From Environmental Simulators to Building-Scale Risk Maps: A Digital Twin Framework for Multi-Hazard Risk Assessment

Thomas Dau[1], Beatriz Moya[2,*], Amine Ammar[3,1], Boris Huljak[1], Sergio Torregrosa[2], and Francisco Chinesta[2,1]

[1]CNRS@CREATE LTD, 1 Create Way, #08-01 CREATE Tower, Singapore 138602
[2]ENSAM Institute of Technology, PIMM, UMR CNRS 8006, ESI/Keysight Chair, 151 Bd de l'Hôpital, 75013 Paris, France
[3]LAMPA Laboratory & ESI Group Chair, Arts et Metiers Institute of Technology, 2 boulevard du Ronceray, BP 93525, 49035 Angers cedex 01, France
[*]Corresponding author: beatriz.moya_garcia@ensam.eu

Abstract

Risk maps are traditionally constructed from historical observations of hazards and impacts. Yet such records are often sparse, incomplete or poorly representative of future conditions, limiting their usefulness for assessing rare or emerging events. This paper proposes a Digital-Twin-based methodological framework for multi-hazard risk analysis that generates synthetic, forward-looking hazard scenarios and translates them into building-scale risk indicators. The framework combines physics-based simulators for atmospheric pollution, wildfire spread and surface flooding with scenario generation informed by environmental drivers such as wind regimes, ignition locations, rainfall intensities and soil parameters. These scenarios are propagated through a probabilistic workflow to derive empirical distributions, exceedance probabilities, tail-oriented indicators and spatial risk maps. Three case studies demonstrate how heterogeneous hazard models can be organised within a common pipeline. Pollution simulations combine deterministic advection with wind-rose frequencies to produce weighted concentration indicators; wildfire simulations use a stochastic cellular-automaton model to estimate burn probabilities; and flood simulations use a rainfall–soil design of experiments to evaluate maximum water-depth exceedance. The results show that synthetic yet physically grounded scenarios can reveal low-probability, high-impact situations that may be absent from historical records. The framework provides a transparent basis for decision support, urban planning and longer-term risk-informed assessment.

Keywords:

Digital twin; Probabilistic risk assessment; Multi-hazard analysis; Environmental hazards; Building-scale risk; Risk mapping.

Key policy points

- The proposed framework translates heterogeneous environmental hazard simulations into building-scale risk indicators that can support spatial planning, emergency preparedness, and local risk-management decisions.
- By combining scenario generation, probabilistic weighting, and interpretable risk maps, the methodology helps decision-makers compare different hazards within a common and transparent analytical structure.

- The use of synthetic yet physically grounded scenarios can reveal low-probability, high-impact situations that may be absent from historical records but remain important for forward-looking risk assessment.
- The framework is presented as a methodological proof of concept and should be further enriched with calibrated exposure, vulnerability, and socio-economic data before operational deployment.

1 Introduction

Urban territories are increasingly exposed to a broad range of environmental hazards, among which air pollution, wildfires, and surface flooding play a major role (L. Li et al., 2025). This exposure is further amplified by urban densification, the growing complexity of interactions between infrastructures and their environment, and evolving climatic conditions, which increase the occurrence of extreme situations and make past observations less representative of future conditions. In this context, the development of risk-assessment tools capable of anticipating plausible but rarely observed situations has become a central challenge for urban planning, risk management, and decision support (Liang, Moya, Seah, et al., 2025).

Risk maps are still largely built from historical observations of hazards and their impacts. However, such records are often sparse, incomplete, heterogeneous, or only weakly representative of future conditions (De Haan and Rigby, 2020; Abdo et al., 2017; Biswas and Zaman, 2019; Brito et al., 2023). They make it difficult to characterise rare, emerging, or highly conditional events driven by specific combinations of environmental factors. As a result, approaches based exclusively on historical data rapidly reach their limits when the objective is to explore extreme situations, unobserved configurations, or contexts for which feedback data remain scarce.

Physics-based models and Digital Twins provide a relevant framework for addressing these limitations. By relying on transport, propagation, or flow models that explicitly represent the underlying physical mechanisms, they make it possible to generate synthetic yet plausible scenarios consistent with the environment under study. In this work, Digital Twins are not regarded only as digital replicas, but rather as a structuring methodological framework that connects environmental data, simulated scenarios, and risk quantification (AIAA Digital Engineering Integration Committee et al., 2020; Chinesta et al., 2020; Pylianidis et al., 2022; Barbie and Hasselbring, 2024).

The main methodological challenge, however, is not only to simulate a given hazard, but also to organise models of different natures within a common framework and to extract comparable indicators from them (Reis et al., 2009; Liang, Moya, Chinesta, et al., 2025). The outputs of a dispersion model, a wildfire spread model, and a hydraulic model differ in form, physical meaning, and spatial scale. The key issue is therefore to build a consistent processing chain from environmental conditions to scenario definition, hazard simulation, spatial aggregation, and mapped indicators at the building scale. In this context, uncertainty treatment and empirical exceedance statistics play a central role in transforming simulated outputs into probabilistic measures of risk (Hafver et al., 2018; Kapteyn et al., 2021; Ghanem et al., 2022; Baraldi and Zio, 2008).

Simulation-based approaches, including scenario analysis and stochastic sampling, have long been used in pollution dispersion, wildfire propagation, and flooding studies (Hassan et al., 2022; Balter and Faminskaya, 2017; Carmel et al., 2009; C. Li et al., 2023; Falter et al., 2015). However, many of these studies remain hazard-specific and application-driven (C. Li et al., 2023; Carmel et al., 2009; Hassan et al., 2022). The objective of this paper is therefore not to introduce another single-hazard workflow, but to propose a common methodological framework that links environmental data, scenario construction, hazard simulation, uncertainty treatment, and building-scale risk mapping across several environmental hazards.

The proposed framework is illustrated through three case studies: urban air-pollution dispersion, wildfire spread at the wildland–urban interface, and urban surface flooding. The objective is not to define a unique risk indicator shared by all hazards, but to show how a common simulation-to-risk-mapping pipeline can transform heterogeneous model outputs into interpretable building-scale indicators and richer risk maps than those derived from limited historical observations alone (Liang, Moya, Seah, et al., 2025; Liang, Moya, Chinesta, et al., 2025).

This paper is primarily methodological and should be understood as a proof of concept. Its purpose is to show that hazard models of different kinds can be organised within a common guiding framework, and that interpretable risk indicators can be derived from them to support cartographic representations, decision-support tools, and longer-term risk assessment once the framework is enriched with realistic exposure, vulnerability, and calibration data. The remainder of the paper is organised as follows. Section 2 presents the proposed methodology and the three case-study models. Section 3 reports the resulting building-scale indicators and risk maps. Section 4 discusses the interpretation, limitations, and possible extensions of the framework.

2 Materials and Methods

2.1 Overview of the Digital-Twin-based risk-assessment framework

The proposed methodology uses Digital Twins (DTs) as simulation environments for exploring environmental hazards under plausible physical conditions. In this work, DTs are not considered only as digital replicas of existing systems, but as a structuring framework that connects environmental data, scenario generation, hazard simulation, uncertainty propagation, and risk-map construction (AIAA Digital Engineering Integration Committee et al., 2020; Chinesta et al., 2020; Pylianidis et al., 2022; Barbie and Hasselbring, 2024).

The framework is primarily designed for offline exploration rather than real-time operation. Historical or empirical data are used, when available, to inform or weight selected input variables, such as wind conditions or rainfall intensities. Other variables, such as ignition locations in the wildfire case, are generated synthetically when no reliable historical information is available. The resulting scenarios are then propagated through physics-based or stochastic models to compute hazard-specific quantities of interest, such as pollution concentration, burned buildings, or maximum water depth.

The proposed approach combines deterministic and probabilistic elements. Some models are deterministic once the scenario is fixed, as in the pollution-dispersion case. Others include intrinsic stochasticity, as in the wildfire-spread model, where repeated runs are required for a given scenario. In both cases, uncertainty is represented through scenario ensembles and propagated to the building scale through empirical distributions, exceedance probabilities, and tail-oriented indicators. This corresponds to a pragmatic probabilistic Digital-Twin setting, in which uncertainty is treated through scenario sampling and Monte Carlo-type propagation rather than through full Bayesian updating (Hafver et al., 2018; Kapteyn et al., 2021; Ghanem et al., 2022; Baraldi and Zio, 2008).

The scenarios considered in this work remain partly synthetic. They are designed to be physically plausible and, when possible, informed by empirical environmental records, but they are not calibrated against observed damage or impact data. The objective is therefore not to deliver fully operational risk predictions, but to demonstrate a reproducible methodology for translating heterogeneous hazard simulations into interpretable building-scale risk indicators.

2.2 Common terminology and four-step pipeline

The study area denotes the geographical domain over which the hazard is simulated and analysed.

Buildings define the target interpretation scale adopted in this work. Depending on the hazard model, indicators are either computed directly at the building level or first obtained on

a raster support and then interpreted or aggregated in relation to building footprints.

A hazard refers to the physical phenomenon under consideration, such as atmospheric pollution emitted from a source, a fire front propagating through a wildland–urban interface, or surface water accumulation during a rainfall event. A scenario denotes a specific hazard configuration, defined by a combination of environmental conditions and source parameters. Depending on the case study, this may involve wind direction and speed, source location, ignition point, rainfall intensity, soil hydraulic conductivity, or initial soil moisture.

The simulation model is the numerical tool that computes the hazard response over the study area for a given scenario. This response may be a continuous field, such as pollutant concentration or water depth, or a binary state, such as whether a building is reached by fire. Risk indicators are summary quantities derived from these outputs at the building scale. They are defined so as to support spatial comparison between buildings and to enable risk mapping.

The transition from environmental data to risk maps follows four common steps:

1. **Analysis of environmental drivers.** Available environmental data are analysed to identify relevant classes or regimes. Examples include wind roses derived from wind time series, rainfall intensity distributions, or soil-parameter ranges.

2. **Scenario construction.** A finite set of representative scenarios is built by combining the main environmental drivers with source or event locations. The goal is not to exhaust all possible configurations, but to cover a meaningful range of plausible situations.

3. **Hazard simulation.** Each scenario is propagated through the corresponding Digital-Twin model. Depending on the hazard, this may produce concentration fields, burn-impact matrices, or raster maps of maximum water depth.

4. **Aggregation and risk mapping.** Simulation outputs are aggregated across scenarios and, when relevant, across stochastic runs. Scenario weights are applied when empirical frequencies are available. The resulting indicators are then mapped at the building scale or on a raster support.

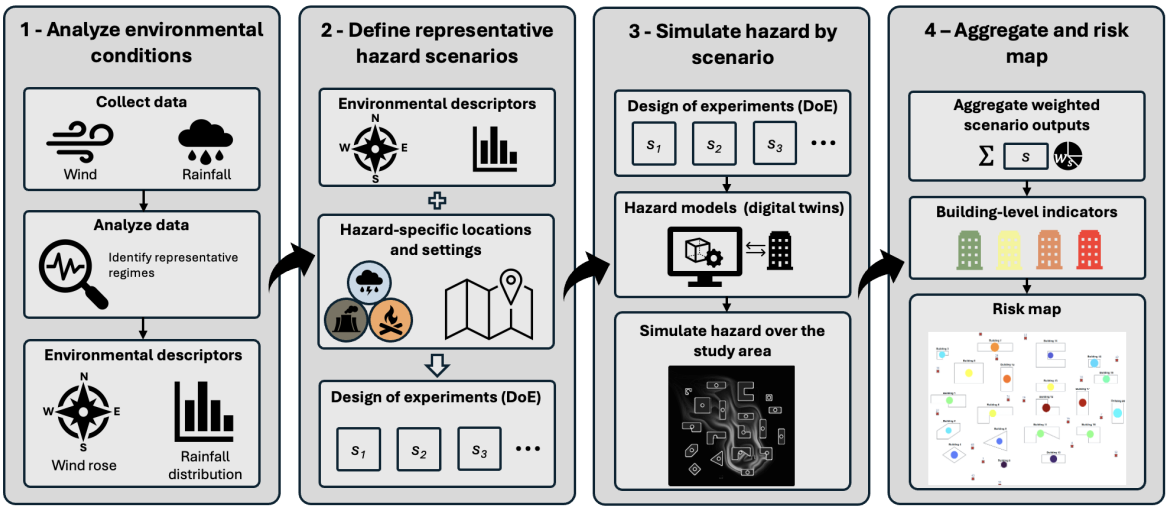


Figure 1: Four-step methodological framework from environmental data analysis to building-level risk mapping.

2.3 Probabilistic formulation and risk indicators

Let $b \in \{1, \ldots, N_b\}$ denote a building and $s \in \{1, \ldots, N_s\}$ a scenario. Each scenario is assigned a weight w_s, with:

$$\sum_{s=1}^{N_s} w_s = 1.$$

For a given building b and scenario s, the simulated continuous hazard-related output is denoted by $X_{b,s}$. Depending on the hazard, $X_{b,s}$ may represent a peak concentration, a maximum water depth aggregated over a building footprint, or more generally any continuous quantity derived from the simulation outputs.

For each building, the weighted sample $\{(X_{b,s}, w_s)\}_{s=1}^{N_s}$ defines an empirical distribution of hazard intensity over the scenario set. This distribution provides the basis for deriving risk indicators that account not only for typical values, but also for rare and severe outcomes. For continuous hazard quantities, the weighted empirical cumulative distribution function (ECDF) is defined as:

$$F_b(z) = \mathbb{P}(X_b \leq z) = \sum_{s=1}^{N_s} w_s \, \mathbf{1}\{X_{b,s} \leq z\},$$

where z denotes a generic hazard-intensity level and $\mathbf{1}\{\cdot\}$ denotes the indicator function.

The ECDF provides a cumulative description of the simulated outcomes and allows order statistics to be read directly. The weighted quantile of order $q \in (0, 1)$ is defined as:

$$Q_{q,b} = \inf\{z \mid F_b(z) \geq q\}.$$

It corresponds to the hazard-intensity level exceeded with weighted probability $1 - q$. More generally, the same ECDF can be interpreted in terms of exceedance probabilities. For a threshold z^*, the exceedance probability is:

$$EP_b(z^*) = \mathbb{P}(X_b > z^*) = 1 - F_b(z^*).$$

When z^* corresponds to an operational or regulatory threshold, $EP_b(z^*)$ directly quantifies the probability that building b exceeds this level over the considered scenario set.

Quantiles and exceedance probabilities identify the location and frequency of severe outcomes, but they do not fully describe their intensity once the threshold is exceeded. To characterise the severity of upper-tail events, we therefore also use the Tail Value at Risk (TVaR). Let:

$$\mathcal{T}_{q,b} = \{s : X_{b,s} \geq Q_{q,b}\}$$

be the set of scenarios belonging to the upper tail beyond the quantile $Q_{q,b}$. The renormalised tail weights are:

$$\bar{w}_{b,s} = \frac{w_s}{\sum_{\ell \in \mathcal{T}_{q,b}} w_\ell}, \qquad s \in \mathcal{T}_{q,b},$$

and the corresponding TVaR is:

$$\mathrm{TVaR}_{q,b} = \sum_{s \in \mathcal{T}_{q,b}} \bar{w}_{b,s} X_{b,s}.$$

Thus, while $Q_{q,b}$ identifies the threshold defining the upper tail, $\mathrm{TVaR}_{q,b}$ measures the weighted mean severity of the simulated outcomes within that tail. This makes it particularly useful for mapping low-probability but high-impact configurations.

The same exceedance logic can also be applied at the pixel scale when the hazard output is naturally represented as a raster. In this case, the generic hazard-intensity level z introduced above corresponds to a water-depth value h, and the generic threshold z^* is denoted h^*. Let

$\boldsymbol{\xi} = (x, y)$ denote a pixel location in the computational domain. Given a threshold h^* and N_s simulated rasters of maximum water depth $h_{\max,s}(\boldsymbol{\xi})$, the empirical exceedance frequency is:

$$EP_h(\boldsymbol{\xi}; h^*) = \sum_{s=1}^{N_s} w_s \, \mathbf{1}\{h_{\max,s}(\boldsymbol{\xi}) > h^*\}.$$

In the flood case study, all rainfall–soil scenarios are treated as equally weighted, so that $w_s = 1/N_s$. The corresponding exceedance map should therefore be interpreted as a relative indicator over the explored parameter space rather than as a true climatological probability.

For binary-impact hazards, such as wildfire exposure, the simulated output is not a continuous intensity but an impact indicator. We define:

$$X_{b,s,r} \in \{0, 1\},$$

where $r \in \{1, \ldots, N_r\}$ denotes an independent stochastic run, and $X_{b,s,r} = 1$ if building b is reached during run r of scenario s, and 0 otherwise. The raw scenario-wise burn probability is estimated by:

$$P_{\text{raw}}(b, s) = \frac{1}{N_r} \sum_{r=1}^{N_r} X_{b,s,r}.$$

The final weighted burn probability is:

$$P_{\text{burn}}(b) = \sum_{s=1}^{N_s} w_s P_{\text{raw}}(b, s).$$

This indicator plays, for binary fire impacts, the same role as an exceedance probability for continuous hazards: it summarises, at the building scale, the probability of being affected over the weighted ensemble of scenarios.

2.4 Case-study models and scenario construction

This section describes how the general formulation introduced above is instantiated for the three hazards considered in this work. For each case study, we specify the environmental data used to construct the scenarios, the numerical model used to simulate the hazard, and the output variable used to derive the risk indicators defined in Section 2.3.

2.4.1 Urban air-pollution dispersion

The first case study considers pollution dispersion in a synthetic urban domain. Historical wind data for Basel, Switzerland, were obtained from the *Meteoblue* [1] website over a one-year period, corresponding to 8,736 hourly observations. Each observation provides wind speed and wind direction at 10 m height.

Wind conditions are discretised into direction and speed classes. Direction classes are indexed by $i \in \{1, \ldots, 8\}$ and correspond to 45° sectors centred on N, NE, E, SE, S, SW, W, and NW. Speed classes are indexed by $j \in \{1, \ldots, 4\}$ and correspond to the intervals [0–5), [5–10), [10–20), and [20–40) km/h. Combining both dimensions yields $8 \times 4 = 32$ theoretical wind modalities, of which 29 are observed in the dataset.

Let $\mathcal{M}_{\text{wind}}$ denote the set of observed wind modalities. The empirical weight of wind modality $(i, j) \in \mathcal{M}_{\text{wind}}$ is defined as:

$$w_{\text{wind},ij} = \frac{N_{ij}}{N}, \qquad (i, j) \in \mathcal{M}_{\text{wind}},$$

[1] https://www.meteoblue.com/fr/meteo/archive/export

where N_{ij} is the number of observations in modality (i, j) and $N = 8{,}736$ is the total number of hourly observations. By construction:

$$\sum_{(i,j)\in\mathcal{M}_{\text{wind}}} w_{\text{wind},ij} = 1.$$

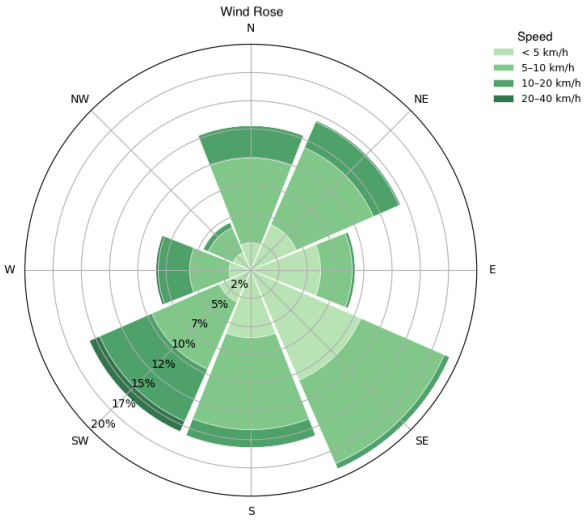


Figure 2: Empirical wind rose built from the *Meteoblue* dataset.

The dispersion model used in this case study, developed in MATLAB, relies on a two-dimensional representation of the urban layout, in which buildings are treated as impermeable obstacles within the computational domain. The airflow is assumed incompressible, which is appropriate for low-speed atmospheric wind at the considered urban scale. A refined 2D mesh is constructed over the study area, with stronger refinement near building walls and narrow passages. For each meteorological configuration, wind direction and wind intensity are imposed as far-field boundary conditions, and the incompressible Navier–Stokes equations are solved to obtain the velocity field over the whole domain.

This velocity field is then used to drive a transient convection–diffusion model for pollutant concentration. Because pollutant transport is typically advection-dominated under realistic wind conditions, the scalar transport equation is stabilised using a Streamline Upwind Petrov–Galerkin (SUPG) formulation in order to avoid non-physical oscillations and preserve a numerically robust plume shape. The simulations are repeated for several source locations and wind boundary conditions, producing time-dependent concentration fields over the entire 2D domain (Mazarakis et al., 2016; Zhuang and Wang, 2014; Ahmad et al., 2025).

Nineteen point sources of identical intensity are distributed over the domain and assumed equiprobable. Let $m \in \{1, \ldots, 19\}$ denote the source index. Each pollution scenario is defined by a triplet $s = (m, i, j)$, where m is the source index and (i, j) is the wind modality.

Direction class i is associated with an angle θ_i, and speed class j with a representative mean value v_j. The corresponding wind vector imposed in the Cartesian coordinate system of the computational domain is:

$$W_x = -v_j \sin(\theta_i), \qquad W_y = -v_j \cos(\theta_i),$$

where θ_i is expressed in radians.

Combining the 19 sources with the 29 observed wind modalities gives $19 \times 29 = 551$ scenarios.

For each scenario s, the model produces a time-dependent concentration field over the domain. Fig. 3 illustrates the type of simulated output obtained for a representative scenario. The figure

is used here to visualise the model output structure, namely the advection and spreading of the concentration plume under the imposed wind condition and the influence of urban obstacles on the plume trajectory.

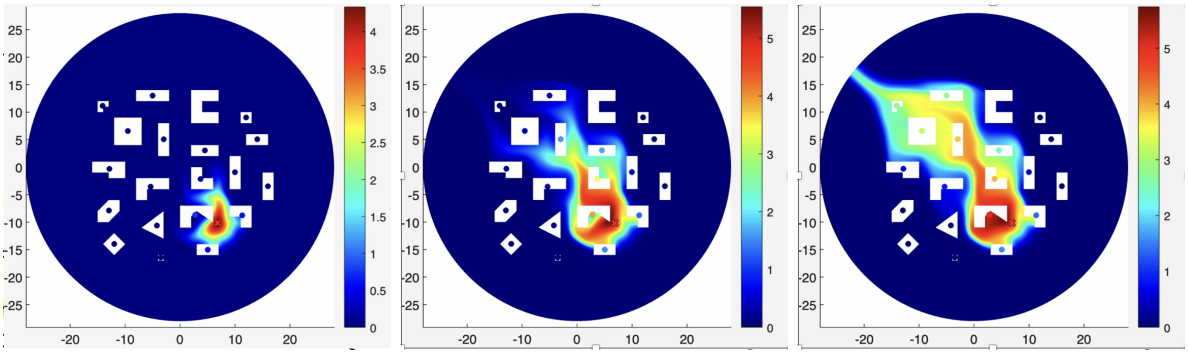


Figure 3: Illustration of the simulated time evolution of the concentration field for a representative pollution scenario.

Let b denote a building, $t \in \{1, \ldots, N_t\}$ a time step, and $\mathcal{N}_b$ the set of mesh nodes belonging to building b. In this case, $N_t = 60$. The building-scale concentration is computed as:

$$\overline{C}_{b,t,s} = \frac{1}{|\mathcal{N}_b|} \sum_{u \in \mathcal{N}_b} C_{u,t,s}.$$

The concentrations $C_{u,t,s}$ and $\overline{C}_{b,t,s}$ are expressed in relative units, with no direct correspondence to absolute physical concentration values.

In the notation of Section 2.3, the generic hazard-intensity level z corresponds here to a concentration level C. The continuous hazard-related output used for risk analysis is the peak concentration:

$$X_{b,s} = \max_{t=1,\ldots,N_t} \overline{C}_{b,t,s}.$$

Since the source locations are assumed equiprobable, the scenario weight is:

$$w_s = \frac{w_{\text{wind},ij}}{19}.$$

By construction, these scenario weights sum to one over the 551 pollution scenarios.

2.4.2 Wildfire spread at the wildland–urban interface

The second case study focuses on wildfire spread at the wildland–urban interface in the neighbourhood of Éoures, east of Marseille, France. The study area contains $N_b = 1{,}090$ buildings and exhibits a discontinuous urban fabric, nearby vegetation, rugged topography, and pronounced wildland–urban interfaces. These characteristics make the area sensitive to wind-driven fire propagation, especially under regional wind regimes such as the Mistral and easterlies. The objective is to estimate, for each building, the probability of being reached by a wildfire front.

Meteorological data were extracted from the ERA5-Land product provided by ECMWF, accessed via the *Open-Meteo Historical Weather API*[2], for the geographical point of Éoures. The analysis period spans from 1 January 2015 to 30 September 2025, corresponding to approximately 94,000 hourly records. Each observation provides wind speed and wind direction at 10 m height using the `FROM` convention.

Wind conditions are discretised into direction and speed classes. Direction classes are indexed by $i \in \{1, \ldots, 8\}$ and correspond to 45° sectors centred on N, NE, E, SE, S, SW, W, and NW.

[2] https://open-meteo.com/en/docs/historical-weather-api

Speed classes are indexed by $j \in \{1, 2, 3\}$ and correspond to the intervals [0–14), [14–29), and [29–47) km/h. Combining both dimensions yields $8 \times 3 = 24$ wind modalities.

For each direction–speed modality (i, j), let N_{ij} denote the number of observations falling in that modality. The empirical wind weight is defined as:

$$w_{\text{wind},ij} = \frac{N_{ij}}{N}, \qquad \sum_{i=1}^{8} \sum_{j=1}^{3} w_{\text{wind},ij} = 1,$$

where N is the total number of hourly observations. In addition to these weights, the observations within each modality are used to define representative wind inputs. Let $\mathcal{I}_{ij}$ denote the set of observations belonging to modality (i, j). The directional mean and the empirical mean speed are computed as:

$$\bar{\theta}_{ij} = \arg\left(\frac{1}{N_{ij}} \sum_{o \in \mathcal{I}_{ij}} e^{\,\mathrm{i}\theta_o} \right), \qquad \bar{v}_{ij} = \frac{1}{N_{ij}} \sum_{o \in \mathcal{I}_{ij}} v_o.$$

The pair $(\bar{\theta}_{ij}, \bar{v}_{ij})$ is used as the representative wind condition for modality (i, j).

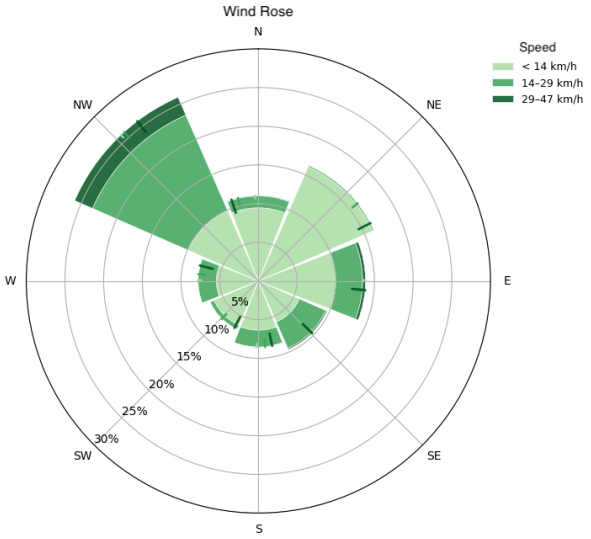
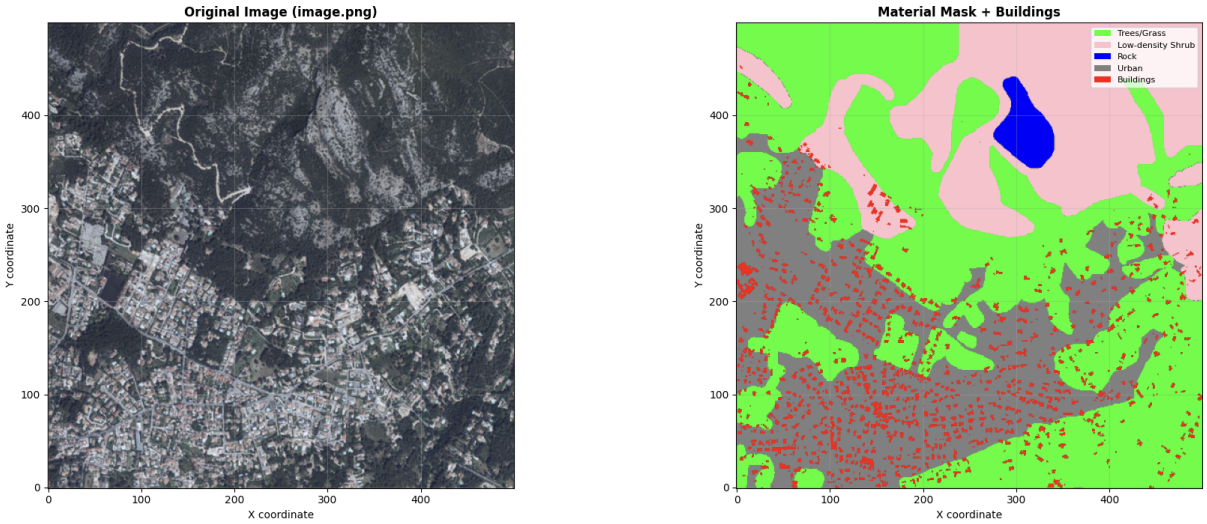


Figure 4: Empirical wind rose built from the *Open-Meteo Historical Weather API*.

Wildfire propagation is simulated using `FIRE-2D`, a stochastic two-dimensional cellular automaton model following the general logic of established cellular-automata approaches for wildfire spread (Freire and DaCamara, 2019). The raw aerial image of the study area is converted into a discretised material grid distinguishing trees/grass, low-density shrub, rock, and urban areas. A separate building mask is superimposed on the same grid in order to identify the buildings reached by the fire front.



(a) Raw aerial image

(b) Material grid and building mask

Figure 5: Conversion of the real-world image into the discrete grid used by the wildfire model.

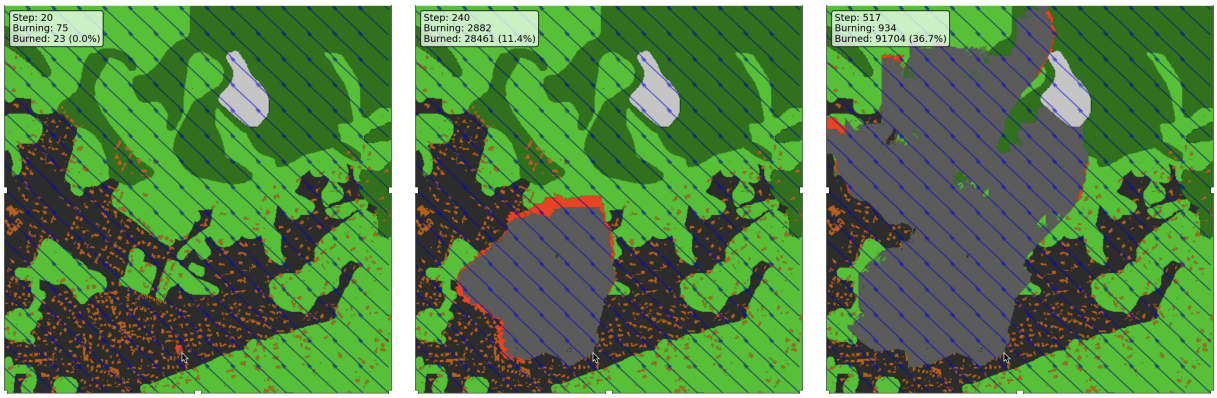
Each material class is assigned a base flammability coefficient and a burn duration, as reported in Table 1. Trees/grass and low-density shrub are the most combustible classes, rock is treated as non-flammable, and urban cells are assigned an intermediate behaviour. Buildings are tracked separately through the building mask and are considered exposed as soon as at least one of their pixels is reached by the fire.

Material class	Flammability	Burn time
Trees / grass	0.10	15
Low-density shrub	0.10	5
Rock	0.00	0
Urban	0.05	10

Table 1: Material-dependent parameters used in the `FIRE-2D` model.

Model setup and physical drivers. In this study, each simulation corresponds to a scenario defined by: (i) a wind modality derived from the wind rose, combining direction and speed, and (ii) an ignition point located within the domain. The model incorporates several physical components that are relevant to wildfire spread: material-dependent combustibility, topography and its directional influence, the built environment through the building mask, and wind conditions, which are assumed to remain constant over the duration of a simulation.

Stochastic spread mechanism. The spread process is stochastic and is governed by local transition rules on the cellular grid. Each cell can be in one of three states: unburned, burning, or burned. At each time step, every burning cell attempts to ignite each of its eight neighbouring cells that are still unburned. For a given burning–unburned pair, a random draw is performed, and ignition occurs if the draw is lower than a probability proportional to the base flammability of the target cell multiplied by a directional influence factor. In the present implementation, this influence factor accounts for the local slope, wind direction, wind intensity, and the relative position of the neighbour, including the larger distance of diagonal neighbours. Once ignited, a cell remains in the burning state for a prescribed number of iterations determined by its burn duration, after which it becomes burned. As a result, two simulations performed under identical input conditions may yield different spread trajectories and different sets of impacted buildings.



(a) $t = 20$ (b) $t = 240$ (c) $t = 517$

Figure 6: Example of the spatio-temporal evolution of a `FIRE-2D` simulation under a fixed scenario.

After each simulation, each building is assigned a binary outcome indicating whether at least one of its pixels was reached by the fire. These elementary outcomes define the variables $X_{b,s,r}$ introduced in Section 2.3, and are subsequently aggregated over repeated stochastic runs and weighted scenarios to estimate the final building-level burn probabilities.

The design of experiments combines wind modalities and ignition locations. Let $n \in \{1, \ldots, 20\}$ denote the ignition-point index. The 20 ignition points are randomly selected within the study area so as to cover different environmental settings, including vegetated areas, wildland–urban interfaces, and locations close to the built environment. Each wildfire scenario is defined by $s = (n, i, j)$, where n is the ignition point and (i, j) is the wind modality. Combining the 20 ignition points with the 24 wind modalities yields $20 \times 24 = 480$ scenarios.

Because the model is stochastic, several independent runs are required for each scenario. A preliminary convergence analysis was conducted on representative scenarios by running 1,250 simulations and assessing the stability of cumulative burn-probability estimates under random reorderings of the runs. Based on these diagnostics, $N_r = 500$ runs per scenario were retained as a practical compromise between statistical stability and computational cost. Detailed convergence diagnostics are reported in Appendix A. The full wildfire design therefore contains $480 \times 500 = 240{,}000$ independent simulations.

For each building b, scenario s, and run $r \in \{1, \ldots, N_r\}$, the binary variable $X_{b,s,r}$ is equal to 1 if building b is reached by the fire during run r, and 0 otherwise, as introduced in Section 2.3. Since ignition locations are assumed equiprobable within each wind modality, the weight of scenario $s = (n, i, j)$ is:

$$w_s = \frac{w_{\text{wind},ij}}{20}.$$

The final building-level burn probability $P_{\text{burn}}(b)$ is then obtained using the weighted aggregation defined in Section 2.3.

In addition to the final burn probability, we compute scenario-contribution diagnostics in order to identify the wind modalities and ignition points that contribute most to the overall risk. For a given scenario s, the scenario risk score is defined as:

$$\text{Risk}(s) = \sum_{b=1}^{N_b} w_s P_{\text{raw}}(b, s).$$

Let $\mathcal{S}_{ij}^{\text{wind}}$ denote the set of scenarios associated with wind modality (i, j), and let $\mathcal{S}_n^{\text{ign}}$ denote the

set of scenarios associated with ignition point n. The normalised wind and ignition contributions are:

$$\mathrm{Contrib}_{\mathrm{wind}}(i,j) = \frac{\sum_{s\in\mathcal{S}_{ij}^{\mathrm{wind}}} \mathrm{Risk}(s)}{\sum_{s=1}^{480} \mathrm{Risk}(s)},$$

and

$$\mathrm{Contrib}_{\mathrm{ign}}(n) = \frac{\sum_{s\in\mathcal{S}_{n}^{\mathrm{ign}}} \mathrm{Risk}(s)}{\sum_{s=1}^{480} \mathrm{Risk}(s)}.$$

These quantities are used in the Results section to relate the final burn-probability map to the dominant wind regimes and ignition locations.

2.4.3 Urban surface flooding

The third case study focuses on surface flooding in an urban area around the neighbourhood of Éoures, near Marseille. Surface runoff is simulated using the open-source *Landlab* framework (Hobley et al., 2017; Barnhart et al., 2020). The objective is to construct a flood-depth exceedance indicator from an ensemble of rainfall–soil scenarios. The hazard variable considered in this case is the maximum water depth reached during a simulation.

Precipitation data were obtained from the public Météo-France dataset "Données climatologiques de base horaires (BASE-HOR)" for the Bouches-du-Rhône department over the 2020–2024 period, downloaded from the *data.gouv.fr* platform[3]. The variable `RR1_mm`, corresponding to one-hour accumulated precipitation, was extracted to characterise the empirical distribution of hourly rainfall intensities and guide the selection of rainfall levels explored in the design of experiments.

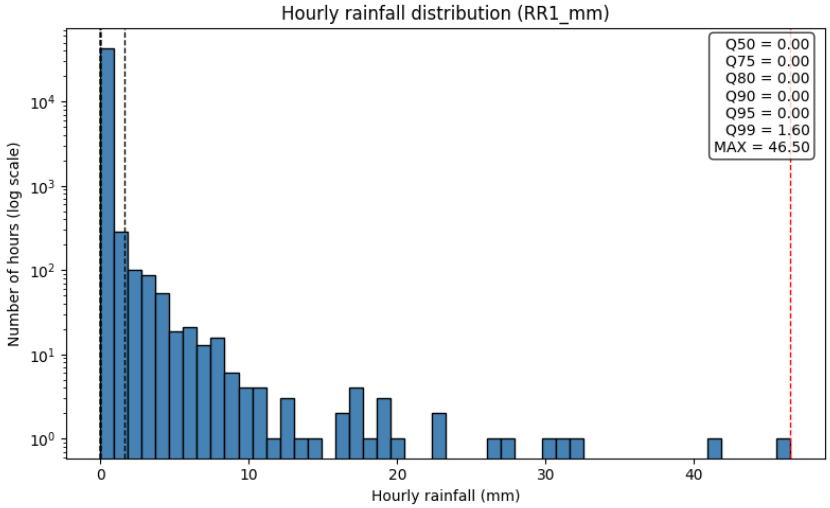


Figure 7: Histogram of hourly rainfall built from the *data.gouv.fr* dataset.

The computational domain is defined from a raster digital elevation model (DEM) describing ground elevation over the study area. The DEM is discretised on a regular grid with resolution $(\Delta x, \Delta y) \approx (6.606, 6.604)$ m in EPSG:3857 projection. Let $\boldsymbol{\xi} = (x, y)$ denote a pixel location in the computational domain. The same raster grid is used to represent topography and hydraulic outputs, ensuring spatial consistency between the input elevation field and the simulated water-depth fields.

Surface runoff is simulated using the *OverlandFlow* component of *Landlab*, which relies on a simplified inertial approximation of the shallow-water equations. Rainfall is prescribed as a spatially uniform intensity and applied for 2 hours. Infiltration losses are represented using a

[3] https://www.data.gouv.fr/datasets/donnees-climatologiques-de-base-horaires

Green–Ampt-type scheme. Domain boundaries are treated as open, allowing water to leave the computational domain, and the time step is automatically adjusted to satisfy the numerical stability constraints of the flow scheme.

The model computes the water depth $h(\boldsymbol{\xi}, t)$ over the raster grid. The output retained for the risk analysis is the maximum water depth:

$$h_{\max}(\boldsymbol{\xi}) = \max_t h(\boldsymbol{\xi}, t).$$

In the notation of Section 2.3, the generic hazard-intensity level z corresponds here to a water-depth value h, and the generic threshold z^* is denoted h^*.

A design of experiments is constructed by combining three parameters: rainfall intensity I, saturated hydraulic conductivity K, and initial soil moisture θ_0. The tested levels are:

$$I \in \{1, 3, 5, 10, 20, 40\}\ \mathrm{mm\,h^{-1}}, \quad K \in \{10^{-7}, 10^{-6}, 10^{-5}\}\ \mathrm{m\,s^{-1}}, \quad \theta_0 \in \{0.1, 0.2, 0.3\}.$$

Each flood scenario is therefore defined by a triplet $s = (I, K, \theta_0)$. The full design contains $6 \times 3 \times 3 = 54$ simulations. In this proof-of-concept version, these rainfall–soil scenarios are treated as equally weighted, so that:

$$w_s = \frac{1}{54}.$$

Each simulation produces one raster field of maximum water depth, denoted $h_{\max,s}(\boldsymbol{\xi})$. Fig. 8 illustrates the type of model output obtained for one rainfall–soil scenario. This raster is not used as a final risk map by itself, but as one element of the ensemble from which threshold-exceedance indicators are computed.

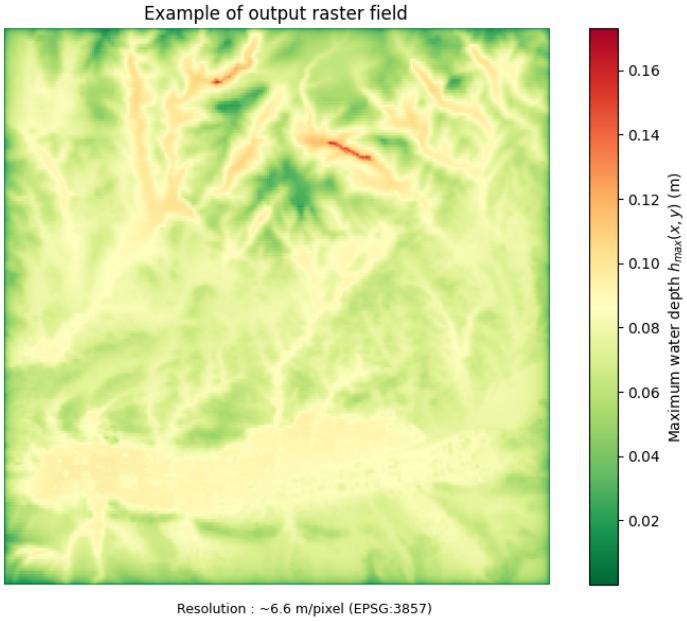


Figure 8: Example model output: raster field of maximum water depth $h_{\max,s}(\boldsymbol{\xi})$, for the combination $(I = 40\,\mathrm{mm\,h^{-1}}, K = 10^{-7}\,\mathrm{m\,s^{-1}}, \theta_0 = 0.3)$.

At each pixel location $\boldsymbol{\xi}$, the ensemble of 54 simulations therefore defines an empirical sample $\{h_{\max,s}(\boldsymbol{\xi})\}_{s=1}^{54}$. From this sample, threshold-exceedance indicators are constructed following the formulation introduced in Section 2.3. For a water-depth threshold h^*, the empirical exceedance frequency is:

$$EP_h(\boldsymbol{\xi}; h^*) = \frac{1}{54} \sum_{s=1}^{54} \mathbf{1}\{h_{\max,s}(\boldsymbol{\xi}) > h^*\}.$$

This quantity represents the fraction of rainfall–soil scenarios for which the maximum water depth exceeds h^* at pixel $\boldsymbol{\xi}$. Since all scenarios are equally weighted in this proof-of-concept implementation, it should be interpreted as a relative indicator over the explored parameter space rather than as a true climatological exceedance probability.

The hydraulic model naturally produces raster-based exceedance fields. These fields are displayed together with building footprints, providing a spatial support for assessing flood exposure at the building scale. In practice, the exposure of an individual building can be inferred from the exceedance values intersecting its footprint, or summarised into a single building-level indicator using zonal statistics such as the maximum or mean exceedance value over the footprint.

3 Results

The following results illustrate how the methodological framework translates heterogeneous simulation outputs into interpretable spatial indicators for each hazard.

3.1 Urban air-pollution dispersion

The 551 weighted pollution scenarios produce, for each building, an empirical distribution of peak concentration values. These distributions make it possible to analyse not only average exposure levels, but also the upper tail of the simulated outcomes, which corresponds to rare but severe pollution episodes.

Fig. 9 illustrates this behaviour for building 17, which is used here as an illustrative example of the weighted empirical distribution obtained at the building scale. The weighted histogram highlights the distribution of peak concentration values across all source–wind scenarios, whereas the weighted ECDF provides a cumulative representation from which quantiles, exceedance probabilities, and tail-oriented indicators can be derived. The distribution is strongly skewed, with most of the probability mass concentrated at low concentration levels and a long upper tail corresponding to less frequent but more severe exposure configurations.

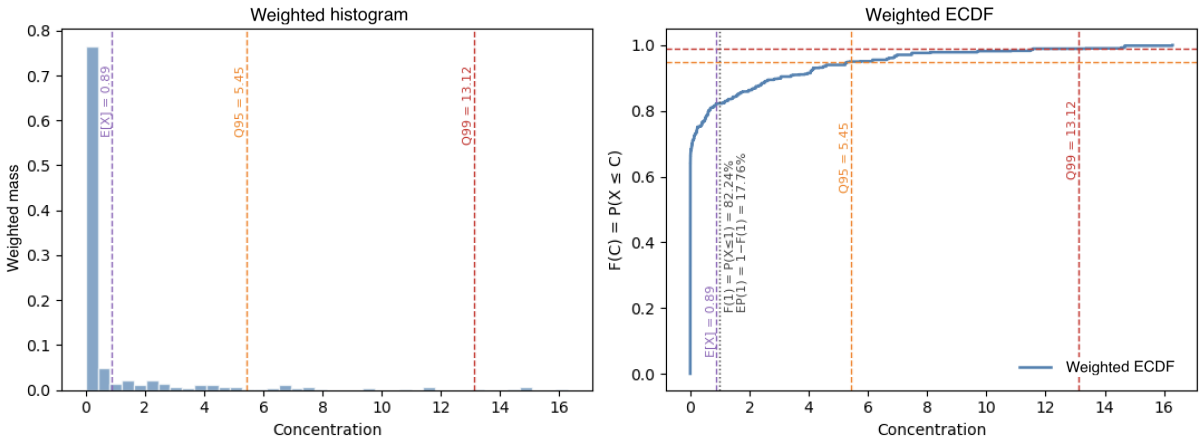


Figure 9: Weighted empirical distribution of peak concentrations for building 17. The left panel shows the weighted histogram, while the right panel shows the corresponding weighted ECDF.

The building-level risk indicators computed from the weighted empirical distributions are shown in Fig. 10. In this map, the circle size represents the empirical maximum peak concentration reached by each building across the 551 scenarios, while the colour represents the mean severity of the worst 5% of cases, measured by $\mathrm{TVaR}_{0.95,b}$ and normalised across buildings.

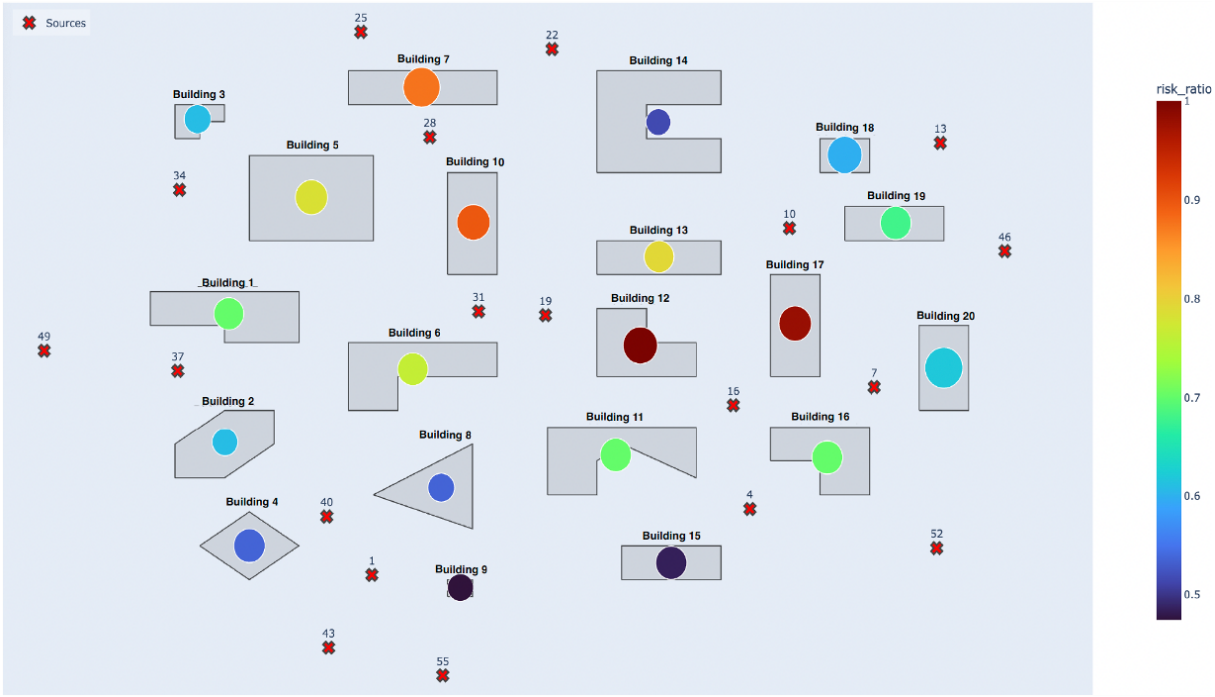


Figure 10: Empirical building-level air-pollution risk map. Circle size represents the empirical maximum peak concentration, while colour represents the normalised $\mathrm{TVaR}_{0.95,b}$.

The resulting map reveals marked spatial contrasts between buildings. The highest tail severities, as measured by $\mathrm{TVaR}_{0.95,b}$, are observed for buildings 12, 17, 10, and 7. By contrast, the largest empirical maximum concentrations are mainly concentrated in buildings 20, 5, 12, and 10. This difference between buildings associated with high tail-average severity and those associated with the largest isolated peak values highlights the interest of using complementary risk indicators. While the maximum identifies the most extreme simulated exposure, $\mathrm{TVaR}_{0.95,b}$ provides a more stable measure of the severity of rare high-impact configurations.

Overall, the pollution case study shows that combining deterministic dispersion simulations with wind-rose-based scenario weighting produces interpretable building-scale indicators. Order statistics derived from the weighted empirical distributions provide consistent indicators to describe, beyond average exposure, both the frequency and the severity of extreme episodes. The resulting map is consistent with the influence of prevailing wind directions and urban geometry, and provides a transparent basis for ranking building exposure to unfavourable source–wind configurations.

3.2 Wildfire spread at the wildland–urban interface

The wildfire simulations produce, for each building, a final burn probability $P_{\mathrm{burn}}(b)$ obtained by aggregating the stochastic runs and the weighted wind–ignition scenarios. This indicator represents the probability that building b is reached by the fire front over the scenario ensemble.

Beyond the final building-level probability, the contribution analysis makes it possible to identify which wind modalities and ignition points drive most of the overall risk. Fig. 11 ranks the wind modalities and ignition points by decreasing contribution to the total weighted burn probability. The results show that a limited number of wind–ignition configurations accounts for a large share of the overall risk, highlighting the strongly directional and scenario-dependent nature of wildfire exposure in the study area.

Wind

Rank	Speed km/h	Direction	Contrib_Vent	Contrib_Cumul
1	5.4	NE -> SO	24.0%	24.0%
2	6.8	E -> O	12.5%	36.4%
3	8.6	NO -> SE	9.8%	46.3%
4	7.9	O -> E	9.6%	55.8%
5	20.2	NO -> SE	8.4%	64.2%
⋮				
23	32.4	NE -> SO	0.0%	100.0%
24	30.6	SO -> NE	0.0%	100.0%

Ignition points

Rank	Start_X	Start_Y	Contrib_Igntion	Contrib_cumul
1	436	165	20.7%	20.7%
2	106	380	20.5%	41.2%
3	314	172	16.7%	57.9%
4	143	292	16.5%	74.4%
5	228	51	14.6%	88.9%
⋮				
19	186	456	0.0%	100.0%
20	46	72	0.0%	100.0%

Figure 11: Wind modalities and ignition points ranked by decreasing contribution to overall wildfire risk.

The final wildfire risk map is shown in Fig. 12. The colour scale represents the building-level burn probability $P_{\text{burn}}(b)$ for the 1,090 buildings in the study area. The five ignition points with the largest contributions to overall risk are indicated by star-shaped symbols. An inset wind rose summarises the relative contributions of the dominant wind modalities.

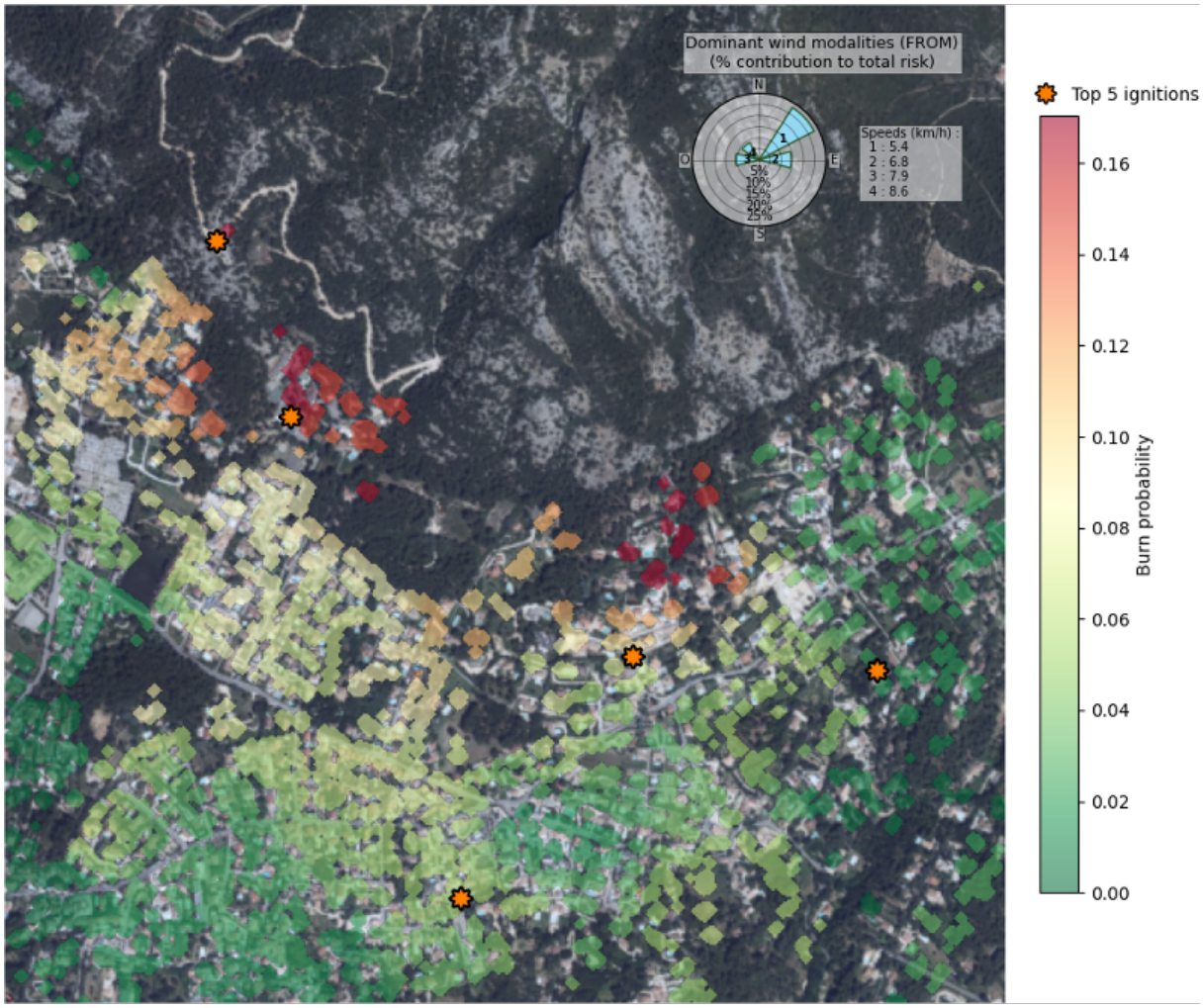


Figure 12: Final wildfire risk map based on the building-level burn probability $P_{\text{burn}}(b)$.

The map highlights spatial contrasts in building-level burn probability across the study area. The star-shaped symbols identify the ignition points with the largest contributions to overall risk, while the inset wind rose relates the mapped probabilities to the dominant wind modalities. Together, these elements make it possible to interpret the final burn-probability map in terms of both ignition location and wind-regime contributions.

Overall, the wildfire case study illustrates how a stochastic spread model can be transformed

into an interpretable building-scale risk map. The final indicator $P_{\text{burn}}(b)$ summarises both model stochasticity and scenario variability, while the contribution analysis provides additional information on the wind regimes and ignition locations that dominate the overall risk.

3.3 Urban surface flooding

The 54 rainfall–soil scenarios produce an ensemble of maximum water-depth rasters $h_{\max,s}(\boldsymbol{\xi})$. To illustrate how the design of experiments translates into local variability of the hydraulic response, three representative pixels are selected within the domain: one associated with low accumulation, one with intermediate accumulation, and one corresponding to the global maximum. For each of these pixels, Fig. 13 shows the distribution of the 54 simulated values of $h_{\max}$.

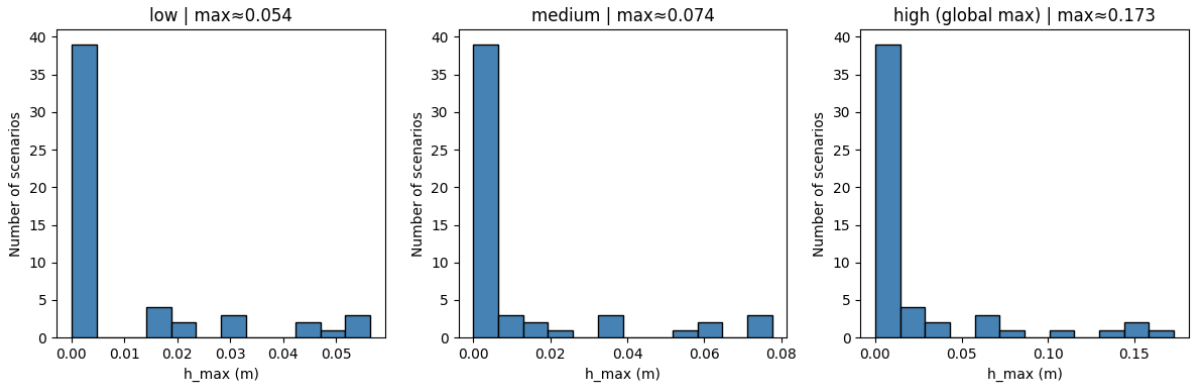


Figure 13: Distributions of $h_{\max}$ values across the 54 rainfall–soil simulations for three representative pixels: low, intermediate, and high accumulation.

The distributions show that the response to the rainfall–soil design of experiments varies strongly across the domain. Pixels located in low-accumulation areas remain associated with very small maximum water depths across most scenarios, whereas pixels located in accumulation zones exhibit a much wider range of responses. This local variability confirms the interest of analysing the full ensemble of simulated rasters rather than relying on a single rainfall–soil configuration.

The resulting exceedance map is shown in Fig. 14 for a threshold $h^* = 5$ cm. The mapped value corresponds to the empirical frequency with which the maximum water depth exceeds this threshold over the 54 rainfall–soil scenarios.

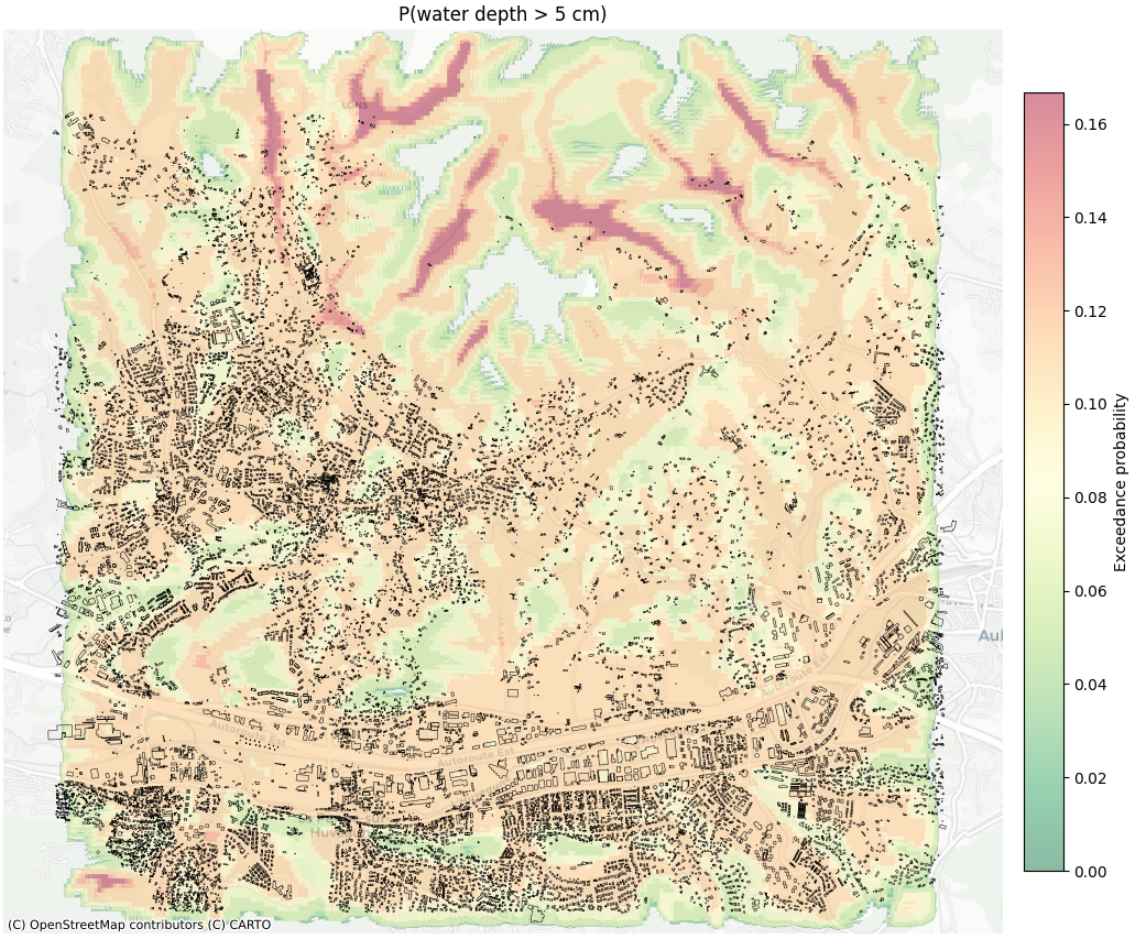


Figure 14: Final flood-risk map based on the empirical exceedance frequency of the threshold $h^* = 5$ cm.

The map highlights the areas where water-depth exceedance is most frequent across the explored parameter space. Since all rainfall–soil scenarios are equally weighted in this proof-of-concept implementation, the mapped indicator should be interpreted as a relative exceedance indicator rather than as a true climatological probability. Nevertheless, it provides a spatially explicit representation of flood-prone zones and identifies locations where surface-water accumulation is repeatedly generated across multiple rainfall and soil configurations.

The final exceedance field is displayed together with building footprints, making it possible to assess the exposure of individual buildings from the exceedance values intersecting their footprint. This supports a building-scale interpretation of the flood indicator while preserving the raster nature of the hydraulic model output.

4 Discussion

Taken together, the three case studies show that heterogeneous environmental hazard models can be organised within a common Digital-Twin-based risk-assessment framework. Although the physical processes, numerical models, and output variables differ substantially between pollution dispersion, wildfire spread, and surface flooding, the same general pipeline can be applied: environmental drivers are discretised into scenarios, hazard models generate spatial outputs, these outputs are transformed into empirical risk indicators, and the resulting indicators are mapped in relation to buildings or building footprints.

A central contribution of this work is therefore methodological. The objective is not to propose a single universal risk metric valid for all hazards, but rather to define a common simulation-to-risk-mapping structure within which hazard-specific indicators can be derived consistently. In

the pollution case, weighted empirical distributions of peak concentrations make it possible to compute order statistics and tail-oriented indicators such as $\mathrm{TVaR}_{0.95,b}$. In the wildfire case, repeated stochastic simulations are aggregated into building-level burn probabilities. In the flooding case, an ensemble of rainfall–soil simulations is transformed into threshold-exceedance fields that can be interpreted in relation to building footprints. These examples illustrate that continuous fields, stochastic binary impacts, and raster-based exceedance indicators can be handled within the same general framework.

The proposed approach also highlights the value of synthetic yet physically grounded scenarios. Historical records of hazard impacts are often sparse, incomplete, or not representative of rare and future conditions. By using physical or stochastic simulators driven by environmental scenarios, the framework makes it possible to explore configurations that may not be present in the historical record but remain relevant for risk assessment. This is particularly important for low-probability high-impact events, for which empirical evidence is limited but consequences may be severe. The resulting maps provide interpretable spatial indicators that can support preliminary diagnosis, comparison between buildings, and prioritisation of areas requiring further analysis.

From a decision-support perspective, the framework provides a first level of information for urban planning and risk management. The maps do not replace detailed operational studies, but they offer a transparent way to identify spatial contrasts in exposure and to relate them to environmental drivers such as wind regimes, source or ignition locations, rainfall intensities, and soil conditions. In this sense, the framework can be seen as an intermediate layer between raw hazard simulation outputs and future decision-support tools incorporating vulnerability, exposure, economic losses, or adaptation costs.

Several limitations must nevertheless be emphasised. First, the present work is a proof of concept and does not include calibration against observed hazard events or impact data. The pollution concentrations are expressed in relative units, which prevents direct comparison with regulatory concentration thresholds. The wildfire model relies on simplified material classes, fixed wind conditions during each simulation, and randomly selected ignition points. The flood case uses equally weighted rainfall–soil scenarios, so the resulting exceedance indicators should be interpreted as relative indicators over the explored parameter space rather than as true climatological probabilities. More generally, the scenario weights used in this work are simplified approximations of environmental occurrence frequencies and should be refined for operational applications.

Second, the framework currently focuses on hazard intensity and exposure, but does not yet include explicit vulnerability or damage models. Buildings are treated primarily as spatial assets for which hazard indicators are computed. To move towards full risk assessment, future developments should integrate building vulnerability, occupancy, asset values, protection measures, and potential consequences. This would make it possible to derive indicators such as expected damage, repair time, disruption, or economic loss, and would strengthen the connection with insurance-oriented and policy-oriented applications.

Third, the hazards are treated independently. In real urban systems, environmental hazards may interact or occur in cascading sequences. For example, drought conditions may increase wildfire susceptibility, wildfire smoke may generate pollution episodes, and flooding may disrupt infrastructure or trigger secondary failures. Extending the proposed framework towards coupled or cascading hazards would require linking several Digital Twins or developing multi-physics and hybrid modelling strategies. Such extensions represent an important direction for future work.

Finally, the current framework is designed for offline scenario exploration. Its operational relevance could be increased by incorporating real-time data streams, sensor measurements, remote-sensing products, or updated meteorological forecasts. Data assimilation, Bayesian updating, surrogate modelling, and machine-learning-based acceleration could support more dynamic probabilistic Digital Twins. These developments would allow the framework to move from

static scenario-based mapping towards adaptive risk forecasting.

Overall, this work demonstrates that Digital-Twin-based scenario modelling can provide a coherent methodology for constructing multi-hazard risk maps at the building scale. The three case studies show that diverse hazard simulators can be connected to a common probabilistic post-processing pipeline, producing interpretable indicators that support spatial risk diagnosis and provide a basis for future decision-support applications.

Acknowledgements

DesCartes: this research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.
BM acknowledges the support of the ANR under the Junior Professor Chair contract ANR-24-CPJ1-0005-01.

During the preparation of this manuscript, the authors used ChatGPT, developed by OpenAI, to assist with language editing, grammar correction, and improvement of clarity and structure in selected parts of the text. The tool was not used to generate scientific results, perform data analysis, or draw conclusions. After using this tool, the authors carefully reviewed and edited the manuscript and take full responsibility for the final content of the publication.

Declaration of interest statement

No potential conflict of interest was reported by the authors.

References

Abdo, H., J.-M. Flaus, and F. Masse (2017). "Uncertainty quantification in risk assessment - Representation, propagation and treatment approaches: Application to atmospheric dispersion modeling". In: *Journal of Loss Prevention in the Process Industries*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/85020681847?origin=scopusAI.

Ahmad, F., D. Majumder, R. Ranjit, and M. Manhart (2025). "Preliminary study on the spread of air-borne pollutants in urban environment: a CFD simulation approach". In: *Scientific Reports*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/105006928598?origin=scopusAI.

AIAA Digital Engineering Integration Committee, S et al. (2020). "Digital Twin: Definition & Value—An AIAA and AIA Position Paper". In: *AIAA: Reston, VA, USA*.

Balter, B. M. and M. V. Faminskaya (2017). "Irregularly Emitting Air Pollution Sources: Acute Health Risk Assessment Using AERMOD and the Monte Carlo Approach to Emission Rate". In: *Air Quality, Atmosphere & Health* 10.4, pp. 401–409. DOI: 10.1007/s11869-016-0428-x.

Baraldi, Piero and Enrico Zio (2008). "A combined Monte Carlo and possibilistic approach to uncertainty propagation in event tree analysis". In: *Risk Analysis: An International Journal* 28.5, pp. 1309–1326.

Barbie, Alexander and Wilhelm Hasselbring (2024). "From digital twins to digital twin prototypes: Concepts, formalization, and applications". In: *IEEE access* 12, pp. 75337–75365.

Barnhart, K. R., E. W. H. Hutton, G. E. Tucker, N. M. Gasparini, E. Istanbulluoglu, D. E. J. Hobley, N. J. Lyons, M. Mouchene, S. S. Nudurupati, J. M. Adams, and C. Bandaragoda (2020). "Short communication: Landlab v2.0: A software package for Earth surface dynamics". In: *Earth Surface Dynamics* 8.2, pp. 379–397. DOI: 10.5194/esurf-8-379-2020.

Biswas, T. K. and K. Zaman (2019). "A Fuzzy-Based Risk Assessment Methodology for Construction Projects Under Epistemic Uncertainty". In: *International Journal of Fuzzy Systems*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/85065850943?origin=scopusAI.

Brito, M. P., M. Stevenson, and C. Bravo (2023). "Subjective machines: Probabilistic risk assessment based on deep learning of soft information". In: *Risk Analysis*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/85128572134?origin=scopusAI.

Carmel, Yohay, Shlomit Paz, Faris Jahshan, and Maxim Shoshany (2009). "Assessing Fire Risk Using Monte Carlo Simulations of Fire Spread". In: *Forest Ecology and Management* 257.1, pp. 370–377. DOI: 10.1016/j.foreco.2008.09.039.

Chinesta, Francisco, Elias Cueto, Emmanuelle Abisset-Chavanne, Jean Louis Duval, and Fouad El Khaldi (2020). "Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data: F. Chinesta et al." In: *Archives of computational methods in engineering* 27.1, pp. 105–134.

De Haan, M. and J. Rigby (2020). "Scenario-based risk analysis in the absence of historical data". In: *Risk Analysis*.

Falter, Daniela, Kai Schröter, Nguyen Viet Dung, Sergiy Vorogushyn, Heidi Kreibich, Yeshewatesfa Hundecha, Heiko Apel, and Bruno Merz (2015). "Spatially Coherent Flood Risk Assessment Based on Long-Term Continuous Simulation with a Coupled Model Chain". In: *Journal of Hydrology* 524, pp. 182–193. DOI: 10.1016/j.jhydrol.2015.02.021.

Freire, Joana Gouveia and Carlos C. DaCamara (2019). "Using cellular automata to simulate wildfire propagation and to assist in fire management". In: *Natural Hazards and Earth System Sciences* 19, pp. 169–179. DOI: 10.5194/nhess-19-169-2019.

Ghanem, Roger, Christian Soize, Loujaine Mehrez, and Venkat Aitharaju (2022). "Probabilistic learning and updating of a digital twin for composite material systems". In: *International Journal for Numerical Methods in Engineering* 123.13, pp. 3004–3020.

Hafver, A, S Eldevik, and FB Pedersen (2018). "Probabilistic digital twins". In: *Position paper, DNV GL*.

Hassan, Mustafa Hamid, Salama A. Mostafa, Aida Mustapha, Mohd Zainuri Saringat, Bander Ali Saleh Al-rimy, Faisal Saeed, A. E. M. Eljialy, and Mohammed Ahmed Jubair (2022). "A New Collaborative Multi-Agent Monte Carlo Simulation Model for Spatial Correlation of Air Pollution Global Risk Assessment". In: *Sustainability* 14.1, p. 510. DOI: 10.3390/su14010510.

Hobley, D. E. J., J. M. Adams, S. S. Nudurupati, E. W. H. Hutton, N. M. Gasparini, E. Istanbulluoglu, and G. E. Tucker (2017). "Creative computing with Landlab: an open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics". In: *Earth Surface Dynamics* 5.1, pp. 21–46. DOI: 10.5194/esurf-5-21-2017.

Kapteyn, Michael G, Jacob VR Pretorius, and Karen E Willcox (2021). "A probabilistic graphical model foundation for enabling predictive digital twins at scale". In: *Nature Computational Science* 1.5, pp. 337–347.

Li, Cailin, Na Sun, Yihui Lu, Baoyun Guo, Yue Wang, Xiaokai Sun, and Yukai Yao (2023). "Review on Urban Flood Risk Assessment". In: *Sustainability* 15.1, p. 765. DOI: 10.3390/su15010765.

Li, L., C.-C. Duan, Y.-W. Gong, and H.-G. Cheng (2025). "Research progress on urban ecological environment risk assessment methods and models". In: *Zhongguo Huanjing Kexue / China Environmental Science*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/105017112627?origin=scopusAI.

Liang, Huangbin, Beatriz Moya, Francisco Chinesta, and Eleni Chatzi (2025). "A multi-model probabilistic framework for seismic risk assessment and retrofit planning of electric power networks". In: *Reliability Engineering & System Safety*, p. 112001.

Liang, Huangbin, Beatriz Moya, Eugene Seah, Ashley Ng Kwok Weng, Dominique Baillargeat, Jonas Joerin, Xiaozheng Zhang, Francisco Chinesta, and Eleni Chatzi (2025). "Harnessing hybrid digital twinning for decision-support in smart infrastructures". In: *Data-Centric Engineering* 6, e43.

Mazarakis, N., E. Kaloudis, A. Nazos, and K.-S. P. Nikas (2016). "LES and RANS comparison of flow and pollutant dispersion in urban environment". In: *International Journal of Environmental Studies*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/84955179486?origin=scopusAI.

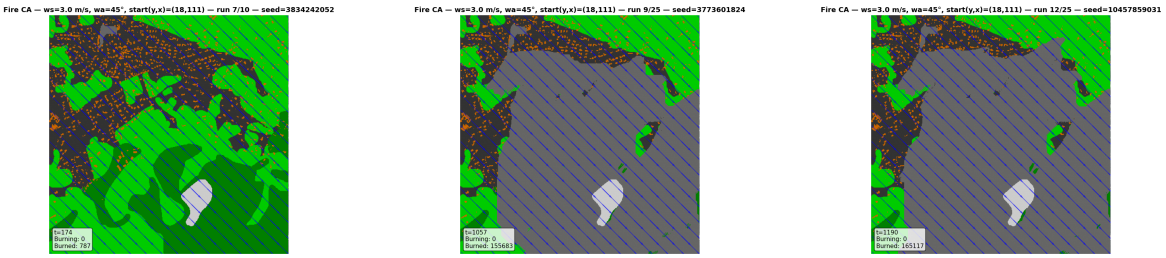
Pylianidis, Christos, Val Snow, Hiske Overweg, Sjoukje Osinga, John Kean, and Ioannis N Athanasiadis (2022). "Simulation-assisted machine learning for operational digital twins". In: *Environmental Modelling & Software* 148, p. 105274.

Reis, S., C. Sabel, and T. Oxley (2009). "Implications of different spatial (and temporal) resolutions for integrated assessment modelling on the regional to local scale — nesting, coupling, or model integration?" In: *18th World IMACS Congress and MODSIM 2009 - International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/85086221192?origin=scopusAI.

Zhuang, X. and X. Wang (2014). "Environment analysis near a highway using computational fluid dynamics". In: *ASME International Mechanical Engineering Congress and Exposition (IMECE)*. URL: https://www-scopus-com.rp1.ensam.eu/pages/publications/84926355750?origin=scopusAI.

A Convergence analysis for the wildfire simulations

Because `FIRE-2D` is stochastic, two simulations performed under identical input conditions may produce markedly different outcomes. Fig. A1 illustrates this intrinsic variability for a fixed scenario: depending on the internal random draw, fire spread may remain limited or, conversely, reach a large portion of the built environment.



(a) Run 1: 22 buildings impacted (b) Run 2: 534 buildings impacted (c) Run 3: 602 buildings impacted

Figure A1: Variability of fire spread for a fixed wildfire scenario in the `FIRE-2D` model.

This variability implies that a single simulation is not sufficient to characterise a scenario. A sufficiently large number of independent stochastic runs is therefore required to obtain stable estimates of the burn probability for each building. The objective of the convergence analysis is to determine how many independent runs per scenario are needed so that the estimated building-level burn probabilities become stable despite sampling variability.

Experimental protocol

To assess the statistical stability of the estimated probabilities, several representative scenarios were selected and simulated 1,250 times. For each scenario, the available runs were randomly re-ordered into multiple shuffled sequences. Each shuffle corresponds to a different possible ordering of the same set of stochastic outcomes.

Let $a \in \{1, \ldots, N_{\text{shuffles}}\}$ denote the shuffle index, b a building, s a scenario, and r an individual stochastic run. For shuffle a, the cumulative burn-probability estimate after k runs is defined as:

$$p_a(k) = \frac{1}{k} \sum_{r=1}^{k} X_{b,s,r}^{(a)},$$

where $X_{b,s,r}^{(a)} \in \{0, 1\}$ is equal to 1 if building b is reached by the fire in run r of scenario s under shuffle a, and 0 otherwise. Thus, for a given building and scenario, each curve $p_a(k)$ describes how the estimated burn probability evolves as the number of runs increases under a different ordering of the stochastic outcomes.

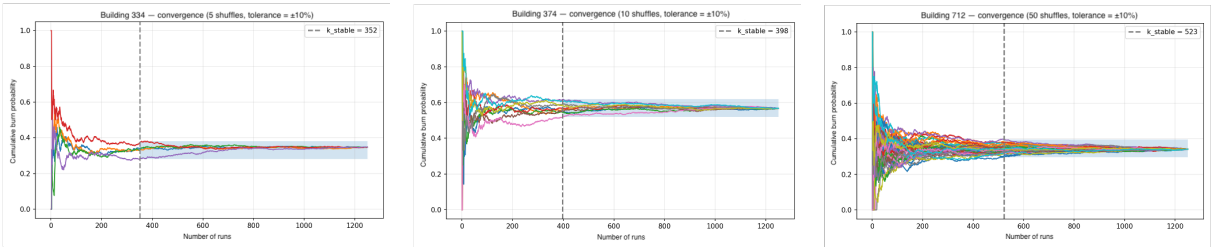
Stability criterion

A tolerance $\varepsilon = 10\%$ is used to assess convergence. For a given building and scenario, the stable number of runs k_{stable} is defined as the smallest integer such that, beyond this value, the cumulative estimates obtained from all shuffles remain within a horizontal band of width at most ε:

$$\max_{\substack{a \in \{1,\ldots,N_{\text{shuffles}}\} \\ k' \geq k_{\text{stable}}}} p_a(k') - \min_{\substack{a \in \{1,\ldots,N_{\text{shuffles}}\} \\ k' \geq k_{\text{stable}}}} p_a(k') \leq \varepsilon.$$

In other words, after k_{stable} runs, the estimated burn probability is considered stable if the remaining variability induced by the ordering of the stochastic runs stays below the prescribed tolerance.

For each scenario, this criterion is applied across buildings. The building exhibiting the slowest convergence is retained as the worst-case building for that scenario. Fig. A2 illustrates the cumulative estimates $p_a(k)$ for one representative scenario and for the corresponding worst-case building, using 5, 10, and 50 shuffles.



(a) 5 shuffles: $k_{\text{stable}} = 352$ (b) 10 shuffles: $k_{\text{stable}} = 398$ (c) 50 shuffles: $k_{\text{stable}} = 523$

Figure A2: Cumulative burn-probability estimates $p_a(k)$ for a representative wildfire scenario, obtained with 5, 10, and 50 shuffles for the worst-case building.

Results

The convergence analysis was conducted on four representative scenarios, with $N_{\text{shuffles}} \in \{5, 10, 50\}$. For each configuration, three quantities were extracted: the maximum value of k_{stable} over buildings, the 90% quantile of the building-level stability values, and the building corresponding to the worst-case convergence. The results are summarised in Table A1.

		Scenario 1			**Scenario 2**			**Scenario 3**			**Scenario 4**		
N_{shuffles}	ε	k_{stable}	q_{90}	worst-case building	k_{stable}	q_{90}	worst-case building	k_{stable}	q_{90}	worst-case building	k_{stable}	q_{90}	worst-case building
5	10%	235	22	253	509	387	718	352	86	334	389	207	1051
10	10%	241	148	245	509	387	718	398	122	374	391	207	1051
50	10%	355	213	253	570	477	459	523	276	712	517	337	584

Table A1: Summary of the convergence diagnostics obtained for the four representative wildfire scenarios as a function of the number of shuffles.

The results show that, even in the most stringent configuration with 50 shuffles, the maximum values of k_{stable} remain of the order of a few hundred runs across the four representative scenarios. The 90% quantiles remain below the worst-case values, indicating that only a limited number of buildings exhibit slower convergence. Based on these diagnostics, $N_r = 500$ independent stochastic runs per scenario are retained in the main analysis as a practical compromise between statistical stability and computational cost. This value provides robust building-level burn-probability estimates for the wildfire case study while keeping the full design computationally tractable.