framework

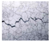
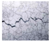
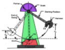
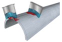
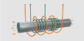
- $\mathbf{x_n} = (x_1, \ldots, x_n) \in \mathcal{X}^n$ = some design points
- Corresponding output $y_i = g_\theta(x_i)$ where g_θ is a model
- How selecting
  $\begin{cases} \text{new informative design points} \\ \text{the most informative design points within } \mathbf{x_n} \end{cases}$

(informative on θ or some function of θ



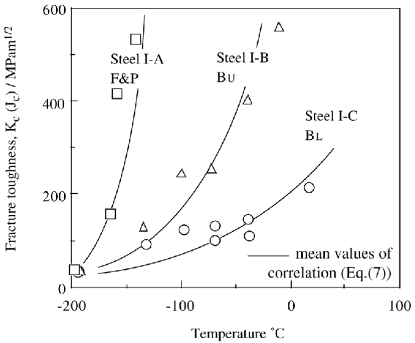
[45]

Close (if not similar) approaches for various situations

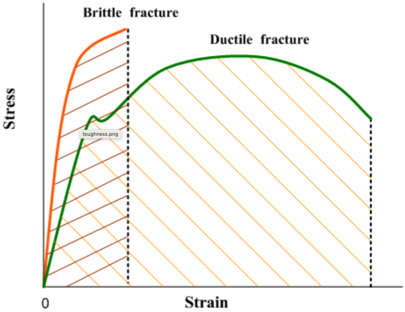
	Destructive experiments	**Nondestructive experiments**	**In silico experiments**
Examples of tested properties	• Fracture toughness • Stress corrosion cracking  	• Stable undercoating defects • Stress corrosion cracking  	• Robustness of an artificial intelligence (AI) tool • Fidelity of a physically-based digital twin
Some experimental techniques	• Charpy-type experiments • ALT chemical testing, etc.  	• Ultrasonic inspections • Eddy current testing  	• Selecting among collected observations • Designing numerical experiments using optimization
Typical cost constraints	• Very limited number of specimen • Selecting stress levels and specimen features • Strong environemental conditions (e.g. T°) • Noise removal requirement • Availability of experts		• Prohibitive training time • Prohibitive inference / simulation time related to model complexity

Fracture toughness of steel (FTOS) characterizes the capacity of the material to resist to cracking through plastic deformation when a load is applied (e.g., a transient cooling such as water injection)

It is part of the most influential material attributes in structural safety studies [55].

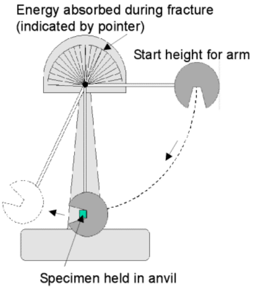


Source : [44]



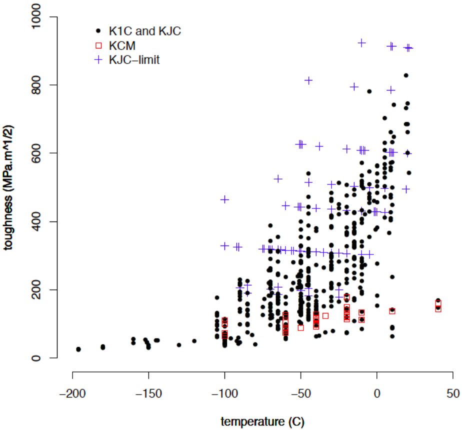
Source : https://www.substech.com

Charpy impact tests [5] ⇒ indirect toughness values (megapascal square root meter) with different qualities



Source : `https://theconstructor.org`

European FTOS database (ferritic steels) from Oak Ridge National Laboratory (ASTM E399-90 [4])



Inspired by [21], define a Bayesian experimental design

1. Consider a well recognized theoretical statistical model (e.g., from *weakest link theory [37]*) linking a FTOS measure y_i^j at a given temperature T_j

$$P(Y_i^j < y | T_j, \theta) \quad = \quad 1 - \exp\left(- \left\{ \frac{y_i^j - \alpha)}{\mu(T_j)} \right\}^{\beta} \right)$$ *(simple Master Curve [59])*

   with $\mu(T_j) = \lambda_1 + \lambda_2 \exp(\lambda_3 T_j)$ and $\theta = (\alpha, \{\lambda_i\}_i, \beta)$

2. Elicit a prior distribution $\Pi(\theta)$
3. Formalize a design of experiments for fixed n standard Charpy specimen [25 mm]

$$\varepsilon = \left\{ J, \left\{ \begin{array}{cccc} T_1 & T_2 & \ldots & T_J \\ \eta_1 & \eta_2 & \ldots & \eta_J \end{array} \right\} \right\}$$

   with $\eta_j = \frac{n_j}{n} \in [0,1]$ for all $j = 1, \ldots, J$ and $\sum_{j=1}^{J} \eta_j = 1$

$U_1(\varepsilon)$ = expected utility function quantifying the expected gain in knowledge about θ provided by data collected under the experimental design ε

$U_2(\varepsilon)$ = expected utility function quantifying the opposite of the expected experimental cost under ε

Generic (compound) weighted (dimensionless) utility [3] (similar idea in [36])

$$U(\varepsilon) = \omega \times \Delta U_1(\varepsilon) + (1-\omega) \times \Delta U_2(\varepsilon)$$

where

$$\Delta U_k(\varepsilon) = \frac{U_k(\varepsilon) - U_k(\varepsilon_0)}{|U_k(\varepsilon_0)|} \qquad \text{for} \qquad k = 1, 2$$

- $\Delta U_k(\varepsilon)$ = relative change in expected utility
- ε_0 = fixed baseline experimental design for which the total expected utility $U(\varepsilon_0)$ is set to zero
- for instance (using typical temp values within the brittle-ductile transition zone)

$$\varepsilon_0 = \left\{ 4, \left\{ \begin{array}{cccc} -150 & -100 & -50 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{array} \right\} \right\}$$

Quantifying the opposite of the number of days of work required for collecting data at a design point :

$$U_2(\varepsilon) = -\sum_{j=1}^{J} n_j \times \left(2 - \mathbf{1}_{\{T^- < T_j < T^+\}}\right)$$

where $T^- = -130^\circ C$, $T^+ = -60^\circ C$

(one day of work to make a test when $T \in [T^-, T^+]$ but two days to homogenize the room temperature in more extreme conditions)

Quantifying the expected gain in knowledge provided by data collected under an experimental design ε about θ

- **[Ex.1]** Posterior-prior KL divergence ⇒ all dimensions of θ

$$U_1^1(\varepsilon) = \iint \log \frac{\pi(\theta|\mathbf{y},\varepsilon)}{\pi(\theta)} \pi(\theta|\mathbf{y},\varepsilon) d\theta d\mathbf{y}$$

$$\overset{\text{asympt.}}{\simeq} cte + \frac{1}{2}\mathbb{E}_{\mathbf{Y}}\left[\log\left(\det(\Sigma(\widehat{\theta},\varepsilon)^{-1})\right)\right] \quad (D\text{−optimal design})$$

- **[Ex.2]** Opposite of the quadratic loss function ⇒ selected linear combination of dimensions of θ

$$U_1^2(\varepsilon) = -\iint (\theta-\hat{\theta})^T A(\theta-\hat{\theta}) f(\mathbf{y},\theta|\varepsilon) d\theta d\mathbf{y}$$

$$\overset{\text{asympt.}}{\simeq} -\mathbb{E}_{\mathbf{Y}}\left[\text{tr}(A\Sigma(\widehat{\theta},\varepsilon))\right] \quad (A\text{−optimal design})$$

(with $\widehat{\theta}$ = posterior mode, A symmetric nonnegative definite matrix, $I(.,.)$ = Fisher matrix and R = prior precision matrix)

1. Gaussian prior computed as an approximation of a posterior from European FTOS data (flat baseline prior)
2. Use simulated annealing [3] or the approximate coordinate exchange algorithm [48]

ω	u_1	J^*	η^*	T^*	$\widetilde{U}(\epsilon^*)$
1	D	J=3	(0.55,0.27,0.18)	(-213.84,-97.52,17.80)	0.046
	A_1	J=2	(0.31,0.69)	(-213.80,9.21)	0.156
	A_2	J=2	(0.58,0.42)	(-213.70,12.48)	0.102
0.9	A_1	J=2	(0.31,0.69)	(-213,91,7.62)	0.126
	A_2	J=3	(0.54,0.10,0.36)	(-213.96,-60.21,17.71)	0.079
0.5	A_1	J=3	(0.49,0.42,0.09)	(-129.51,-60.10,17.92)	0.164
	A_2	J=2	(0.92,0.08)	(-129.97,-60.37)	0.200

Howver... avoiding pseudo-Bayesian designs if possible

- The addition of noises, model errors and measurements limits will decrease the quantity of information yielded by planned experiments
- Asymptotic assumptions and prior (Gaussian) assumptions behind A- and D-optimal design criteria can be strongly unrealistic and lead to degenerate situations [38, 36]
- Modern computational techniques become capable of tackling the problem of computing repeatedly posteriors to solve the optimization problem of the design ε
  - multi-stages mixing stochastic gradient optimisation and automatic differentiation [49]

$\Rightarrow$ Good prior is required !

We should focus more on prior modeling

- Priors ("best guesses") can significantly help to produce useful designs (e.g., [12] for clinical studies)
- All the more when the planned design is small-sized (since costly)

Producing defensible priors take part in a more general, growing approach of questioning the formalization of prior choices

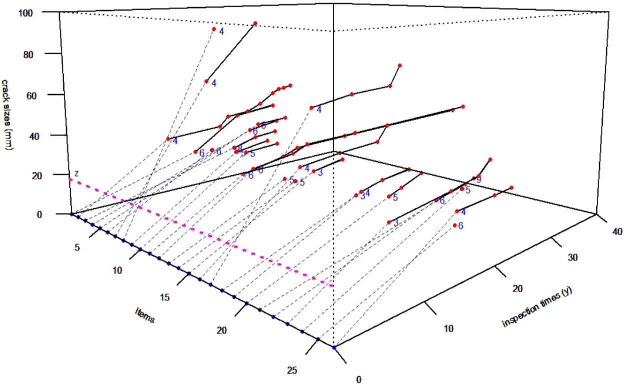
- A. Gelman and J. Sprenger on the objectivity and reproducibility of Bayesian assessments : [Holes in Bayesian Statistics] [34, 54, 35]
- Contemporary concerns for the auditability of deep learning [30] and artificial intelligence [62]

What could be a good prior ? Use an illustration !

A crack size $Z_{k,t}$ on a component k is monotonically increasing with time t

The increments (assumed independent) $X_{k,i} = Z_{k,t_i} - Z_{k,t_{i-1}}$ are assumed to be gamma distributed

$$f_{\alpha(t-s),\beta}(x) = \frac{1}{\Gamma(\alpha(t-s))} \cdot \frac{x^{\alpha(t-s)-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha(t-s)}} \mathbb{1}_{\{x \geq 0\}}$$



Consider Jeffreys' prior $\pi^J(\alpha,\beta) \propto \frac{1}{\beta}\sqrt{\alpha\Psi_1(\alpha)-1}$

$\pi(\theta)$ beneath is the first-order (Taylor) approximation of the posterior of an imaginary sample of crack increments $\tilde{\mathrm{x}}_{\mathrm{m}} = (\tilde{x}_1,\ldots,\tilde{x}_m)$ observed at times $\tilde{\mathrm{t}}_{\mathrm{m}} = (\tilde{t}_1,\ldots,\tilde{t}_m)$:

$$
\begin{aligned}
\beta|\alpha &\sim \mathcal{IG}\left(\alpha m\tilde{t}_{e,1}, m\tilde{x}_e\right)\\
\alpha &\sim \mathcal{G}\left(m/2, m\tilde{t}_{e,2}\right)
\end{aligned}
$$

with the meanings

$$
\begin{aligned}
\tilde{t}_{e,1} &= \frac{1}{m}\sum_{i=1}^{m}\tilde{t}_i && \textit{(mean observation time)}\\
\tilde{x}_e &= \frac{1}{m}\sum_{i=1}^{m}\tilde{x}_i && \textit{(mean increase)}\\
\tilde{t}_{e,2} &= \frac{1}{m}\sum_{i=1}^{m}\tilde{t}_i \log\frac{\sum_{j=1}^{m}\tilde{x}_j/\tilde{x}_i}{\sum_{j=1}^{m}\tilde{t}_j/\tilde{t}_i} && \textit{(tuning hyperparameter)}
\end{aligned}
$$

Other similar ideas can come from the rich literature on Edgeworth expansions for posterior densities [40]

- **Static or incremental Space filling designs** based on discrepancies [50]
- **Sequential Bayesian designs** produced by Stepwise Uncertainty Reduction (SUR) strategies [10]
- **Quantization techniques** like Maximum Mean Discrepancy minimization (ie., using kernel herding, grid search or Sequential Bayesian Quadrature [51])
- **Selection of subsamples (prototypes) in a database**
  - A concern shared with machine learning tasks confronted with huge cardinality and dimension (e.g. [26, 14])
  - Related current works (EDF-IRSN-CEA-Université de Toulouse) linked to some improvements of the **SAPIUM** project [7] (2017-2019) : Establishing the relevance of an experimental database

SAPIUM: a generic framework for a practical and transparent quantification of thermal hydraulic code model input uncertainty

Jean Baccou, Jinzhao Zhang, Philippe Fillion, Guillaume Damblin, Alessandro Petruzzi, Rafael Mendizabal, Francesc Reventos, Tomasz Skorek, Mathieu Couplet, Bertrand Iooss, et al.

Conservative and robust risk assessments

Sometimes, the small size of measurement samples makes it difficult to use a physical or statistical model to describe them

Ex : Radiological characterization of contaminated elements (e.g., walls, grounds, objects)

- Too few measurements $x_1, \ldots,$ to make classical hypotheses (e.g., Gaussian distribution)
- How determining **risk prediction bounds** on the level of contamination ?

$$P(X > x_s) \leq \alpha$$

The authors [13] propose to strongly limit model assumption $X \sim f$ and use nonasymptotic (concentration) inequalities tools

- **Camp-Meidell inequality** (if f unimodal)

$$P(X \geq \mu + t) \quad \leq \quad \left(1 + \frac{9}{4}\frac{t^4}{\sigma^2}\right)^{-1}$$

Sometimes, the small size of measurement samples makes it difficult to use a physical or statistical model to describe them

Ex : Radiological characterization of contaminated elements (e.g., walls, grounds, objects)

- Too few measurements $x_1, \ldots,$ to make classical hypotheses (e.g., Gaussian distribution)
- How determining **risk prediction bounds** on the level of contamination ?

$$P(X > x_s) \leq \alpha$$

The authors [13] propose to strongly limit model assumption $X \sim f$ and use nonasymptotic (concentration) inequalities tools

- in addition : **Wilks' formula** (no assumption on f), requiring a minimal size n such that

$$P(P(X \geq \max X_i \geq \gamma) \geq \beta$$

Took from [13]

$$P(P(X \geq \max X_i \geq \gamma) \geq \beta$$

γ	0.9	0.9	0.9	0.95	0.95	0.95	0.95	0.99	0.99
β	0.5	0.9	0.95	0.4	0.5	0.78	0.95	0.95	0.99
n	7	22	29	10	14	30	59	299	459

Applications to non-iid data, through martingale-based inequalities (e.g., Azuma-Hoeffding)

Our decisional variable is

$$Y = g(X)$$

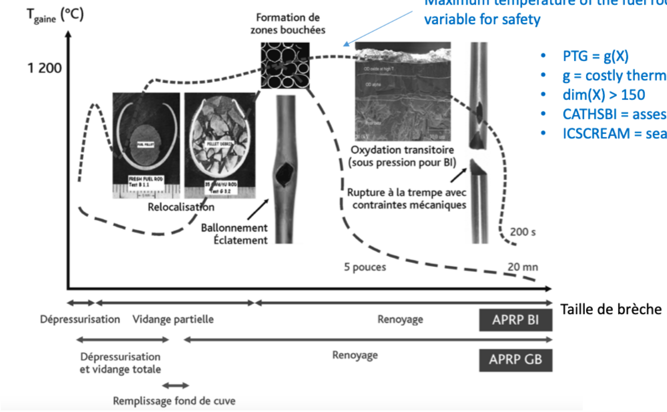
where $X \sim f(x)$ is a set of incertain input parameters

- Safety analysis based on specific designs of experiments
- Complementary structural reliability analysis based on indicators as

$$p = P(Y \geq y_0) = \int \mathbb{1}_{\{g(x) \geq y_0\}} f(x) dx \quad \text{(failure probability)},$$

or quantiles associated to extreme levels of risk, etc.

The risk and reliability (R&R) indicators can theoretically be computed by Monte Carlo sampling



Source : IRSN

Maximum temperature of the fuel rod cladding (PTG)= dimensioning variable for safety

- PTG = g(X)
- g = costly thermo-hydraulical computer code
- dim(X) > 150
- CATHSBI = assessing uncertainties on PTG
- ICSCREAM = searching for penalizing X

The risk and reliability (R&R) indicators can theoretically be computed by Monte Carlo sampling

But computationally unfeasible in practice

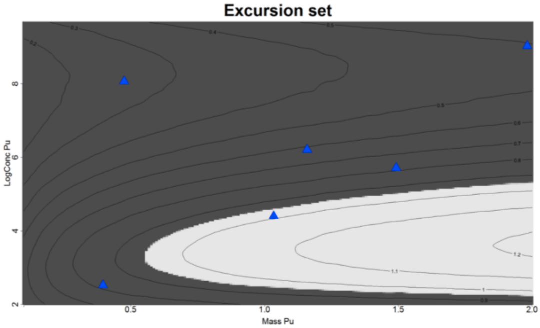
- $\Rightarrow$ Monte Carlo variance reduction techniques (e.g., splitting, line sampling)
- use of surrogates if relevant (e.g. Gaussian process-based kriging, physics-informed neural networks...)

Issues

- What is a good metamodel ?
- How should I deal with its error for computing my indicators ?

Excursion set of g above y_0

$$\Gamma^* \quad = \quad \{x \in \mathcal{X} : \; g(x) \geq y_0\}$$



Excursion set (light gray) of a nuclear criticality safety coefficient depending on two design parameters. Blue triangles : initial experiments [22]

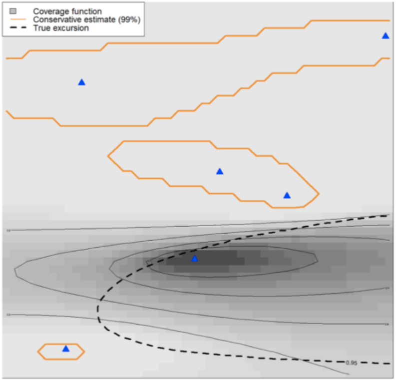
⇒ Bounding such sets can provide conservative assessments of reliability and risk indicators

Excursion set of g above y_0

$$\Gamma^* = \{x \in \mathcal{X} : g(x) \geq y_0\}$$

- Stepwise Uncertainty Reduction (SUR) strategies + Gaussian process [10]
- Authors [6] recently provide strategies to ensure a (very) good conservative estimate of an excursion set
- Could maybe be adapted to neural networks using **conformalized prediction**
- **Remaining issues** : high dimension + results obtained conditionally to small (meta)model error

Conservative estimate at 99% (6 observations)



What would we like to have ideally ?

Computing (R&R) indicators, using metamodels, with fine conservatism

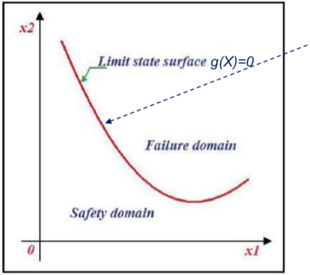
Consider the target probability of *feared situations*

$$p = P(X \in A_{y_0}) = \int_{A_{y_0}} dP(x) = \int_{\mathcal{X}} \mathbb{1}_{\{g(x)\geq y_0\}} dP(x).$$

estimated as $\hat{p}_{d_n} = \int_{\mathcal{X}} \mathbb{1}_{\{\hat{g}(x|d_n)\geq y_0\}} dP(x)$ where $\hat{g}(x|d_n)$ is a metamodel of g assessed from a (training) design d_n

We could rather propose another estimator based on a metamodel $\hat{\Gamma}_n$ of the limit state (classification) surface

$$\Gamma = \{x \in, \; g(x) \in \partial A_{y_0}\}$$



Computing (R&R) indicators, using metamodels, with fine conservatism

Consider the target probability of *feared situations*

$$p \quad = \quad P(X \in A_{y_0}) \;=\; \int_{A_{y_0}} dP(x) \;=\; \int_{\mathcal{X}} \mathbb{1}_{\{g(x) \geq y_0\}} dP(x).$$

estimated as $\hat{p}_{\mathrm{d_n}} = \int_{\mathcal{X}} \mathbb{1}_{\{\hat{g}(x|\mathrm{d_n}) \geq y_0\}} dP(x)$ where $\hat{g}(x|\mathrm{d_n})$ is a metamodel of g assessed from a (training) design $\mathrm{d_n}$

For any function $h(g)$ (as p), let $\hat{h}_m(g)$ be an estimator computed from m simulations, and denote $\hat{h}_m(\hat{g}_n)$ the metamodel-based approximation of $\hat{h}_m(g)$

Weak (minimal) guarantee ($\sim$ universal approx. theorem)

$$\hat{h}_n\left(\hat{f}_m\right) \quad \xrightarrow[n,m\to\infty]{a.s.} \quad h(f) \qquad \text{(general tool : random set theory)}$$

Computing (R&R) indicators, using metamodels, with fine conservatism

Consider the target probability of *feared situations*

$$p = P(X \in A_{y_0}) = \int_{A_{y_0}} dP(x) = \int_{\mathcal{X}} \mathbb{1}_{\{g(x) \geq y_0\}} dP(x).$$

estimated as $\hat{p}_{d_n} = \int_{\mathcal{X}} \mathbb{1}_{\{\hat{g}(x|d_n) \geq y_0\}} dP(x)$ where $\hat{g}(x|d_n)$ is a metamodel of g assessed from a (training) design d_n

Instead, reuse the concept from Ducoffe *et al.* [28] (then Gauffriau *et al.* [33]) of the probability of a *safe* surrogate by

$$q = P\left(g(X) \geq \hat{g}(X|d_n)\right)$$

$\hat{g}(X|d_n)$ **is safe with probability** $q = 1 \Leftrightarrow \boxed{p \leq \hat{p}_{d_n}}$

Upper bound (lemma)

- Assume there exists $\alpha \in [0,1[$ such that $P(E_{y_0}) \leq \alpha$
- Denote

$$\beta = P(g(X) \geq \hat{g}(X) | \hat{g}(X) \geq y_0).$$

Then

$$p \leq \beta \hat{p}_{d_n} + \frac{(1-\beta\hat{p}_{d_n})}{(2-\hat{p}_{d_n}-q)} \left[\hat{p}_{d_n}(1-q) + \frac{q\alpha(1-\hat{p}_{d_n})}{q-\hat{p}_{d_n}} \right].$$

- **Tools** : concentration inequalities, among others (for iid and non iid samplings)
- Ongoing work at EDF

Exploration of monotonicity of models and limit state surfaces ⇒ constrained surrogates [16, 24, 19, 42, 57], etc.

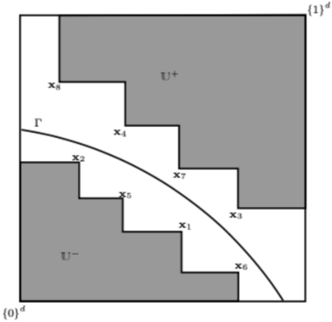
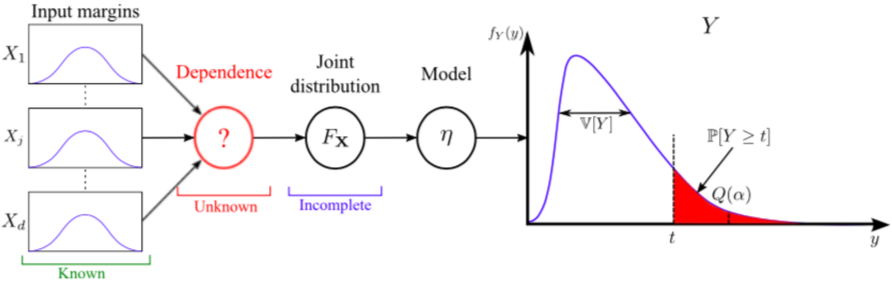


Figure 8 – Illustration for $d = 2$ of the designs $\mathbb{U}^-(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_6)$ and $\mathbb{U}^+(\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_7, \mathbf{x}_8)$, where the $\mathbf{x}_i$ are sampled over $\mathbb{U} = [0,1]^2$. The probability p corresponds to the volume under the limit state surface Γ.

- Monotonicity can be possibly considered as a conservative assessment or a way of respecting physics
- But obtaining strong guarantees is uneasy (e.g. [9])
- Could be interesting to define the closest monotonic surrogate (and controlling its error to be unsafe) (e.g., linear variational surrogates)
- Other ideas : use quasi-convexity properties [47]



We are looking for a parameterized dependence structure C_θ between the X_i

In [11], algorithms to select penalizing dependence structures (copulas) are provided

- Penalizing $\Leftrightarrow$ minimizing the output quantile value of order α
- Requires parsimony hypotheses

Today, the validity of metamodels is somewhat entangled with their calibration : both are based on *training sets of simulation data*

⇒ Clarifying the generalization properties

A metamodeling constraint should be : conclusions produced with its help should be similar to the ones provided using the "real" (most accurate) model

It means for instance that sensitivity analyses should produce the same results

⇒ the metamodel error should have the lowest SA indice($e.g.$, $Sobol$, $Shapley$, $etc.$)

Ensuring this is not an easy task [17]

$$Y = g(X) \quad \text{with } X \sim P$$

How risk indicators computed over Y react to model misspecifications on P ?

Idea : minimizing a distance under constraints

$$Q^* \quad = \quad \arg\min_{Q} \mathcal{D}(P, Q)$$

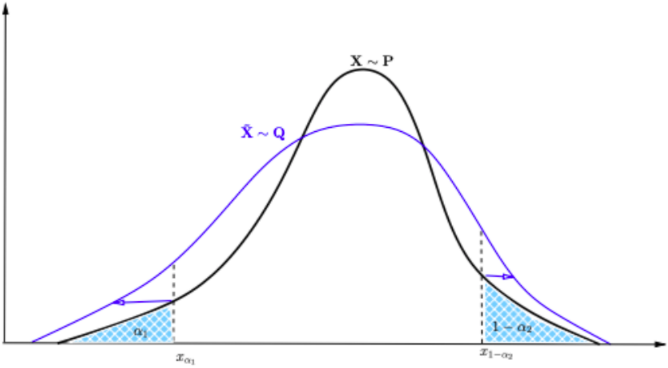
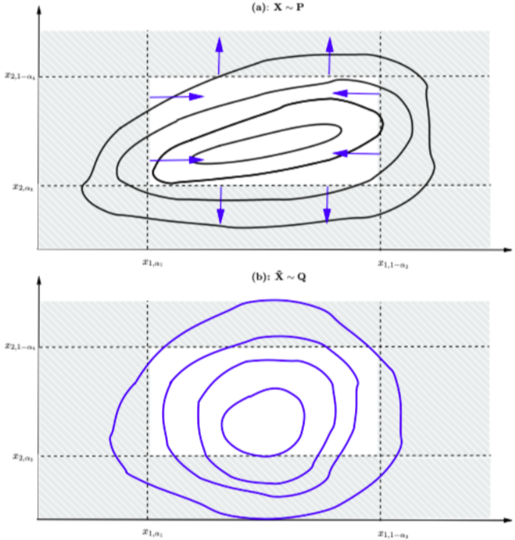


FIG 1. One-dimensional illustration of covariate shift in the UQ framework. A postulated density distribution for $X \sim P$ (solid black line) is dilated by modifying the tail order of low and high quantiles defining an application domain (blue arrows), resulting in a new distribution Q.

Ex : quantile constraints



The p-Wasserstein distance between Q and P on respective supports $\mathcal{X}_K$ and $\mathcal{X}$ is the quantity defined by

$$W_p(Q,P) = \inf_{f_c \in \Pi_c(Q,P)} \left\{ \int_{\mathcal{X}_K \times \mathcal{X}} \|x - y\|_p^p df_c(x,y) \right\} \quad (1)$$

where $\|.\|_p$ denotes the ℓ^p norm and $\Pi_c(Q,P)$ the set of probability couplings, with Q and P as its marginals

Theorem ([39] using a result from [1])

If Q and P share the same dependence structure (copula), then

$$W_p^p(Q,P) = \sum_{i=1}^{d} W_p^p(Q_i, P_i). \quad (2)$$

Working on the real line (each dimension of X), the choice of the 2−Wasserstein distance (W_2) leads to

$$W_2(Q_i, P_i) = \sqrt{\int_0^1 \left(F^{\rightarrow}_{Q,i}(x) - F^{\rightarrow}_i(x)\right)^2 dx}$$

with $F^{\rightarrow}$ denoting the generalized inverse cdf, which

- metricizes weak convergence on $\mathcal{P}_2(\mathbb{R}) \Leftrightarrow W_2$ is a measure of proximity on a broad set of probability measures
- simplifies solving the minimization problem by
  - estimating the $F^{\rightarrow}_{Q,i}$ using isotonic polynomials between marginal quantiles, with controlled regularity, which requires to solve a convex quadratic program
  - using gradient descent

Technical details in our recent preprint [39]

Key subjective messages and suggestions

1. Producing clear, auditable rules for prior (Bayesian) modeling should be a more lively research axis
   - Relying on a huge objective corpus
   - Improving the quality of designs
   - Useful for model inversion and sensitivity and robustness analyses
   - Improves the overall interpretability of UQ and decisions
2. Developing / improving nonparametric characterizations in small samples situations
3. Obtaining stronger guarantees on the use of metamodels and ML approaches
   - Relying on nonasymptotic statistics
   - Relying on topological/geometrical constraints of models
   - Relying on conservatisms on uncertain dependences
   - Relying on biased designs and metamodels
4. Sensitivity analyses for extreme situations

[1] Alfonsi, A. and Jourdain, B. (2014). A remark on the optimal transport between two probability measures sharing the same copula. *Statistics & Probability Letters*, 84 :131–134.

[2] Alvarado, J., Steinmann, P., and Estier et al., S. (2014). Anthropogenic radionuclides in atmospheric air over switzerland during the last few decades. *Nature Communications*, 5 :3030.

[3] Ancelet, S., Bousquet, N., and Parent, E. (2022). Optimal nonlinear bayesian designs for structural reliability estimation under experimental cost constraints. application to a functional weibull model of steel fracture toughness. *submitted*.

[4] ASTM (1997). *E399-90 : Standard Test Method for Plane-Strain Fracture Toughness of Metallic Materials. Annual Book of ASTM Standards*. American Society for Testing and Materials International.

[5] AWS (2008). Best practices : destructive testing for material toughness. *Inspection Trends, American Welding Society*, pages 30–31.

[6] Azzimonti, D., Ginsbourger, D., Chevalier, C., Bect, J., and Richet, Y. (2021). Adaptive design of experiments for conservative estimation of excursion sets. *Technometrics*, 63 :13–26.

[7] Baccou, J., Zhang, J., Fillion, P., Damblin, G., Petruzzi, A., Mendizabal, R., Reventos, F., Skorek, T., Couplet, M., Iooss, B., Oh, D.-Y., Takeda, T., and Sandberg, N. (2020). SAPIUM : a generic framework for a practical and transparent quantification of thermal hydraulic code model input uncertainty. *Nuclear Science and Engineering*, 194 :721–736.

[8] Barbillon, P., Celeux, G., Grimaud, A., Lefebvre, Y., and Rocquigny (de), E. (2011). Non linear methods for inverse statistical problems. 55 :132–142.

[9] Bect, J., B. N. I. B., Liu, S., Mabille, A., Popelin, A.-L., Rivière, T., Stroh, R., Sueur, R., and Vazquez, E. (2014). Uncertainty quantification and reduction for the monotonicity properties of expensive-to-evaluate computer models. *Proceedings of UCM 14, Sheffield*.

[10] Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22 :773–793.

[11] Benoumechiara, N., Michel, B., Saint-Pierre, P., and Bousquet, N. (2020). Detecting and modeling critical dependence structures between random inputs of computer models. *In revision for Dependence Modeling, manuscript available on demand*.

[12] Berry, S., Carlin, B., Lee, J., and Muller, P. (2011). *Bayesian Adaptive Methods for Clinical Trials*. FDA Guidelines.

[13] Blatman, G., Delage, T., Iooss, B., and Perot, N. (2017). Probabilistic risk bounds for the characterization of radiological contamination. *EPJ N - Nuclear Sciences Technologies*.

[14] Borovicka, T., Jr., M. J., Kordik, P., and Jirina, M. (2012). Selecting representative data sets. In Karahoca, A., editor, *Advances in Data Mining Knowledge Discovery and Applications*, chapter 2. IntechOpen, Rijeka.

[15] Bousquet, N. (2008). Diagnostics of prior-data agreement in applied Bayesian analysis. *Journal of Applied Statistics*, 35 :1011–1029.

[16] Bousquet, N. (2012). Accelerated Monte Carlo estimation of exceedance probabilities under monotonicity constraints. *Annales de la Faculté des Sciences de Toulouse*, 21 :557–591.

[17] Bousquet, N., Blazère, M., and Cerbelaud, T. (2022). Prior constraints of well-posedness in stochastic inversion problems of computer models. *In revision for the Electronic Journal of Statistics*.

[18] Bousquet, N., Fouladirad, M., Grall, A., and Paroissin, C. (2015). Bayesian gamma processes for optimizing condition-based maintenance under uncertainty. *Applied Stochastic Models in Business and Industry*, 31 :360–379.

[19] Bousquet, N., Klein, T., and Moutoussamy, V. (2018). Approximation of limit state surfaces in monotonic Monte Carlo settings, with applications to classification. *SIAM Journal on Uncertainty Quantification*, 6 :1–33.

[20] Celeux, G., Grimaud, A., Lefebvre, Y., and Rocquigny (de), E. (2010). Identifying intrinsic variability in multivariate systems through linearised inverse methods. 18 :401–415.

[21] Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design : a review. *Statistical Science*, 10(3) :273–304.

[22] Chevalier, C. (2013). *Fast uncertainty reduction strategies relying on Gaussian process models*. Theses, Universität Bern.

[23] Clemmensen, L. and Kjaersgaard, R. (2022). Data representativity for machine learning an ai systems. *arXiv :2203.04706v1*.

[24] Da Veiga, S. and Marrel, A. (2012). Gaussian process modeling with inequality constraints. *Annales de la Faculté des Sciences de Toulouse*, 21 :529–555.

[25] Darach, O. and Garot, J. (2011). Aradiation-induced heart disease. *Circulation : Heart Failure*, 4.

[26] Daszykowski, M., Walczak, B., and Massart, D. (2002). Representative subset selection. *Analytica Chimica Acta*, 468(1) :91–103.

[27] de Valpine, P. and Hilborn, R. (2005). State-space likelihoods for nonlinear fisheries time-series. *Can. J. Fish. Aqu. Sc.*, 62 :1937–1952.

[28] Ducoffe, M., Gerchinovitz, S., and Sen Gupta, J. (2020). A High-Probability Safety Guarantee for Shifted Neural Network Surrogates. *Proceedings of SafeAI 2020*, pages 74–82.

[29] Fillion, N. and Corless, R. (2014). On the epistemological analysis of modeling and computational error in the mathematical sciences. *Synthese*, 191 :1451–1467.

[30] Fortuin, V. (2022). Priors in Bayesian Deep Learning : A Review. *International Statistical Review*.

[31] Fu, S., Celeux, G., Bousquet, N., and Couplet, M. (2015). Bayesian inference for inverse problems occurring in uncertainty analysis. 5 :73–98.

[32] Fu, S., Couplet, M., and Bousquet, N. (2017). An adaptive kriging method for solving nonlinear inverse statistical problems. 28.

[33] Gauffriau, A., Malgouyres, F., and Ducoffe, M. (2021). Overestimation learning with guarantees. In *AAAI-21, workshop on safeAI*, Valence (Virtual), Spain.

[34] Gelman, A., Simpson, D., and Betancourt, M. (2017). The Prior can often only be understood in the context of the Likelihood. *Entropy*, 19(10).

[35] Gelman, A. and Yao, Y. (2020). Holes in Bayesian statistics. *Journal of Physics G : Nuclear and Particle Physics*, 48(1) :014002.

[36] Giovagnoli, A. and Verdinelli, I. (2018). Bayesian randomized adaptive designs with a compound utility function. *Proceedings of the CIRM Conference on New Challenges on Designs of Experiments*.

[37] Hasofer, A. (1968). A statistical theory of the brittle fracture of steel. *International Journal of Fracture*, 4 :439–452.

[38] Hassler, E. (2015). Bayesian d-optimal issues and optimal design construction.

[39] Il Idrissi, M., Bousquet, N., Loubes, J.-M., Iooss, B., and Gamboa, F. (2022). Quantile-constrained Wasserstein projections for robust interpretability of numerical and machine learning models. *https://arxiv.org/abs/2209.11539*.

[40] Kolassa, J. and Kuffner, T. (2020). On the validity of the formal edgeworth expansion for posterior densities. *Annals of Statistics*, 48(4) :1940–1958.

[41] Lazor, R., Lovis, A., Nicod, L., and Cornuz, J. (2012). Dépistage du cancer pulmonaire par scanner thoracique. *Rev Med Suisse*, 363 :2206–2211.

[42] López-Lopera, A. F., Bachoc, F., and Roustant, O. (2022). High-dimensional additive gaussian processes under monotonicity constraints.

[43] Mahmoud, M. A. M. (2013). Reliability of using standard penetration test (spt) in predicting properties of silty clay with sand soil. *International Journal of Civil and Structural Engineering*, 3 :545–556.

[44] Miyata, T. and Tagawa, T. (2002). Mezzo-scopic analysis of fracture toughness in steels. *Materials Research*, 5.

[45] Molnar, C. (2022). Interpretable machine learning : A guide for making black-box models explainable (2nd ed.). https://christophm.github.io/interpretable-ml-book/.

[46] Monte, L. (2013). Predicting the effect of ionising radiation on biological populations : testing of a non-linear leslie model applied to a small mammal population. *Journal of Environmental Radioactivity*, 122 :63–69.

[47] Moutoussamy, V. (2015). *Contributions to structural reliability analysis : accounting for monotonicity constraints in numerical models*. Ph.D. Thesis, Université Toulouse III.

[48] Overstall, A., McGree, J., and Drovandi, C. (2018). An approach for finding fully bayesian optimal designs using normal-based approximations to loss functions. *Statistics and Computing*.

[49] Prangle, D., Harbisher, S., and Gillespie, C. S. (2022). Bayesian Experimental Design Without Posterior Calculations : An Adversarial Approach. *Bayesian Analysis*, pages 1 – 31.

[50] Pronzato, L. (2017). Minimax and maximin space-filling designs : some properties and methods for construction. *Journal de la Socité Française de Statistique*, 158 :7–36.

[51] Pronzato, L. (2021). Performance analysis of greedy algorithms for minimising a maximum mean discrepancy. *arXiv :2101.07564*.

[52] Robert, C. (2007). *The Bayesian Choice : From Decision-Theoretic Foundations to Computational Implementation (2nd edition)*. Springer.

[53] Song, F., Liu, J., Zhang, T., Guo, J., Tian, S., and Xiong, D. (2020). The Grey Forecasting Model for the Medium-and Long-Term Load Forecasting. *Journal of Physics : Conference Series*, 1654 :012104.

[54] Sprenger, J. (2018). The objectivity of subjective Bayesianism. *European Journal for Philosophy of Science*, 69 :539–558.

[55] Strnadel, B. and Haušild, P. (2008). Statistical scatter in the fracture toughness and charpy impact energy of pearlitic steel. *Material Science and Engineering*, 486 :208–214.

[56] Tong, M., Duan, H., and He, L. (1973). A novel Grey Verhulst model and its application in forecasting co2 emissions. *Environmental Science and Pollution Research*, 28 :31370–31379.

[57] Vidot, G., Ducoffe, M., Gabreau, C., Ober, I., and Ober, I. (2022). Formal monotony analysis ofÂ neural networks withÂ mixed inputs : An asset forÂ certification. In Groote, J. F. and Huisman, M., editors, *Formal Methods for Industrial Critical Systems*, pages 15–31, Cham. Springer International Publishing.

[58] Vives i Batlle, J., Sazykina, T., Kryshev, A., Wood, M., Smith, K., Copplestone, D., and Biermans, G. (2020). Modelling the effects of ionising radiation on a vole population from the chernobyl red forest in an ecological context. *Ecological Modelling*, 438 :109306.

[59] Wallin, K. (2002). Master curve analysis of the "euro" fracture toughness dataset. *Engineering Fracture Mechanics*, 69 :451–481.

[60] Watanabe, Y., Dahlman, E., Leder, K., and Hui, S. (2016). A mathematical model of tumor growth and its response to single irradiation. *Theoretical Biology and Medical Modelling*, 13 :6.

[61] Xian, W., Long, B., Li, M., and Wang, H. (2014). Prognostics of lithium-ion batteries based on the verhulst model, particle swarm optimization and particle filter. *IEEE Transactions on Instrumentation and Measurement*, 63(1) :2–17.

[62] Yang, S., Vong, W., Sojitra, R., Folke, T., and Shafto, P. (2021). Mitigating belief projection in explainable artificial intelligence via bayesian teaching. *Scientific Reports*, 11.