14th International Conference on Structural Safety and Reliability - ICOSSAR'25
1-6 June, 2025 - Los Angeles California, USA
R. Ghanem, E. Johnson, S. Masri and A. Olivier (Eds)

GUIDING ADAPTIVE FIRE TESTING THROUGH EXPECTED INFORMATION GAIN

ANDREA FRANCHINI[1] AND RUBEN VAN COILE[2]

[1]Ghent University
Technologiepark-Zwijnaarde 60, 9052 Ghent (BE)
Andrea.Franchini@UGent.be

[2]Ghent University
Technologiepark-Zwijnaarde 60, 9052 Ghent (BE)
Ruben.VanCoile@UGent.be

Key words: Fire Safety, Experimental Design, Expected Information Gain, Flammability Testing, Number of Trials

Abstract. This paper presents initial insights into a fire design acceptance framework that applies Adaptive Fire Testing (AFireTest) principles to support more informed and efficient planning of testing campaigns. A core feature of AFireTest is the pre-execution quantification of a test's expected utility, evaluated in terms of campaign objectives such as uncertainty reduction, cost-efficiency, and environmental impact. We illustrate this evaluation with a bench-scale flammability testing case study. In this context, we use Expected Information Gain (EIG) to estimate how varying trial numbers are expected to reduce uncertainty in ignition time predictions. Results show that EIG stabilises between 10 and 15 trials, suggesting a rational stopping point for the test campaign. Validation with experimental data demonstrates the reliability of EIG as a tool for guiding test design.

1 INTRODUCTION

The current fire safety paradigm relies on standardised tests developed as part of a prescriptive design framework, which fails to provide an in-depth understanding of construction products' fire performance. Design solutions which comply with the prescriptive guidance are assumed to achieve an adequate level of fire safety [1]. In the wake of demonstrated inadequate performance (i.e. following fire disasters), the prescriptive guidance is updated [2]. The resulting incomplete fire performance characterisation hampers people's safety and the much-needed innovation in the built environment. Furthermore, in its current format, standardised testing does not provide the necessary in-depth understanding of construction products' fire performance needed for performance-based design (e.g. [3,4])

To cope with these limitations, the Adaptive Fire Testing (AFireTest) project at Ghent University aims to develop a fire design acceptance framework which exploits AFireTest principles. Figure 1 provides an overview of the overall acceptance framework. We envision the prescriptive design of the 21st century to rely on an explicit fire performance evaluation through surrogate models. A design's fire risk can then be assessed quasi-instantaneously and taking into account the best currently available data. If the submitted design does not meet the acceptance thresholds, a number of actions are possible to the architect/designer: (i) the design can be modified; (ii) the performance of the design can be evaluated in depth by a fire engineer using advanced Fire Safety Science and Engineering (FSSE) models; (iii) targeted fire tests can be performed to reduce the uncertainty in fire performance. Naturally, these options can be combined. Upon acceptance, the advanced evaluations performed under (ii) and the fire tests performed under (iii) provide input to the Authority Having Jurisdiction (AHJ) to extend the surrogate model's applicability range.

At the core of this design acceptance framework is the identification and definition of optimal testing protocols which maximise the test's utility. By "utility", we mean a measure of the benefit of conducting the test—which raises the following question: how should the expected utility of a fire test be quantified before its execution?

This paper presents initial insights into this quantification process. Specifically, we propose quantifying a fire test's expected utility in terms of several metrics of interest to a stakeholder (e.g. uncertainty, cost, or embodied carbon reduction). In this quantification, we use Bayesian inference (e.g. [5-7]) to incorporate: the prior beliefs on the probabilistic distribution of uncertain parameters (e.g. fire resistance, time to ignition, heat release rate); and the information that may be gained from potential experimental outcomes. We demonstrate this approach through an application in bench-scale flammability testing, investigating how uncertainty in predicting PMMA ignition time is expected to reduce with the number of tests performed.

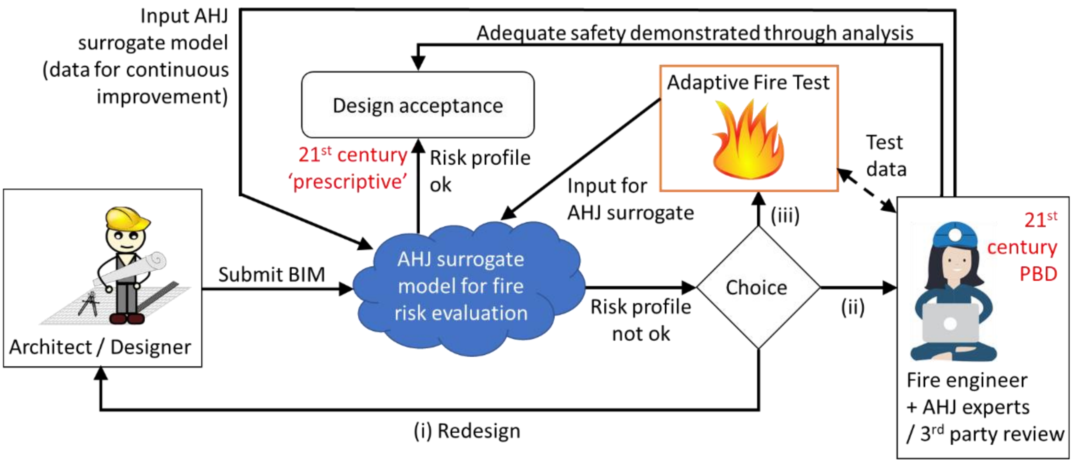


Figure 1: Envisioned fire design acceptance in the 21st century by comprehensive performance evaluation through adaptive fire testing.

2 FIRE TEST UTILITY QUANTIFICATION

Bayesian inference is a model-based process by which a current state of knowledge on uncertain parameters of interest $\boldsymbol{\theta}$ is updated based on evidence from newly collected data (e.g. [5]). Knowledge states are expressed through probabilistic distributions that we generally refer to as $p(\cdot)$. After performing a test with testing parameters $\boldsymbol{\xi}$ (e.g. testing temperatures, specimen dimensions, loading forces), and observing the data $\boldsymbol{y}$, we can update our prior state of knowledge using Bayes' theorem in the context of posterior Bayesian analysis:

$$p(\boldsymbol{\theta}|\boldsymbol{\xi}\,,\boldsymbol{y},\mathcal{M}_{exp}) = \frac{p(\boldsymbol{\theta})\,p(\boldsymbol{y}|\boldsymbol{\xi},\boldsymbol{\theta},\mathcal{M}_{exp})}{p(\boldsymbol{y}|\boldsymbol{\xi},\mathcal{M}_{exp})} \tag{1}$$

In this equation, $p(\boldsymbol{\theta}|\boldsymbol{\xi}\,,\boldsymbol{y},\mathcal{M}_{exp})$ is the posterior distribution, reflecting the updated state of knowledge; $\mathcal{M}_{exp}$ is an experimental model that simulates the outcomes of the test we are considering; $p(\boldsymbol{\theta})$ is the prior distribution, incorporating available knowledge prior to the test; $p(\boldsymbol{y}|\boldsymbol{\xi},\boldsymbol{\theta},\mathcal{M}_{exp})$ is the likelihood model, which provides the probability of observing the data $\boldsymbol{y}$; $p(\boldsymbol{y}|\boldsymbol{\xi},\mathcal{M}_{exp})$ is the marginal probability of the data.

When planning a testing campaign, the data $\boldsymbol{y}$ are not yet available. Therefore, Eq. (1) cannot be directly applied. However, potential data outcomes can be simulated using the prior predictive distribution [5], i.e. the expected distribution of the data according to the prior and the likelihood. Then, posterior distributions can be obtained for each simulated data—a process named pre-posterior Bayesian analysis.

Eventually, we want to use the prior and pre-posterior distributions to define "utility functions" U, i.e. mathematical representations that assign numerical values to possible testing outcomes based on their relative desirability to the decision-maker. The expected utility of a fire test using testing parameters $\boldsymbol{\xi}$ is then obtained as follows:

$$U_{test}(\boldsymbol{\xi}) = U_{pre-post}\left(p(\boldsymbol{\theta}|\boldsymbol{\xi}\,,\boldsymbol{y},\mathcal{M}_{exp})\right) - U_{prior}(p(\boldsymbol{\theta})) \tag{2}$$

The pre-posterior utility function is usually obtained as an expected value with respect to the prior predictive distribution. Within the AFireTest project, we are formulating utility functions that quantify the potential benefit of a test for diverse experimental goals, such as:

- Reducing uncertainty in the distribution of target parameters entering FSSE predictive models (e.g. material properties, ignition time, heat release rates, fuel load);
- Reducing uncertainty in estimating system-level performance metrics (e.g. failure probability, risk) for performance-based fire assessment;
- Enhancing confidence in the classification of construction products or identifying test specifications that most effectively demonstrate the likelihood of a product belonging to a particular class.
- Maximising the net lifecycle economic benefit of conducting the test, where "net" refers to the benefit after deducting the cost of the experiment
- Minimising fire safety's environmental impact, accounting for testing and lifecycle emissions.

Without loss of generality, in the following, we focus on quantifying how much a test is expected to reduce uncertainty on a target parameter of interest. To that end, the prior and pre-posterior utility functions can be defined as the negative of Shannon entropies $\mathcal{H}$ [8] of the prior and pre-posterior distributions, with a larger $\mathcal{H}$-value corresponding with a higher measure of uncertainty. The difference in utility is then the reduction of the entropy between the prior and the posterior, which is named information gain IG [7]. For pre-posterior Bayesian analysis, (i.e. before conducting the test), the test utility function can be defined as the expected value of the information gain, named expected information gain, EIG:

$$U_{test}(\boldsymbol{\xi}) = \mathbb{E}\left[\mathcal{H}[p(\boldsymbol{\theta})] - \mathcal{H}[p(\boldsymbol{\theta}|\boldsymbol{\xi}\,,\boldsymbol{y},\mathcal{M}_{exp})]\right] = EIG(\boldsymbol{\xi}) \tag{3}$$

3 ILLUSTRATIVE EXAMPLE

3.1 Example description

Assume we want to experimentally reduce uncertainty in predicting the ignition time t_{ig} of a polymethyl methacrylate (PMMA) batch for an industrial application: specifically, the PMMA is to be used as the cover of an item at risk of exposure to heat flux up to 20 kW/m^2. The ignition time is critical for modelling flame spread over combustible surfaces [9]. A testing approach is chosen consisting of a series of bench-scale flammability tests through a cone calorimeter with an exposure of 20 kW/m^2. The Cone Calorimeter experiments are to be conducted with 6 mm thick black PMMA plates with a heated area of 100 x 100 mm and following the ASTM E1354 standard. Further details on the testing procedure can be found in the work by Morrisset et al. [10]. As suggested in the latter study, we assume the ignition time is normally distributed—i.e. $t_{ig} \sim Normal(\mu_{tig}, \sigma_{tig})$, with mean μ_{tig} and standard deviation σ_{tig}. The mean and standard deviation are unknown, and we take them as the uncertain parameters of interest ($\boldsymbol{\theta}$ in Eq. (1-3)). The uncertainty in these parameters propagates to the ignition time prediction. We assume the only available testing parameter is the number of trials, which is denoted as ξ hereinafter. How does the expected uncertainty in ignition time decrease with the number of trials?

3.2 Pre-posterior analysis

We use pre-posterior Bayesian analysis to tackle the question above. Table 1 summarises the mathematical formulations of the prior, likelihood, and experimental model. As a starting point, a prior distribution for the uncertain parameters of interest needs to be defined. Using an ignition time model for thermally thin solids and material parameters from [11], we estimate that the ignition time lies in the range $[t_{ig,min}, t_{ig,max}] = [139\text{ s}, 214\text{ s}]$. Then, assuming that μ_{tig} and σ_{tig} are uncorrelated, we define their priors as uninformative uniform distributions. Because of the assumption on the ignition time distribution (see Sec. 3.1), the experimental model linking the experimental observations (i.e. ignition time measurements) to the uncertain parameters of interest is, again, the normal distribution model. Then, the likelihood model, which provides the probability of observing a set of cone calorimeter measurements $\mathbf{y} = [t_{ig,1}, \cdots, t_{ig,\xi}]$, is a product of normal distributions.

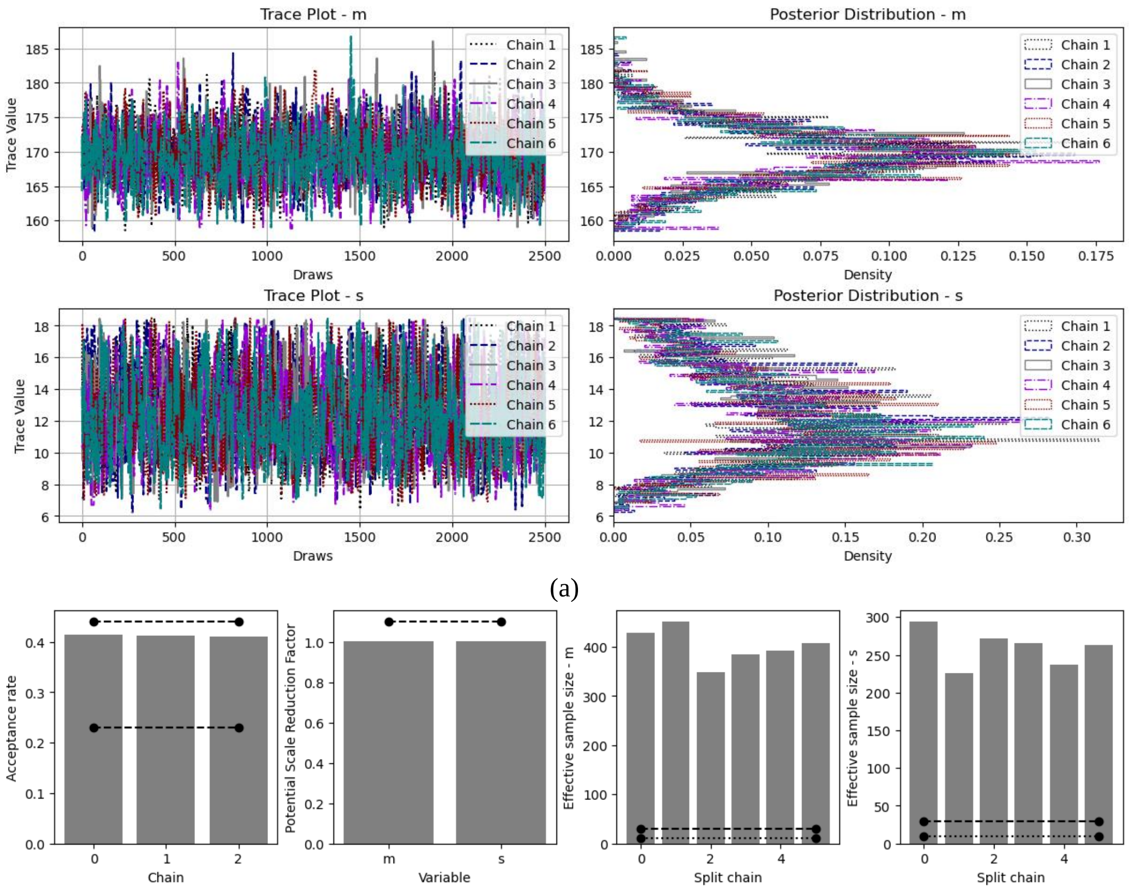
Table 1: Mathematical formulations for Bayesian modelling. $\Delta t = t_{ig,max} - t_{ig,min}$. f_{normal} is the probability density function of the normal distribution.

Prior distribution	$\mathrm{p}(\boldsymbol{\theta}) = [\mathrm{Uniform}(t_{ig,min} + \Delta t/4, t_{ig,min} + 3\Delta t/4), Uniform(0, \Delta t/4)]$
Experimental model	$\mathcal{M}_{exp}(\mu_{tig}, \sigma_{tig}) = Normal(\mu_{tig}, \sigma_{tig})$
Likelihood model	$p(\boldsymbol{y}\mid\xi, \boldsymbol{\theta}, \mathcal{M}_{exp}) = \prod_{i=1}^{\xi} f_{normal}(y_i, \mu_{tig}, \sigma_{tig})$

Generally, a closed-form solution of Eq. (3) does not exist. Thus, we seek an approximate numerical solution. Specifically, we simulate realisations $\boldsymbol{y}_i$ of the cone calorimeter data using the prior predictive distribution. Then, we use these data to perform Bayesian updating and estimate the expected information gain:

$$EIG(\xi) \approx \frac{1}{N_{IGsamples}} \sum_{i=1}^{N_{IGsamples}} \left(\mathcal{H}[p(\boldsymbol{\theta})] - \mathcal{H}[p(\boldsymbol{\theta}|\xi, \boldsymbol{y}_i, \mathcal{M}_{exp})]\right) = \frac{1}{N_{IGsamples}} \sum_{i=1}^{N_{IGsamples}} IG_i \quad (4)$$

For each realisation $\boldsymbol{y}_i$, we solve Eq. (1) using Markov Chain Monte Carlo sampling via the Metropolis-Hastings algorithm [5], which enables drawing samples from the posterior distribution.



(a)

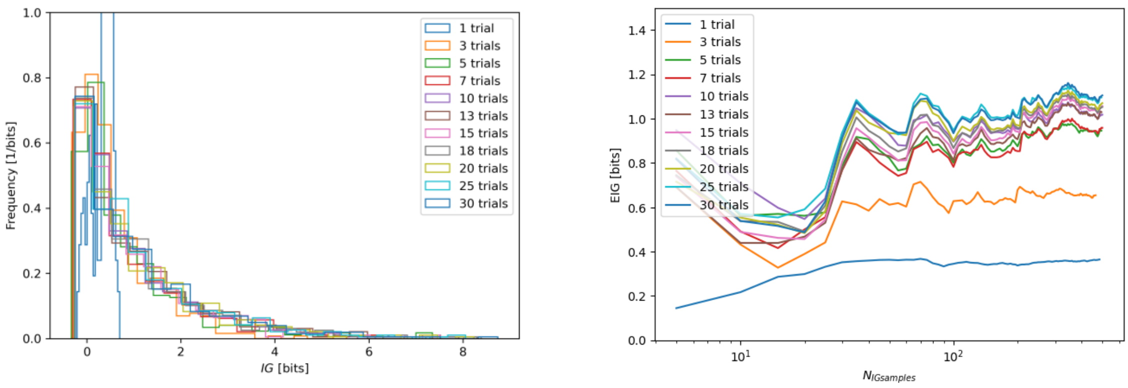
(b)

Figure 2: Example convergence check for a Bayesian update using Markov-Chain Monte Carlo. (a) Trace plots and posterior samples. (b) Convergence metrics. Note: $m = \mu_{tig}; s = \sigma_{tig}$.

For each Bayesian update, we assess convergence of the posterior distributions for all uncertain variables by examining mixing and stationarity [5]. More precisely, for a given data realisation $\boldsymbol{y}_i$, we initiate three chains with random starting points. We discard the initial half of each chain as warm-up to mitigate the influence of starting values, then split the remaining samples to jointly assess mixing and stationarity. Convergence is diagnosed using the potential scale reduction factor (PSRF), which compares between- and within-chain variances, and the effective sample size (ESS), which accounts for autocorrelation. We proceed once PSRF<1.01 and the ESS exceeds 5 times the number of split chains for all the uncertain variables of interest.

An ESS of 10 per sequence typically corresponds to stability. Figure 2 provides a convergence check example.

We conduct the information gain analysis for $\xi = 1, 3, 5, 7, 10, 13, 15, 18, 20, 25, 30$ trials. Figure 3a shows the resulting information gain distributions (IG_i in Eq. (5)), which appear right-skewed. The expected value of these distributions provides an estimate of the EIG for the considered number of trials. The number of samples $N_{IGsamples}$ was chosen to achieve stability of EIG (see Figure 3b).



(a) (b)

Figure 3: Pre-posterior analysis results. (a) Information gain (IG) distributions for different number of trials. (b) Convergence check on the expected information gain (EIG).

Figure 4 shows the expected information gain for the considered number of trials. The EIG grows with the number of trials, but the rate of growth decreases as the number of trials becomes larger, indicating diminishing returns. The EIG stabilises around 10 to 15 trials, beyond which the marginal expected information gain becomes negligible. The maximum observed EIG occurs at 30 trials (the highest number considered). Based on these results, we conclude that commissioning between 10 and 15 trials would maximise information gain while maintaining testing efficiency. More accurate decision-making criteria (e.g. a threshold on the marginal EIG) could be defined, but they fall outside the scope of this paper.

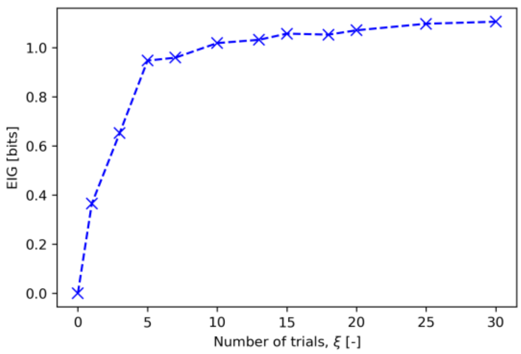
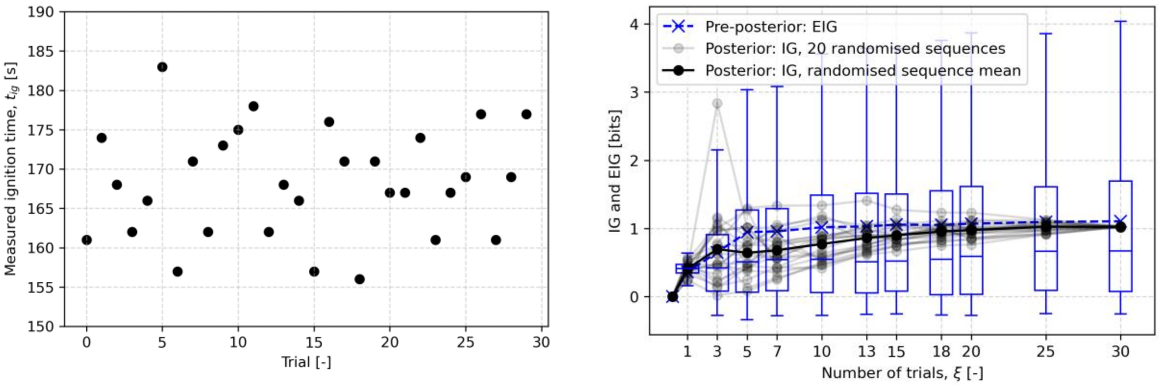


Figure 4: Expected information gain (EIG).

3.3 Result validation

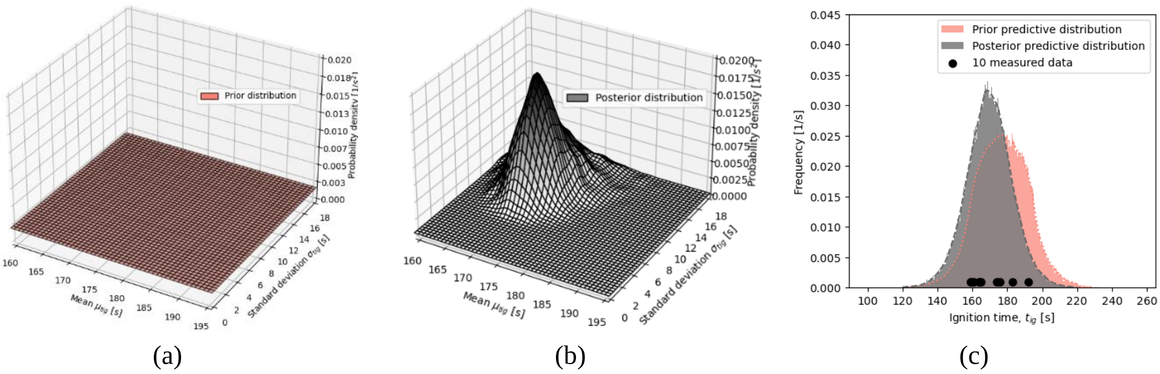
We now assume that the tests have been conducted and the 30 ignition time measurements shown in Figure 5a have been obtained. These data are drawn from the experimental results reported by Morrisset et al. [10]. We randomly shuffle these measurements and generate 20 distinct sequences. Next, for each sequence, we iteratively update the prior distribution of μ_{tig} and σ_{tig} after each considered trial, using all measurements collected up to that point. Finally, for each updated distribution, we calculate the information gain as the difference between the prior and the posterior entropies $\mathcal{H}$ of the ignition time predictive distributions.



(a) (b)

Figure 5: Pre-posterior analysis validation. (a) Considered dataset. (b) Comparison between pre-posterior and posterior Bayesian analysis.

Figure 6 provides an example of posterior Bayesian updating performed for one of the 20 sequences and considering 10 trials. The ignition time predictive distributions (Figure 6c) are obtained by sampling realisations of μ_{tig} and σ_{tig} from the corresponding prior and posterior distributions (Figure 6a-b, respectively) and using the experimental model $\mathcal{M}_{exp}$. The information extracted from the data reduces uncertainty in the ignition time prediction, resulting in a narrower posterior ignition time predictive distribution and a positive information gain.



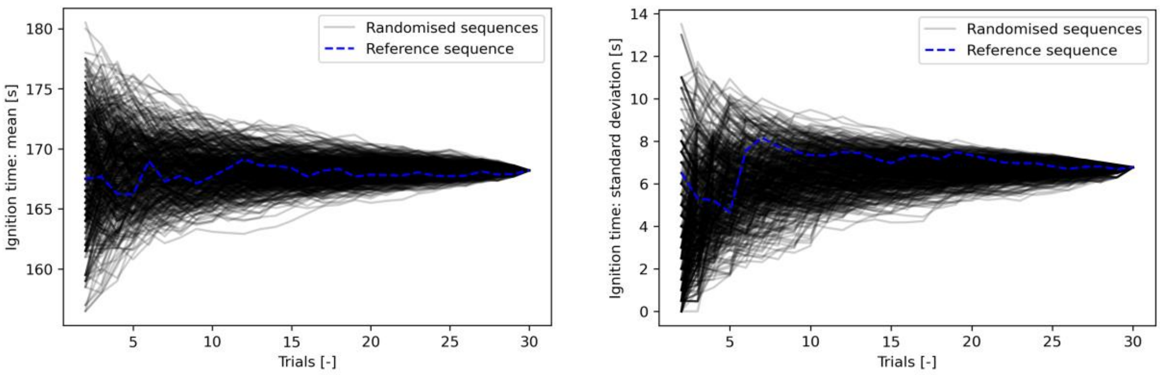
(a) (b) (c)

Figure 6: Example of posterior Bayesian updating with 10 measured data, i.e. with a possible outcome of 10 trials.

The obtained IG values for the 20 randomised sequences are presented in Figure 5b. This figure also includes the expected information gain (EIG) estimates from the pre-posterior analysis (see Figure 4), along with boxplots summarising the distributional statistics of the pre-posterior IG values presented in Figure 3a. It can be observed that the posterior IG values generally lie within the interquartile range predicted by the pre-posterior analysis. Posterior results are more scattered at low trials. However, their variability reduces as the number of trials increases. All sequences provide the same value at 30 trials because, at that point, they account for all the available data. The mean value of the posterior IG aligns well with the trend predicted by the pre-posterior analysis, validating the pre-posterior analysis. These findings demonstrate the effectiveness of the proposed approach in predicting experimental outcomes and informing decision-making. Morrisset et al. [10] reached conclusions similar to those presented in this paper regarding the required number of trials, using *a-posteriori* Gaussian statistics in their analysis.

3.4 Influence of observation order

In this section, we discuss how the order in which the measurement are obtained affects the information gain from posterior analysis. Indeed, the information gain trajectories in Figure 5b exhibit considerable variability in the early stages, primarily due to the order in which measurements are observed. For example, a new data point that is consistent with previous observations typically leads to an increase in IG, whereas a more divergent measurement may cause a decrease. As the number of trials grows, the impact of observation order diminishes, and all trajectories gradually converge toward a stable value of IG. This asymptotic value reflects an intrinsic property of the specific dataset under consideration; different datasets would exhibit distinct total information content and corresponding IG trajectories. In contrast, the pre-posterior analysis yields a smoother and more stable trend for the expected information gain, as it averages over many possible data sequences, effectively capturing the mean behaviour of the posterior-analysis IG trajectories.



(a) (b)

Figure 7: Influence of measurement order on a-posteriori mean and standard deviation with Gaussian analysis.

The order in which the measurements are obtained also affects conventional a-posteriori Gaussian analysis. For example, Figure 7 shows how the sample mean and standard deviation of ignition time evolve for the dataset described in Figure 5a as more data points are included, across 500 randomised measurement orders. By “reference sequence,” we mean the data in the same order as Figure 5a.

The mean estimates (Figure 7a) show significant variation in early trials, as each added data point considerably affects the average. However, as the number of trials increases, the curves converge toward a stable mean value. A similar pattern is visible for the standard deviation (Figure 7b). Thus, for both frameworks (frequentist in Figure 7 and posterior Bayesian in Figure 5b), the high variability at a low number of trials diminishes with more data, leading to convergence toward a stable value. Information gain trajectories can decrease if new data increases uncertainty (e.g. an outlier), while the standard deviation is more symmetric in how it responds to such inputs. However, the posterior IG values begin to stabilise much earlier. Notably, the EIG curve is consistently smooth because it marginalises over all possible observations expected (but yet-not-observed) from planned tests, capturing the average behaviour across all data orderings.

4 CONCLUSIONS

In this paper, we presented initial insights into a Bayesian decision-theoretic framework for quantifying the expected utility of fire tests before their execution. The main conclusions are the following:

- Bayesian pre-posterior analysis allows quantifying a fire test utility before data are collected, supporting optimised *a-priori* selection of testing parameters (e.g. number of trials, testing temperatures, specimen dimensions, loading forces).
- Within the AFireTest project at UGent, we are developing utility functions to guide fire test planning across diverse goals, including uncertainty reduction in predictive models and system-level metrics (e.g. safety, risk, resilience), improved product classification, net lifecycle economic benefit optimisation, and minimisation of fire safety's environmental impact.
- We demonstrated the proposed framework in the context of bench-scale flammability testing. The Expected Information Gain (EIG) increases with the number of trials, but exhibits diminishing returns, stabilising between 10–15 trials, which suggests a rational stopping point for the test campaign.
- Validation with posterior analysis using real ignition time data confirms that EIG estimates align with actual information gain from testing, demonstrating the method's predictive capacity.
- Data ordering significantly affects early-stage posterior Information Gain (for Bayesian analysis) and measures of spread around the mean (for frequentist analysis). The EIG is consistently smooth because it averages over all possible observations that could result from future, yet-unobserved test data.

ACKNOWLEDGEMENTS

This work is funded by the European Union (ERC, AFireTest, 101075556). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] Van Coile, R., Hopkin, D., Lange, D., Jomaas, G., & Bisby, L. (2019). The need for hierarchies of acceptance criteria for probabilistic risk assessments in fire engineering. Fire Technology, 55(4), 1111-1146.
[2] Spinardi, G., Bisby, L., & Torero, J. (2017). A review of sociological issues in fire safety regulation. Fire Technology, 53(3), 1011-1037.
[3] Bisby, L., Gales, J., & Maluk, C. (2013). A contemporary review of large-scale non-standard structural fire testing. Fire Science Reviews, 2(1), 1.
[4] Torero, J. L. (2013). Scaling-up fire. Proceedings of the Combustion Institute, 34(1), 99-124.
[5] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2021) Bayesian data analysis. Third edition. Electronic edition.
[6] Rainforth, T., Foster, A., Ivanova, D. R., & Bickford Smith, F. (2024). Modern Bayesian experimental design. Statistical Science, 39(1), 100-114.
[7] Lindley, D.V. (1956). On a measure of the information provided by an experiment. The Annals of Mathematical Statistics, 27(4), 986-1005.
[8] Shannon, C.E. (1948). A mathematical theory of communication. The Bell system technical journal, 27(3), 379-423.
[9] Hittini, W., Wiesner, F., Lange, D., & Hidalgo, J. P. (2024). A taxonomy of concurrent upward flame spread models and sources of uncertainty. International Journal of Thermal Sciences, 204, 109218.
[10] Morrisset, D., Thorncroft, G., Hadden, R., Law, A., & Emberley, R. (2021). Statistical uncertainty in bench-scale flammability tests. Fire safety journal, 122, 103335.
[11] Drysdale, D. (2011). An introduction to fire dynamics. John Wiley & Sons.