Open **and** Interoperable **Spatial Decision Support Systems** for Multi-Hazard **Resilient** Cities

Research Proposal prepared by **Justin Schembri**

Go/No-Go Meeting: **9 December 2024 (last version document: 5 March 2025)**

Promoter: **Prof.dr.ir. P.J.M. van Oosterom**
Co/Promoters: **Prof Dr Mauro Overend, Dr Azarakhsh Rafiee, Dr Simona Bianchi**

TU Delft

Contents

1 RESEARCH SUMMARY

Urban areas must contend with a swelling risk envelope driven by the increased frequency and intensity of climate and weather-related hazards, aging building stocks, land use change, and an overall trend toward increased social vulnerability. Urban vulnerability is defined as the inability of buildings and urban populations to perform under induced hazard stresses. To a greater extent, reducing vulnerability is the primary mitigable component of risk. Urban and social vulnerability is modeled in several ways, though all models attempt to correlate an expected level of harm to the building stock and populations when impacted by a hazard strike or disaster event. Risk and vulnerability modeling at the urban scale invariably involves assumptions, aggregations, and uncertainties. Decision-makers utilize the results of these models to make choices such as how to distribute funding to civil protection, or wether to fund infrastructural projects such as dikes. Thus, model efficacy crucially impacts and possibly biases the final decision taken. Due to the complexity of considering all aspects, all hazards and all the decision pathways possible, decision-makers (e.g., urban planners, emergency response) are often aided by geospatial software systems known as Spatial Decision Support Systems (SDSS). The absolute majority of Natural Hazard DSS developed focus on the risks posed by the *single hazard* and be somewhat narrow in scope.

DSSs capable of considering the potential risks posed by multiple hazards (multi-hazard) have received sparse attention for various reasons (see Section 5), even though such a multi-hazard, multi-criteria system has the potential to facilitate more holistic decision-making. Multi hazard SDSS' offer potential gains, primarily that it facilitates more holistic approaches to risk reduction to be taken. However, the complex distributed systems, methodologies, and standards required for the development of durable multi-hazard DSSs remain underdeveloped and present a significant research gap. This research gap arises from several factors, explored elsewhere, but is largely attributed to a lack of semantic and technical interoperability between hazard domains.

In response, this work proposes to investigate the **design and development of a standards-based information and processing infrastructure that leverages heterogeneous data sources and services within a unified environment, supporting multi-hazard decision support systems.** The research is divided into two major components: the first involves the development of an integrated, standards-based data and processing (digital) system for conducting *multi*-hazard risk assessments, while the second focuses on leveraging this system to make novel contributions to multi-hazard risk assessments.

1.1 The Multi-Hazard Domain Model and Risk Assessment Library

In the first portion of this research, and following a careful consideration of three major risk domains—seismic, flooding, and overheating—a novel and suitable multi-hazard domain model (MHDM) will be investigated and developed. A *domain model* is a collection of technical and ontological abstractions that describe the relationships between different components (e.g., data, simulations) forming part of a multi-hazard SDSS. It shall broadly define the following:

1. a collection of technical, semantic, and ontological goals or *requirements* that a multi-hazard SDSS should satisfy,
2. how urban *multi*-vulnerability is modeled,
3. reusable and shareable *multi*-hazard simulation processes from a software architecture perspective,
4. the representation and dissemination of *multi*-risk simulation results, and
5. the integration of *multiple* data sources and emerging technologies into the risk assessment process for more comprehensive approaches.

This domain model will leverage and be built around existing geoweb standards, data formats, and data sources, as well as the particularities of individual hazard domains. The domain model shall be implemented and developed into an OpenAPI-compliant service and digital library, enabling users to serve and execute multi-risk simulation processes as a geoweb service. The novel domain model facilitates the integration of existing open-source packages catering to individual hazards, ensuring that well-supported and established developments contribute to the work. At this stage of the research, the library will provide a clear approach to reconciling the many technical differences between multi-hazard domains and will explore the problem in the following concrete ways:

1. The research and development of a **shared semantics and conceptual MHDM** defining the input and output data for seismic, flood, and overheating risk, including semantic and technical solutions for multi-hazard database management systems (DBMS), data exchange, processing services and CLI or GUI all from a multi-hazard perspective.
2. The selection and extension of existing geoweb standards, most notably Open Geospatial Consortium (OGC) standards, and other data formats suitable for the multi-hazard problem, enabling standardized encodings for effective data exchange in the multi-hazard domain.
3. The development of standards or schemas for new geospatial processes (e.g., simulations) from a multi-hazard perspective, enabling interoperable processing services.
4. The creation of standards or schemas for incorporating and leveraging real-time sources such as sensors into the DSS, enabling sensor-web technologies to contribute to the multi-hazard problem.
5. Interoperability guidelines with existing, strongly supported single-hazard open-source tools, encouraging the reuse and reconciliation of single-hazard domain models and actual data and servies into a novel multi-hazard domain model.

1.2 Novel Contributions Made Possible by the Domain Model

The domain model will be assessed and used to develop a prototype multi-hazard SDSS. Building upon the integrated and harmonized approach the multi-hazard domain model offers will enable the development of several novel contributions to natural hazard decision-making processes. These contributions may be broadly categorized into two types:

1. The development of data-driven multi-hazard vulnerability models (Subsection 1.3).
2. Novel methods for capturing or reducing multi-hazard uncertainty (Subsections 1.4 and 1.5).

1.3 Baseline Vulnerability, Consequence, and Loss Datasets

The component of the domain model that describes data encoding and the technical exchange model will be applied to existing historical records of performance and damage to the built environment related to natural hazard events. These records, normalized to the domain model, will be statistically leveraged to form a baseline dataset of regional expected losses for varying intensities of natural hazards from a multi-hazard perspective. Statistical and machine-learning methods will be used to develop an initial suite of cross-building, multi-hazard fragility and/or vulnerability functions, conditioned on relevant predictor variables.

A sensitivity analysis will assess statistical dependencies and sensitivities to explicitly quantify uncertainties within building performance models. Additionally, methods will be developed to enrich the vulnerability functions when more detailed local information is available. A hybrid approach for incorporating numerical methods will complement long-return-period hazards (e.g., seismic) in the vulnerability functions.

1.4 Reducing Uncertainty through Hybrid Simulation and Observation Techniques

The underlying models used to perform risk analysis may be broadly divided into empirical or numerical methods, the merits of which will be discussed in Section 5. The domain model will support robust harmonization of static data (e.g., flood maps or seismic hazard maps) with sensors, IoT, and other real-time emergent technologies to improve risk assessment capabilities and reduce uncertainties. Geostatistical techniques will enhance the models by enabling real-time sensor data to update and refine the baseline through Bayesian techniques, particularly for low-return-period hazards such as building overheating. Simulation processes will be served

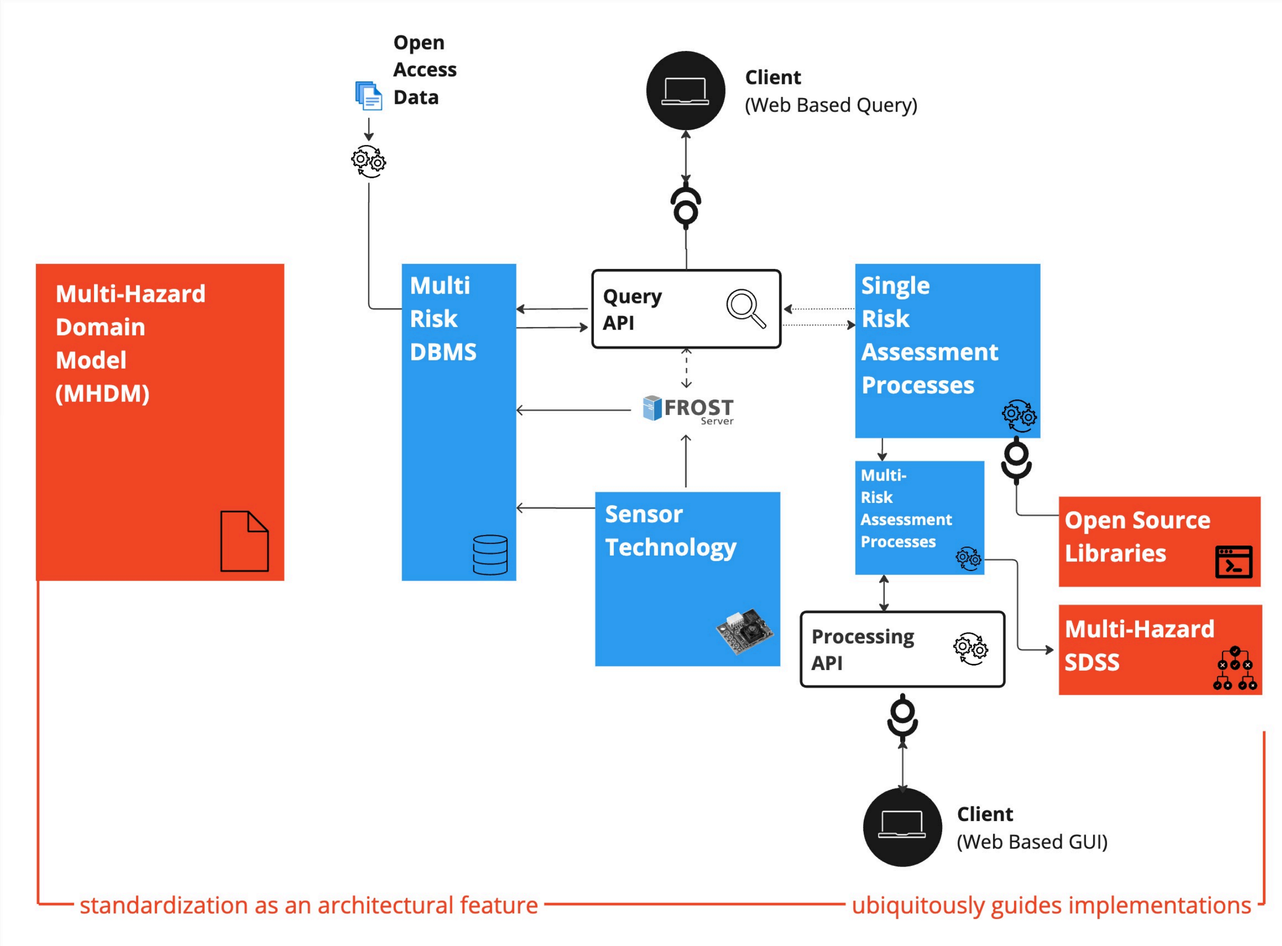

Fig. 1. Simplified architectural schema showing the components of a multi-hazard SDSS.

in an interoperable geospatial compliant API, namely the OGC Processes API. An example of this work has already been carried out in the first year (see Appendix A).

1.5 Uncertainty Propagation and Model Performance Comparisons

Risk assessments inherently embed several aleatory and epistemic uncertainties, both of which are discussed in Section 5. The propagation of these uncertainties across models remains an area of active research. The integrated approach will facilitate the development of methods for numerically tracking uncertainty at each step and from each input. Often, the uncertainties of specific components in a risk assessment are unknown, requiring the use of baseline figures as a reference (see Section 1.4). The integrated domain model will enable the development of novel methods for estimating multi-hazard uncertainty with respect to hazard, vulnerability, and exposure components.

1.6 Final Contribution and Novelty Summary

The final contribution of this research is an open-source geospatial variant of the DSS—a Spatial Decision Support System (SDSS)—integrated with existing open-source tools (e.g., OpenQuake, EnergyPlus) to provide robust multi-hazard decision-making support. A sketch of the architectural schema is shown in Fig 1. This system will utilize the **novel harmonized MHDM**, allowing users to conduct multi-risk urban assessments using heterogeneous data and model sources. The SDSS will be further enhanced by a normalized baseline vulnerability, consequence, and loss dataset derived from existing sources, while capturing the propagation of uncertainty throughout the entire risk assessment pipeline.

2 BACKGROUND AND MOTIVATION

The assessment of urban vulnerability to geo-risks is a fundamental component of disaster risk reduction (DRR) and efforts to mitigate the worst impacts of natural hazards. Vulnerability, in broad terms, is defined as the capacity of an entity to resist an applied stressor or force. In the context of geo-risks, *vulnerability* refers to the ability of individuals, groups, societies, buildings, infrastructure, and environments to withstand the effects of natural hazard events such as earthquakes, storms, or floods. Weather-generated hazards (e.g., floods, winds, storms) are currently being exacerbated by anthropogenic climate change, while non-weather-related geo-risks (e.g., seismic events, tsunamis, volcanic activity) are increasing as urban environments densify and building stocks age [5].

Numerous pathways exist to address rising urban vulnerability, including building retrofits, reconstruction, or engineering interventions (e.g., [12], [16]). These interventions span a wide spectrum of socioeconomic costs. Therefore, before developing action plans or strategies, the concept of 'risk' must first be qualified—or ideally quantified—to enable meaningful comparisons before and after the intervention. A *risk assessment* process thus underpins the decision-making process [17], ensuring that applied interventions effectively reduce societal vulnerability.

Human civilization is now more complex and interconnected than ever before, underscoring the need for robust risk assessments. Contemporary natural hazard risk assessments increasingly strive to address the complete risk profile by considering multiple hazards (i.e., *multi-hazard*) that could plausibly affect the area of interest. Furthermore, the time horizon of these assessments varies significantly, ranging from hours to years or even decades. Although multiple hazards are generally unlikely to occur simultaneously, scenarios where a second hazard strikes during the recovery period are plausible [30].

One additional aspect to consider when conducting such risk assessments is that the potential harm or losses a society may face extend beyond purely economic impacts. More recently, the concept of 'loss' has evolved into a broader, more robust framework of societal *resilience*. In the context of this work, urban resilience is viewed as an extension of vulnerability—one that integrates tangible losses (e.g., monetary losses or casualties) with peripheral losses, such as loss of functionality, environmental degradation, and time to recover. Resilience encompasses not only the immediate negative impacts of a hazard event but also the initial loss of functionality of the affected asset and the time required to restore that functionality (see, for example, [11]). Consequently, within this work, references to *risk* inherently include considerations of *resilience*.

It follows that the assumptions, models, simulations, theoretical frameworks, and all the components that form integral parts of the risk assessment process have a critical impact on the selection of mitigation strategies by decision-makers. Poor assessments lead to poor decisions, while the inverse should, at least in theory, hold true. Consequently, the underlying data and models guiding multi-hazard risk and resilience assessments play a decisive role in the outcomes of decision-making processes. Such data takes several forms but is generally divided into three key partitions: hazard, exposure, and vulnerability.

1. **Hazard datasets** include models, simulations, or empirical data describing the likelihood of various hazard intensities, e.g., flood inundation depth, earthquake magnitude, storm wind speed, or peak ambient temperature.
2. **Exposure datasets** describe the geospatial—and possibly temporal—distribution of individuals, societies, and buildings within a given area.
3. **Vulnerability models** are modeling approaches that transform the impact of a hazard into an expected loss. The simplest vulnerability model may take the form of a ratio, e.g., €200 of economic losses per 10 cm of flood inundation per square meter of property. More complex approaches are often multivariate. As previously discussed, loss has several potential metrics, such as economic, human casualties, environmental degradation, or resilience loss. Vulnerability models are hazard-dependent and vary across hazard types. These models are typically developed using heuristic, empirical, probabilistic, numerical, hybrid, or machine learning (ML) approaches.

Among these three partitions, modeling vulnerability is perhaps the most sensitive component of the assessment process, as it addresses the complex interactions between the stressor (the hazard) and the entity of interest, sometimes referred to as the *asset*. Vulnerability modeling remains an active area of research and is likely to continue as such (e.g., [27], [32], [22]). Non-empirical techniques, which do not require the prior occurrence of an event, hold significant potential for advancing vulnerability modeling. While empirical vulnerability models serve as valuable validation tools—and more recently as integral parts of hybrid machine learning techniques—the availability of such data is often limited for various reasons, despite its utility. This scarcity is frequently cited as a limiting factor in data-driven approaches to vulnerability modeling.

However, with respect to the empirical datasets that could contribute to improved vulnerability modeling, it is proposed that the problem lies not in data scarcity but rather in data fragmentation, lack of harmonization, and the absence of supporting digital infrastructure and legislation. In terms of infrastructure, the European Union and its member states have, for example, legislated and implemented the INSPIRE framework [44], which defines the geospatial infrastructure that member states must implement. Nevertheless, these developments have been relatively light on legislating for the collection of improved urban vulnerability data—data that is more organized, refined, and interoperable—supported by long-term collection infrastructure.

Historically, there have been significant and successful efforts to unify hazard datasets from global disaster events. These efforts have enabled the development of seminal predictive models describing hazard generation and associated probabilities, such as earthquake ground shaking predicted by ground motion prediction equations (e.g., [9]). Equivalent efforts applied to vulnerability modeling, however, are less prevalent and often constrained by regional boundaries. Exceptions, particularly in the seismological domain, have been limited by relatively small datasets (e.g., [47], [51]).

Despite the current limitations of empirical vulnerability modeling, other modeling techniques continue to play a central role in the risk assessment process by filling critical data gaps. These alternative approaches remain indispensable in providing actionable insights for disaster risk reduction and resilience planning.

Assessing, qualifying, and quantifying vulnerability and risk represents one side of the challenge faced by decision-makers. Ultimately, efforts are made to qualify vulnerability to facilitate the qualification of risk, with the end goal of guiding actionable plans. These plans aim to optimize limited economic resources to implement modifications to the physical environment, enhancing societal preparedness for events that may never occur. To navigate the complexity of such decisions—often involving dozens of variables—decision-makers frequently rely on software packages known as **Spatial Decision Support Systems (SDSS)**.

In the context of this research, an SDSS is defined as any software tool that allows decision-makers to explore the potential outcomes of Disaster Risk Reduction (DRR) strategies, ideally (but not necessarily) through interactive means. While an SDSS does not inherently require the capability to calculate risk internally, it must fulfill the following core functions:

1. Enable decision-makers to model one or more DRR mitigation measures.
2. Recalculate, using a defined risk quantification methodology, the impact of proposed measures on key performance indicators (KPIs), risk metrics, or decision variables.
3. Provide reports, visualizations, and feedback on the efficacy of proposed measures.

A multi-hazard SDSS, as opposed to a single-hazard SDSS, offers the advantage of providing a more holistic view of societal risks. This broader approach to risk calculation, as argued in this research, represents an ideal pathway for making robust and equitable decisions. This is particularly pertinent for decisions with national or regional impacts, such as those required to address climate change. Despite the strengths of the multi-hazard perspective, there has been limited development within the multi-hazard domain, especially concerning multi-hazard SDSS. Historically, SDSS tools have been developed to address specific hazard domains, such as seismic, flood, or landslide risks. This is understandable, given the inherent complexity and uncertainty (see Section 5) involved in achieving acceptable assessments even for single-hazard risks.

For instance, within the same hazard domain, there are often multiple methodologies for modeling fundamental components of the hazard:

- **Seismic hazard** assessments are typically carried out using empirically derived *Ground Motion Prediction Equations* (GMPEs), of which several sets exist based on research.

- **River flood (fluvial) hazard** modeling may involve hydraulic inundation models (1D, 2D, or 3D) or be driven by empirical data.
- Urban **overheating hazard** is a multifaceted problem characterized by various metrics at different scales, such as land surface temperature, canopy air temperature, heat stress indices, and indoor thermal comfort indices.

Similarly, the exposure component of risk varies significantly across hazard domains:

- **Seismic exposure** modeling requires data on structural characteristics, such as building type, construction year, and height.
- **Flood exposure** modeling necessitates information about building contents, the presence of basements, and drainage capabilities in the affected area.
- Urban **overheating exposure** modeling depends on demographic factors (e.g., resident age) and building characteristics (e.g., thermal properties).

Finally, vulnerability modeling differs across hazards:

- **Seismic vulnerability** primarily depends on structural performance.
- **Flood vulnerability** is influenced by variables such as inundation duration, construction typology, water depth, and the value of finishes and fixtures.
- Urban **overheating vulnerability** is conditioned by factors such as population age and the availability of mitigation measures (e.g., air conditioning).

It is evident that developing an SDSS capable of assessing a comprehensive risk profile and re-evaluating risk following the hypothetical implementation of mitigation measures requires reconciliation across the diverse approaches inherent to each individual hazard domain. **Multi-hazard and multi-criteria SDSS tools hold substantial potential for informing decision-making at the most significant scales, provided semantic and technical reconciliations are achieved.**

Throughout this proposal, reference will be made to the absence of a suitable multi-hazard **domain model** as one of the key research gaps. This domain model, this research will argue, is the foundation on which any multi-criteria SDSS must be developed around. A domain model is an software design abstraction which broadly defines:

1. The data requirements and encoding formats needed by the system.
2. The approach to modeling vulnerability.
3. The interactions between modeling and risk assessment processes.
4. The methods for serving and disseminating results.

In response to these challenges, this research proposes the development and implementation of **a novel multi-hazard domain model**, culminating in the creation of an **open-source multi-hazard SDSS**.

3 PROBLEM DEFINITION

The multi-hazard DSS is a relatively novel concept which encounters a series of barriers. These research limits may contribute, in part, to the relatively sparse availability of multi-hazard SDSS' as packages or methodological approaches. The pertinent problem this research tackles is thus categorized into challenges of:

- multi-hazard **domain model design**, development and uptake
- interoperability, incompatibility and technical exchange limitations
- uncertainty quantification, propagation and management
- vulnerability, consequence and result validation
- reflection in assessment of added value

3.1 Multi-Hazard Domain Model Design

There is no widely accepted, nor deeply investigated MHDM. The limited and existing multi-hazard SDSS software reviewed, (e.g. RiskScape [45], HAZUS [50]) takes a relatively ad-hoc approach to the modeling component. There is a notable lack of emphasis, and attempt to build upon existing standards, protocols or projects; every project seems to "start from scratch" to some greater or lesser extent. Developing a suitable domain model which is built on existing standards increases long-term stakeholder input. This is a recognized problem, and recently the Open Geospatial Consortium (OGC) began running pilot projects such as the '*Climate and Disaster Resilience Pilot 2024*'[1] in order to begin considering standards which will help practitioners re-use their data, models and results.

3.2 Lagging Data and Process Interoperability in Disaster Risk Reduction Generally; Other Incompatibilities

Despite the ever-increasing importance of data, models and processes in DRR research, interoperability remains a lagging concern among many stakeholders [34]. The problem is a broad one that extends past the multi-hazard domain, and is an aspect affecting many DRR efforts, driven primarily by:

1. **lack of standardization**: varying actors have different standards for data collection, processing and distribution, creating an initial friction and hindering cooperation
2. at the technical level, disaster data modeling is a relatively new form of research. The **lack of clear and enforced data-structures** or **process models** for the many data components required as inputs to the SDSS limits interoperability

Furthermore, incompatibilities exist between the different domain considerations that may be difficult to resolve. Different hazards have widely varying return periods, and the data collected for each reflects this. For example, while ample records exist for the damage caused by seismic events, most capture low- to medium-intensity events due to the low return period of high-magnitude earthquakes. Consequently, while there is substantial data for lower damage states, the scarcity of high-magnitude events limits references.

3.3 Uncertainty Propagation: Quantification, Estimation and Communication

Every component that contributes to an underlying risk model in an SDSS includes uncertainty. At smaller scales of risk analysis, even detailed approaches such as expert engineering analysis' of buildings must tackle various uncertainties embedded into assumptions, especially when considering existing situations or buildings. Some examples of such uncertainties may include the quantity and detailing of structural reinforcement at critical building joints (seismic risk); the true U-value or energy-consumption of a building (over-heating or energy risk).

At the meso- (neighborhood) or macro-scale (city or larger), detailed simulations become unfeasible. Thus, analysts must often approach the problem through simplifications and aggregations, often leveraging statistical models or empirical data (Section 3.4) to reach useful risk-conclusions. Uncertainty propagates more broadly for large analysis areas, e.g.: uncertainty on the properties of buildings in a neighborhood (exposure) is carried over and amplified by uncertainty about how buildings with those given properties behave (vulnerability) under an uncertain hazard profile (hazard). In an integrated multi-hazard SDSS, the uncertainty from *all* aspects of all hazards come together and a reasonable question to ask is: *'how certain are these results?'*

3.4 Prediction's Sensitivity to Vulnerability Models

Vulnerability data and / or models are critical components of risk modeling and vital to decision making processes. Such models may be guided by analytical or empirical methods. Empirical methods are driven by records of observed building damage or performance, which are generated in a fragmented manner by individual researchers and institutions for isolated events. While there has been concerted effort to collate hazard statistics, similar efforts for vulnerability and damage statistics have received less scrutiny. Furthermore, the efforts that have been made are one-time analyses and do not possess the supporting infrastructural capacity to consume new data as it becomes available.

Empirical vulnerability data typically falls into three main categories: 1) post-event damage records, 2) instrument or sensor readings, and 3) experimental data, usually at the component level. Practitioners use these observations to create vulnerability curves that estimate the probability of a building reaching a specific damage state (e.g., light, medium, or high damage) at various levels of natural hazard intensity (e.g., ground motion during an earthquake). Such prediction models, which are straightforward to apply in assessments, can estimate expected damage for a specified event (deterministic hazard assessment) or for a range of possible events (probabilistic hazard assessment). However, the empirical approach to classifying vulnerability has certain limitations, as described in the forthcoming paragraphs.

1. The first limitation arises from the broad variation in building typologies, even within small populations. Empirical damage prediction models are generally based on high-level building attributes, such as height, structural type, water-tightness, etc. In reality, no two buildings are identical, but modelers must simplify a building stock into sets of archetypical buildings. The level of detail (LoD) within these archetypes can vary significantly, and simplified categorizations are common. This reductionist approach addresses both computational constraints and data scarcity. For example, post-earthquake damage data might categorize buildings into three classes: "high-rise," "medium-rise," and "low-rise." Empirical damage observations can then be used to develop probability density functions (PDFs) describing the likelihood of each category reaching a certain damage state given a level of ground shaking. However, due to limited granularity in these models, predictions based on such PDFs are likely to exhibit substantial variance. For small data samples, these simplifications become necessary, but they introduce significant uncertainties, thus small datasets or limited detail compel the use of coarse building archetypes, increasing uncertainty and variance in predictions.

A database of cross-regional records of building damage from natural hazards would enable more granular predictions—if another hurdle could be overcome: the lack of standardized semantics for describing buildings, damage states, and performance levels. Even basic, seemingly objective descriptors of building attributes can lack uniformity; for instance, building height may be recorded in metric or imperial units, or simply by floor count. Without a well-defined schema, a database of this kind would quickly become difficult to manage and limited in utility. Thus, expanding empirical damage datasets with cross-regional sources is challenging due to a lack of semantic uniformity.

Even with a common modeling language for building typologies, damage states, and performance metrics, another challenge persists: the significant variability in data resolution. Some data collections are highly specific, while others remain sparse, leading to inconsistencies in data granularity. This variability complicates statistical analyses, introducing confounding variables that obscure genuine patterns and correlations. Consequently, inconsistent data resolution confounds statistical processes, limits predictive model reliability, and propagates uncertainty.

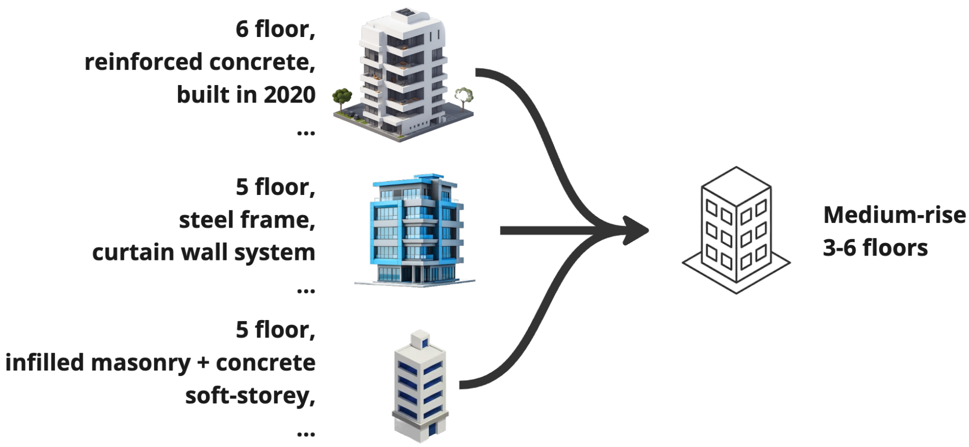

Fig. 2. Archetyping buildings makes computation on large datasets feasible, however it can lead to over-fitting of predictions.

While the empirical data described up to this point has historically provided the foundation for vulnerability assessments, real-time data is also a valuable complement. As the prevalence of digital infrastructure expands, and the Internet of Things (IoT) domain grows, real-time observations have the potential to enhance vulnerability assessments and provide critical validation layers. This sensor-driven data introduces a dynamic aspect, particularly useful in rapid vulnerability assessments. However, like traditional observational data, sensor networks are often fragmented, multi-format, and lack standardization. Although using sensor data for building-level vulnerability predictions is gaining traction as such systems become more widely accessible, this field is still in its infancy, and standardization challenges persist. Thus, methods for deriving vulnerability from sensor-based observations are underdeveloped, with standardization issues remaining a significant challenge.

In conclusion, data scarcity, semantic incompatibilities, and infrastructural shortcomings drive errors due to over-simplification, limit access, and introduce data management problems, increasing the overall uncertainty (See Figure 2). A concerted effort at managing the empirical data described thus far offers the potential to address situations of data scarcity in multi-hazard resilience-based assessments. It would be desirable for such an effort to include provision for suitable digital infrastructure for handling new data as it is produced. As natural-hazard impacts become more severe and frequent, the observation of ground-truth could serve as a fundamental reference tool for validation, planning, and distribution of limited resources. Perhaps most critically, equitable and open access to empirical data would finally reduce the need for costly, resource-intensive vulnerability assessments where data is scarce.

3.5 Added Value of Approach

As part of the reflection process, the MHDM must inevitably be compared to the predominant "single-hazard" approach. The single-hazard approach assesses risk on a hazard-by-hazard basis, with the results then merged and compared—often with minimal inter-hazard or interdisciplinary collaboration facilitated by the software. In contrast, it is proposed that the MHDM (and its implementations) will prove to be *greater than the sum of its parts*. This critical proposition will be demonstrated through comparative case studies, where the same fictional brief is analyzed using the MHDM alongside other available approaches.

4 RESEARCH SCOPE

The scope of this research encompasses the design, development, and practical application of a **multi-hazard domain model**, culminating in the implementation of an **integrated multi-hazard spatial decision support system (SDSS)**. The hazards to be integrated into the SDSS are **seismic, flood and urban over-heating**. The design and development of the domain model include the following key sub-scopes:

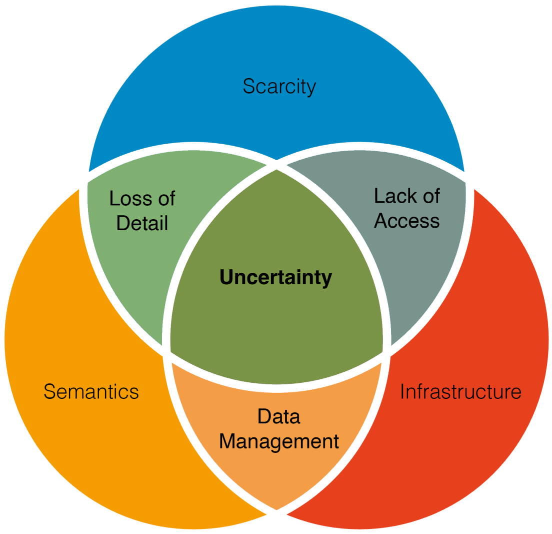

Fig. 3

- **Exposure harmonization** – Developing models to align and harmonize the diverse requirements for representing building and population exposure across multiple hazards.
- **Multi-hazard vulnerability modeling and record keeping** – Establishing effective approaches to model vulnerability across multiple hazards; utilizing a common approach to vulnerability modeling to re-purpose historical datasets into new vulnerability products,
- **Process harmonization** - Development of technical process models using existing standards for use with existing or new risk related simulations or processes,
- **Multi-hazard risk representation** – Designing methodologies to accurately represent the complex and interdependent nature of multi-hazard risk.
- **Enhancement of multi-hazard risk assessments** – Improving the accuracy and robustness of risk assessments by adopting an integrated approach that accounts for multi-hazard interactions.
- **Integration of multi-source data** – Combining diverse sources of risk-related data, such as static hazard maps, with emerging technologies, including sensor and IoT systems, to enhance the quality and timeliness of risk assessments.
- **Delivery, interaction and extendibility** – Developing novel methods and tools for serving, visualizing, and interacting with the outputs of multi-hazard SDSS, given their nascent development stage. Furthermore, what long-term design approaches will allow the SDSS to receive updates and be extensible?

Some additional detail for the sub-scopes 1) and 2) are presented in the upcoming subsections.

4.1 Exposure harmonization

The harmonization sub-scope consists of approaches for ensuring **domain** and **digital** interoperability. Domain interoperability addresses challenges in information modeling, ensuring that data types and processes serve multiple hazard domains. Similarly, digital interoperability deals with issues related to architectural and data structure choices (e.g., through the use of a common querying language) to achieve the same goal. The harmonization effort will build upon existing efforts in open data provision, standardization, and infrastructural developments, namely:

- Open Geospatial Consortium Standards (OGC),

- Humanitarian Open Street Map (HOT),
- The EU INSPIRE framework,

The harmonization approaches developed shall be applied throughout the work, i.e., outputs from other sub-scopes are themselves constrained by the harmonization approach (??).

4.2 Multi-hazard vulnerability modeling and record keeping

The harmonization efforts will guide a process whereby un-harmonized historical records associated with fragility or vulnerability or collated into a backwards facing vulnerability database. This data will be processed, and normalizing available open-source data will be developed, possibly with the use of machine learning and AI. A database and associated end points will be investigated such that a methodology for the extension of new information as it is collected by various researchers may be added to the multi-hazard vulnerability database. An open and automated approach that encourages the submission of new data to the database will be developed, focusing on validation and data processing. The database will be served with automated methods of utilizing the database in statistical processes to drive the development of new vulnerability tools, namely: 1) multi-hazard vulnerability curves and 2) building performance prediction models.

5 RELATED WORK

5.1 Decision Support Systems (DSS)

A Decision Support System (DSS) may be defined as an integrated methodology that assists users in selecting one or more actionable decisions from a set of possible choices [25]. An actionable decision is one that facilitates the formation of a plan and the execution of a practical intervention. Complex decision-making processes are often supported by software, and thus a DSS is typically packaged as a software application or digital toolkit.

In the context of Disaster Risk Reduction (DRR), a DRR-DSS generally follows a methodological framework for the quantification and/or qualification of risk. The risk component in DRR is commonly abstracted as the intersection between a hazard event and the vulnerability of the exposed assets. A Spatial Decision Support System (SDSS) for Natural Hazard Risk Reduction is a domain-specific DSS designed to provide geospatial decision support for natural hazard risk mitigation measures.

Within the scope of this research, a software suite qualifies as an SDSS if it includes three base characteristics:

1. The ability to analyze hazards and exposed assets geospatially and at macro-scales (e.g. block, neighborhood or city level)
2. A methodological framework for the quantification and/or qualification of natural hazard risk,
3. An interface that allows the exploration of potential risk-reduction strategies (as proposed by [40]),

An SDSS essentially provides a two-step feedback loop whereby the existing risk is first analyzed, driving the design and application of mitigation strategies such that the resulting risk may be re-analyzed, and the process continues iteratively. At its core, an SDSS is supported by a suite of theoretical models and datasets through which the risk-reward feedback loop can be quantified. These models are broadly sub-categorized into models and datasets for considering:

1. The **natural hazard**, in terms of:
   - The probabilistic profile of various hazard intensities manifesting (e.g., flood depth, peak ground acceleration, or wind-speed probabilities),
   - Return periods of certain events (e.g., a 1-in-100-year return period storm),
   - The geospatial distribution of the hazard in terms of intensity levels,
   - The physical mechanisms driving the hazard (e.g., plate tectonics and geological faulting for earthquakes, hydrometeorology for flooding).
2. The **vulnerability** of exposed assets, with respect to:
   - The ability of the built environment to withstand varying levels of hazard intensity over a given period for a specific duration,
   - The physical mechanisms at various building components and levels that govern the response to hazards (e.g., structural aspects of a building with respect to earthquake vulnerability),
   - The socio-economic aspects, which are often difficult to quantify, associated with social resilience and response capabilities. These include, but are not limited to:
     - Disaster response capability,
     - The availability of economic recovery funds,
     - Cultural issues and inequality.
3. The geospatial distribution of various **exposed elements** of society at risk, namely:
   - Individuals and communities,
   - The general building stock,

  - Key infrastructure (e.g., transport networks, energy networks, and areas of industry),
  - The temporal variability of these elements.

4. The ways in which mitigation measures interact with and reduce or modify the various components of risk, thereby altering the overall risk,

5. The multi-criteria decision-making process that ultimately leads to one or more optimal decisions being reached.

These models, often packaged as software modules, come together to form the core SDSS pipeline. It is evident that the reliability and inherent uncertainty of each of these models significantly affect the quality of the decisions reached. Thus, the integration of state-of-the-art risk science into an SDSS applications is critical to the success of such a system. In the upcoming sections, the current approaches to four major risks, Earthquake, Flood, Wind and Heat are explored. Before proceeding, it should be noted that many of the models introduced in the upcoming sections remain an ongoing subject of scientific research and debate.

5.2 Earthquake Hazard Modeling Techniques

Earthquakes are a natural part of geotectonic processes. The primary physical process that drives earthquakes is the motion of tectonic plates against, away from, or into each other. The speed of movement of these plates is varied and relatively slow, with maximums of around 10 cm per year relative to one another. Irrespective of this relative slowness, these incompatible motions give rise to huge internal forces and pressures, which in turn drive the formation of seismic faults that geologists and seismologists map out geospatially.

Seismic faults come in several flavors and are described by the direction of motion of the plates relative to each other. For example, a strike-slip fault involves two plates sliding longitudinally against one another. There are also normal and reverse faults, which involve plates moving away from or into each other, respectively; it should be noted that the actual movements are far more complex than this oversimplification. Earthquakes predominantly occur at these seismic faults ([55]), and seismic faults predominantly (but not exclusively) occur at the junctions between significant geological features such as:

1. The junctions between the thick continental crust and thinner oceanic crust (e.g., the Pacific ”Ring of Fire”),

2. Along geologically active mountain ranges (e.g., the Apennines),

3. Along the borders of smaller tectonic plates,

Seismology is a mature field, and while the mechanics that cause earthquakes are well understood, the problem of estimating future ground motion is largely supported by probabilistic methodologies, most notably the Probabilistic Seismic Hazard Assessment (PSHA, [14]). More recently, there is ongoing debate about the validity of the PSHA approach (e.g., [35]), even though it has been used for over 50 years. Seismologists develop methods of ”characterizing” faults through geological studies and are greatly supported by historical records of ground motion, collected through free-field seismic equipment such as accelerometers and global seismic networks ([15]). These historical records allow the reconstruction of a statistical relationship via regression techniques, where the expected ground motion caused by some earthquake is conditioned by several attributes such as:

- The earthquake magnitude,
- The distance (of the site of interest) to the fault,
- The fault type,
- The depth of the earthquake hypocenter,
- The soil conditions and site effects at the point of interest, which attenuate the seismic waves,

The damage caused by earthquakes is the seismic shaking driven by a wide range of complex wave motions the ground undergoes, which travels to buildings that in turn absorb, react to, and damp the seismic waves. There are broadly three wave types generated by earthquakes, namely P-waves, S-waves, and surface waves. P-waves are pressure waves that travel through both solids and fluids, compressing and expanding the ground as they move away from the earthquake origin. S-waves are shear waves that only travel in solids, moving perpendicular to the direction of travel, while surface waves, which are the last to arrive, tend to contribute the most to structural damage due to high amplitude and long durations.

The relationships derived from these historical records are broadly characterized by Ground Motion Prediction Equations (or GMPEs, [15], [10]). From a hazard modeling perspective, the main driver of structural risk is the intensity measure (IM) that an earthquake may generate. Two earthquake IMs are predominantly used by engineers in assessing seismic risk, namely Peak Ground Acceleration (PGA) and Pseudo Spectral Acceleration (PSA). PGA is the expected maximum horizontal acceleration that the ground experiences. Pseudo Spectral Acceleration, on the other hand, is the expected maximum acceleration caused by a given event that a damped oscillator would experience at some given period. PSA is particularly useful for seismic engineering as it correlates the most with what buildings experience.

5.3 Flood Hazard Modeling Techniques

Flooding is a hydrodynamic phenomenon driven by interlinked hydrological and hydraulic processes [53]. Hydrology is the study of the circulation of water through the hydrological cycle and is concerned with how moisture in its various states flows through this system, with precipitation being the main driver of many kinds of flooding. Hydraulic processes, on the other hand, are concerned with the mechanical behavior of water within a system. For example, the mechanics of water flowing down a culvert and flash-flood water flowing through an urban environment are not entirely dissimilar.

There are three primary categories of floods, each with its own hydraulic properties and modeling processes: fluvial, coastal, and pluvial floods. Fluvial and coastal floods require the presence of an existing body of water (rivers, seas, or oceans), while pluvial floods do not. Fluvial floods, sometimes referred to as riverine flooding, are caused by the overflow of rivers, primarily driven by extreme precipitation or snow-melt along the river network [31]. Coastal flooding occurs in low-lying coastal areas, where a combination of storm-driven waters and high tides temporarily inundates large areas [13]. Tsunami flooding is also considered a type of coastal flooding. Finally, pluvial flooding, often associated with flash flooding, is driven by intense precipitation and the inability of an area to drain water quickly enough [43]. This type of flooding is commonly associated with urban environments, where hard, impermeable surface materials are prevalent and the topography has been significantly altered by human development.

As precipitation is one of the primary drivers of flood hazard, hydrometeorology is often the starting point of most flood risk assessments. This field of hydrology is concerned with the transfer of moisture between the land and atmosphere and thus is the main branch concerned with making precipitation forecasts. Hydrometeorological predictions, forecasts and now-casts use observations which are remotely sensed as inputs into models which attempt to describe the physical properties of the atmosphere. Generally, predictions of weather over large areas and moderate time-frame (e.g., one to five days) perform better than predictions for smaller areas in the very short-term [19]. Predictions are geospatial in nature, as rainfall forecasts will include the total rainfall over a given area. They are also, quite critically, temporal: the amount of precipitation in a given period is equally important for risk mapping.

Precipitation predictions become in the input variables of flood-hazard models which may be categorized [23] into:

1. Hydrologic and Hydraulic Modeling
2. Numerical Flood Modeling
3. Runoff Modeling
4. Remotely Sensed GIS Flood Models

Hydrologic and hydraulic modeling is the process by which the flow in water channels (i.e., culverts, streams, rivers, etc) are modeled using computer software. This may be 1D or 2D models, where 1D models work on a cross sectional basis while 2D models also consider longitudinal flow. Several input parameters for these models include a precipitation forecast, evaporation rates, soil moisture along the channel, channel shape, roughness and estimated water velocity. The results of these models are used to simulate the flow rate, volume and water level spatiotemporally (e.g., [3] and [21]). Numerical flood modeling is an physics driven approach that builds upon the principles of hydrologic and hydraulic modeling, except that they are driven by numerical models, algorithms and equations that describe the flow of water.

Runoff modeling is an approach focusing on the transformation of rainwater into runoff and is typically used for flood-plain or urban-flooding. The approaches may be driven by simplified conceptual models (e.g. [6]), empirical models or physics based process models (e.g. [7]).

Arguably GIS methodologies contribute to other modeling techniques to some extent.

5.4 Urban Over-heating Hazard Modeling

Urban overheating may, by comparison to the catastrophic effects of seismic events and floods, appear a mild hazard but this is a mistaken view. Indeed, while its affects are less dramatic, overheating will be experienced much more frequently and by many more populations. The seismic and flood hazards described earlier are "old", familiar hazards. Overheating is, by comparison, a lurching hazard which has crept into urbanized societies which have previously temperate climates such as for example the United Kingdom [26]. Climate change is driving an increase in the intensity and frequency of heatwaves globally, aggravated in urban areas by the urban heat island (UHI) effect. Substantial portions of older building stocks, especially within the EU were designed primarily for heat retention rather than heat removal [39]. Urban over-heating is a hazard which leads to excess deaths, especially among vulnerable members of society [29], as well as significant implications on psychology and the livability of global cities [4].

There are several metrics used to measure the intensity of urban heat beyond solely surface air temperature, the use of which depends on the scale under consideration. *Land surface temperature*, for example is relevant at the City scale[56], *mean radiant temperature* at the street level [33], *indoor air temperature* at the building scale [46] and finally various *heat stress indices* are particular useful at the human scale [24]. The motives that drives considerations at different scales are many, but may be summarized as either issues of 1) human health and comfort or, 2) urban energy efficiency and consumption.

Unlike the other hazards, observing and recording temperature levels should, at face value, be easier. However, heat as experienced by a human population is far more nuanced than can be represented by a simple measurement such as air-temperature. A comprehensive suite of real aspects impacting human reaction to overheating may include:

1. Air surface temperature
2. Humidity
3. Radiation
4. Transitions from indoor to outdoor space
5. Activity levels
6. Wind speed
7. Shading
8. Clothing level

Several sensing technologies can contribute to the collection of environmental data, though all have their respective limits. Weather stations for example are fitted with sophisticated equipment for measuring temperature and humidity levels, but these are often far from urban centers and thus not really representative. An alternative is a network of urban temperature sensors which may be fixed or mobile. Such a network may offer a more comprehensive and granular measurements, yet these again do not capture the real lived experience of populations. Remote sensing has played a significant role in making measurements (e.g., LANDSAT) but these have limitations specifically to surface temperatures.

More comprehensive indices such as Physiological Equivalent Temperature, Predicted Mean Vote, Universal Thermal Climate Index, Standard Effective Temperature and Wet Bulb Globe Temperature are more robust - but still fail to some extent to measure the full impact of over-heating as part of the lived human experience and outlook which may effect health, productivity and livability of a city. This lacuna thus makes it more difficult for decision makers to plan and mitigate in effective ways. A relatively recent development of great promise is the use of the Internet of Things (IoT), which allows for wearable devices which may provide both the physiological aspect and creates, effectively mobile temperature stations.

While observations are particularly useful in analyzing the existing situation, they have also an extended use in the validation of numerical models and simulations. Such models are critical when considering 1) future environmental scenarios and 2) the potential impacts of mitigation measures. Numerical simulations of the environment occur at three major scales, namely:

1. Micro-scale: Closest to the human-scale, considers buildings, streets and walkways. Includes complex phenomena associated with airflow, shading, radiation, air-temperature and more.
2. Meso-scale: Larger scale, considering up to hundreds of square kilometers of area and broader atmospheric conditions. Such models generally provide air-temperature and humidity estimations.
3. Global-scale: Complex weather and heat transfer patterns at a global scale, particularly suited to the climate change modeling and the initial inputs for the meso-scale models.

The integration of such models into DSS' is complex owing to their differing use-cases. Global and meso-scale models have been used to make broader assessments of larger and more widespread mitigation strategies; but micro-scale models take into account the finer urban detail and a truer reflection of potential local improvements that can be made. Thus it would appear that all models have a potential role to play in the development of a multi-hazard DSS.

5.5 Fragility, Vulnerability and Consequence Modeling

In the preceding section, a brief overview of the modeling techniques used for seismic, flood and thermal hazard was given. As described earlier, risk is the intersection of the hazard with a set of exposed and vulnerable societies and the built environment.

Hazard and vulnerability are inextricably linked; risk is null in the absence of vulnerable assets and individuals. Quantifying vulnerability at the urban scale is, as with hazard, inadvertently uncertain and broadly divided into empirical or modeling approaches.

5.6 Seismic Vulnerability

The empirical data which has been used to derive general functions for estimating the seismic demand parameters, fragility, and vulnerability broadly includes:

1. manual post-event surveys:
   - rapid surveys
   - engineering surveys
2. remotely sensed data:
   - satellite imagery

- street-side imagery
- sensor data

3. administrative data:

- governmental data
- insurance data

Recent advances in computational capabilities have seen AI complementing the data generated through remote sensing (e.g., [41]). Post-event damage records offer an opportunity, at the least, of validating numerical models and, at best, making future predictions. High-quality predictions require high-quality databases, which can only be supported by a common framework for data collection and management. More often than not, research produces fragmented databases with variable levels of refinement, ad-hoc damage scales, or absent data. Recent research still cites the fragmentation of such databases [2]. This represents a missed opportunity to leverage such data and support decision makers. Fragmentation appears thus to be a uniform feature.

Without a common framework for the qualification and quantification of post-event damage, significant empirical uncertainties are introduced into the underlying models. Sensed data, that is, through instrumentation, does not suffer from lack of interpretive subjectivity but faces its own limitations, specifically in terms of deployment cost. The Global Earthquake Model Foundation (GEM) has been at the forefront of the most significant effort at general harmonization of seismic risk data and processes in recent years. GEM has undertaken substantial review of vulnerability functions provided by literature [48], where the limitation cited is the lack of consistency among the approaches. The OpenQuake engine [42], an open-source seismic-risk software developed by GEM, provides support for the use of a database of fragility and vulnerability functions [28]. These functions are developed in a fully analytical manner and make no use of empirical data.

5.7 Flood Vulnerability

Conversely to seismic damage, damaging floods are significantly more frequent and thus more data is available. Unlike seismic vulnerability, where greater weight is given to the building (or object) performance, flood vulnerability is more dependent on issues of land-use and topography [49]. Empirical vulnerability curves constitute a valid approach to flood vulnerability [38]. In a building-centric context, floods are more likely to cause damage to contents and finishing; although less likely, structural failures remain plausible [36]. Inundation depth is the most common intensity measure used in flood vulnerability curves, although flow velocity and especially duration related to land-use and the exposed building type. The EU has relatively recently developed a supra-national database of flood damage [20]. The database provides high-level baseline data for estimated economic losses for all global regions in the form of vulnerability curves. It is apparent that some simplifications are necessary for such databases, for example, the only two categories available for South Africa are "small-house", "medium-house" and "large-house". The database is offered as a simple spreadsheet which cannot be easily augmented or enriched by additional data should it be available.

5.8 Heat Vulnerability

The vulnerability of a population to heat risks, much like seismic or flood vulnerability, includes a physical component. This component is largely determined by the characteristics of the built indoor and outdoor environment. For indoor spaces, the key factor is a building's overall thermal performance—its ability to (1) keep heat out and (2) mitigate heat gain through cooling mechanisms [37, 8].

The extent to which heat is excluded depends primarily on material properties, particularly those of the facade, such as thermal conductivity and surface reflectivity. Heat retention is also a crucial factor; for example, thick concrete ceiling slabs absorb heat during the day and gradually release it at night [52, 37]. Insulating materials with low thermal conductivity can serve as effective barriers, but their implementation depends on various factors, including the age of the building stock. The greater the temperature differential between outdoor and indoor environments, the harder passive or active cooling systems must work to maintain comfortable indoor conditions.

The perception of thermal comfort varies significantly in ways other vulnerabilities do not. Physiology, psychology, and cultural factors all influence heat-related risks, particularly those linked to less tangible aspects such as urban livability. Heatwave casualties predominantly include older individuals and those with lower physical fitness. Adaptation and coping mechanisms—such as access to water bodies for cooling, active cooling systems like air conditioning, and passive features like ventilation stacks or operable windows—affect overall heat vulnerability. As climate change intensifies heatwaves in terms of both ambient air temperature and duration, regions historically accustomed to high temperatures, despite having adaptation mechanisms in place, may struggle to keep pace. Meanwhile, temperate climates with little prior need for heat adaptation strategies face heightened vulnerability to increasing heat loads.

6 RESEARCH GAPS

The following research gaps have been identified:

6.1 Lack of a Common Multi-Hazard Domain Model

A well-defined and robust multi-hazard domain model is currently lacking in both research and subsequent digital tools. Existing studies predominantly focus on developing decision support systems for single-hazard contexts. This research posits that the development of a multi-hazard domain model requires more than simply aggregating individual hazard models; rather, it represents an entirely new domain that is "greater than the sum of its parts." Further investigation is necessary to achieve a comparable level of rigor for the multi-hazard and multi-criteria domains. Such a domain model should serve as a comprehensive, well-defined, and interoperable foundation to frame analysis, collaboration, data collection, and processes in the multi-hazard domain.

6.2 Undeveloped Methodologies for Integrating the Multi-Hazard Approach into an SDSS

The development of multi-hazard Spatial Decision Support Systems (SDSS) remains limited (as discussed in Section 5), despite growing interest from prominent disaster risk reduction (DRR) organizations, such as FEMA, with the release of HAZUS 7 [18]. The current approach to the multi-hazard problem primarily involves creating tools capable of calculating risks for individual hazards on a shared platform. However, there is no established SDSS that adopts a truly holistic multi-hazard perspective, which would encompass:

- **Multi-hazard cost-benefit analysis:** Evaluating how to allocate resources between various components of the risk profile.
- **Multi-solution impact analysis:** Modeling and exploring retrofits that address multiple hazards simultaneously.
- **Intervention prioritization plans:** Providing interactive insights and guidance on how to sequence or prioritize interventions.

6.3 Underdeveloped Methodologies for Supporting Assessments with Real-Time Instrument Data

The integration of real-time data from sensing instruments (e.g., accelerometers) into an SDSS remains an underdeveloped area of research. For example, it has been demonstrated that prediction models based on sensor-based observations of building performance are viable for forecasting seismic performance during hazard events (e.g., [51]). However, such work is constrained by the availability of historical datasets, largely due to issues of harmonization. Similar methodologies for flood and thermal hazards remain largely unexplored. Furthermore, there are no digital tools or infrastructure available for the real-time integration and processing of a broad network of multi-hazard building performance observations.

6.4 Lack of Digital Infrastructure for the Generation of Empirical Multi-Hazard Vulnerability Products

The current body of research does not support a long-term approach to collecting, collating, and distributing multi-hazard vulnerability data on a regional or global scale. Existing efforts are hindered by the lack of robust digital infrastructure, leading to delays between the collection of new data and its integration with historical datasets. Open digital infrastructure to facilitate the collection, processing, and dissemination of empirical multi-hazard vulnerability data is unavailable. Existing databases of empirical vulnerability data are:

1. Single-hazard-centric,
2. Fragmented, consisting of multiple data schemas, information models, and levels of granularity,
3. Limited in scope,
4. Focused on damage rather than resilience.

Notably, the limited research efforts that aim to harmonize outputs are primarily analytical and rarely incorporate empirical data at sufficient granularity. This lack of granularity, coupled with an inability to efficiently update datasets, results in significant "up-front" work, hindering the use of such data in novel applications and potentially explaining the scarcity of related research.

6.5 Lack of Statistical Leveraging Methodologies for Historical and Future Vulnerability Data

In the absence of comprehensive, granular, historical databases of empirical building performance and vulnerability, significant statistical leveraging—whether through regression or machine learning (ML) techniques—is currently infeasible or fragmented. Consequently, the potential utility of large, unified datasets remains untapped, largely due to the previously identified gaps.

7 RESEARCH QUESTIONS AND OBJECTIVES

Guided by the research gaps identified, the principal research question proposed by this work is:

> **Research Question**
>
> How can integrated domain driven design facilitate the development of spatial decision support systems (SDSS) for tackling the multi-hazard, multi-domain, and multi-criteria risk problem, while improving the quality of risk assessments and decisions?

This research question is subdivided into the following sub-questions and objectives:

7.1 Sub Question 1 → A Suitable Multi-Hazard Domain Model?

What are the *requirements* and *architectural features* of a novel multi-hazard domain model that successfully navigates the current and potential future needs of a multi-hazard domain?

Building on existing digital standards and harmonization efforts, the underlying domain model that adequately serves the various requirements of the multi-hazard domain will be designed and developed. The components of such a model include data schemas, process models, and technical integration protocols. A successful answer to this sub-question would achieve the following objectives:

- **Objective 1.1** → Develop an ubiquitous ontological model and translate it into a technical model in both written and graphical (UML) formats.
- **Objective 1.2** → Translate the technical domain model into a software library that successfully integrates and builds upon other hazard software and is capable of performing, at a minimum, a multi-hazard risk assessment.
- **Objective 1.3** → Ensure the library complies with existing technical standards, such as OGC and OpenAPI.
- **Objective 1.4** → Leverage existing data sources, formats, and digital infrastructures (e.g., the INSPIRE framework) within the library.

7.2 Sub Question 2 → Improvements in Decision Making?

How does an integrated multi-hazard domain model contribute to the development of a multi-hazard SDSS?

The model forms the foundation on which novel techniques can be developed to 1) improve the quality of multi-hazard risk assessments and 2) encourage more robust decision-making processes. A successful answer to this sub-question would achieve the following objectives:

- **Objective 2.1** → Normalize fragmented historical records of building performance to the multi-hazard domain model and leverage them to create new vulnerability products, such as empirical vulnerability curves and multi-hazard performance prediction models.
- **Objective 2.2** → Integrate vulnerability models at various scales (e.g., building scale) into the SDSS through an interface, enabling them to be automatically scaled up to meso- or macro-scales.
- **Objective 2.3** → Successfully integrate static and real-time data into the SDSS and develop novel methodologies (e.g., as outlined in Appendix A) to address prediction uncertainties.

7.3 Sub Question 3 → Serving, Extensibility, and Future Work

How can web-based technologies enable multi-hazard SDSS products to be served, shared, and support long-term interoperable decision-making?

The long-term collection and maintenance of the multi-hazard domain model, data, and related products require uptake by the research community and decision-makers. Thus, serving both the data and its outcomes must be carefully studied and implemented. The final research objective is to investigate the most suitable approaches toward the extensibility of this work. This involves decisions on interoperability with existing open-source software (e.g., OpenQuake, EnergyPlus) and Geographic Information Systems (GIS) standards and services (e.g., OGC and OSM). The objectives for this sub-question are as follows:

- **Objective 3.1** → Demonstrate, through a testing regime, that the library and SDSS are interoperable across multiple domains.
- **Objective 3.2** → Make the library and SDSS openly available, adhering to the FAIR [54] principles for data management and stewardship.

7.4 Sub Question 4 → Added Value of Research

What benefits does the MHDM have over conventional approaches for assessing single-risk?

The MHDM will provide a set of standardized 'specialist interfaces' which allow the different hazard domains to collaborate and develop their assessments in parallel, effectively creating a 'Common Data Environment'. The effectiveness of the MHDM and systems will be analyzed using qualitative feedback from domain experts.

8 METHODOLOGY

The research methodology is divided into three phases (see Figure 5), each envisaged as a self-contained feedback loop.

8.1 Phase I-A → Development of Multi-Hazard Domain Model

Following the broad literature review of the first year, and while building upon on existing standards (e.g., OGC, OSM), Phase I-A involves the proposition and development of the multi-hazard domain model, tackling the following aspects:

- An ubiquitous (UML) information model for building and city modeling from the multi-hazard and risk lens,
- Database schema designs for geospatial, risk-centric, relational database management systems (RDBMS)
- Process modeling (OGC API compliant) for designing, building, extending and serving multi-hazard geo-risk simulations,
- Methodology for the integration of multi-source datasets, from static to real-time data,
- Interoperability guidelines, technical and methodological connections with other open-source software,

A feedback loop with Phase I-B is expected, where details found in existing datasets contributes, but not necessarily limits, the modeling language. The result of this phase shall culminate in technical documentation and the digital library that allows users to run multi-hazard risk assessments. The digital library at this point is not a fully fledged novel multi-hazard SDSS.

8.2 Phase I-B → Data Procurement, Transformation, and Normalization

The library will be built upon and novel methods of improving assessment capabilities will be built upon it. The domain model will be applied to existing multi-hazard open-source data, in terms of exposure, vulnerability and loss datasets, as well as data from the MultiCare project (see Section 9.3). Depending on the nature of the datasets, this will be achieved either through automated web-scraping processes or through more manual efforts. An abstract normalization pipeline (partly developed, see Appendix B) that will facilitate both initial and future data ingestion processes will be fine-tuned. In this phase, theoretical and technical decisions regarding the handling of incomplete or low-integrity data will be considered. The result of this phase will be an initial pilot databases serving, most notably, vulnerability products but may include existing material normalized to the schema and case-study runs of the library.

The pilot database, duly normalized, will undergo statistical evaluation. The suitable statistical techniques (regressions or ML-assisted methods) will be investigated. Sensor-based data will be used to develop prediction models that provide guidance on building performance. The geospatial correlation between demand parameters will also be investigated and, if possible, leveraged. Prediction models will be constructed at variable levels of detail. For example, a seismic demand parameter model conditioned on only two variables, such as height and fundamental period, will be created. More variables will be incorporated if the data allows, such as structure type and the presence of specific ground conditions. The associated uncertainty will be communicated, with the expectation that more conditioning variables will result in less uncertainty. Building damage records will be treated similarly, with the intention of developing vulnerability curves up to the limit that the records allow. For instance, high-damage states are unlikely to be plausibly derived through empirical data analysis.

8.3 Phase II → Leveraging the Model: Novel Approaches using the Integrated Approach

The domain model and library will be leveraged to develop novel ways of modeling multi-hazard decision making. The following avenues shall be investigated:

1. Leveraging normalized (to the domain model) sets of historical data pertaining to building performance and / or damage when exposed to multi-hazard to develop new vulnerability products (e.g., building performance models, vulnerability curves, baseline loss data). An graphical abstract of showing a pipeline for such a system is shown in 6
2. How to scale up detailed analysis at single-building or component scale to the neighborhood or urban scale and handling the uncertainty embedded in such procedures,
3. How real-time data, measurements and observations can contribute to reducing uncertainty in estimations: e.g., through a simulation-observation feedback loop

8.4 Phase III → Serving, Integration, and Decision Support

The penultimate phase is concerned with the deployment and serving as a fully integrated SDSS. Finally, the library and an OGC-compliant API will be built around the database to both serve and ingest vulnerability data.

The methodology can also be viewed from a digital architectural perspective, where specific layers of the system are developed. Figure 6 presents a high-level, non-technical schema of the envisaged research pipeline, which consists of:

1. Natural hazard events occur and vulnerability data is collected.
2. Vulnerability data may be sensed (remotely or through instruments) or recorded via inspection or photographs; furthermore, both historical and future data are relevant.
3. The data ingestion and transformation layer, governed by the vulnerability schema (Phase I), processes, transforms, and normalizes the data.
4. The datasets are subdivided into relevant archetypes at varying levels of detail.
5. An automated processing layer generates statistical data products in the form of:
   - Demand prediction models (for sensed data),
   - Vulnerability curves (for recorded data).
6. The outputs feed an integration layer, which works with other open-source software.
7. The decision-making layer provides guidance to decision-makers.
8. The serving layer delivers information to the end-user.

8.5 Phase IV → Research Reflection and Analysis of Added Value

The final, reflective phase of the work will set out to establish the benefits and limitations of the MHDM, the SDSS and associated products. Assessing the added value of of the research will be carried out via a number of methodologies:

1. Requirements tracing - The research outcomes will be mapped back to the original requirements defined by the MHDM and SDSS; the ability of other approaches to satisfy the requirements will also be assessed
2. Qualitative / Interview Based Key Performance Indicators (KPI) - A number of user centric KPIs, specifically 1) Task Success Rate, System Usability Scale (SUS) and 3) User Experience Questionnaire will be carried out to assess the benefits and limits of the research and its outcomes,

8.6 Planning

The research project will posses a number of global milestones, and a rolling 3-month look-ahead will map out tasks, sub-tasks and development goals in further detail. For the upcoming 2nd year, two papers are proposed:

1. Journal Paper 1: Sets out to define the Domain Model, the work is a cross between a literature review that considers risk from a software point of view.

Research Planning and Milestones

Task / Period	Work Load	Parent	Description	Dependencies	Risks	Start	End
Journal Paper #1: The Multi-Hazard Domain Model	50% - 80%	Milestone				11 Jan 2025	18 Apr 2025
Introduction Development	50%	Journal Paper #1: The Multi-Hazard Domain Model				10 Feb 2025	21 Feb 2025
Technical Compilation of Domain Knowledge - Paradigms	50%	Journal Paper #1: The Multi-Hazard Domain Model	Continue building upon earlier literature review, organize a set of tables describing all the single-hazard risk paradigms (i.e., the state of the art for each risk).			22 Feb 2025	28 Feb 2025
Single Model Drafts	50%	Journal Paper #1: The Multi-Hazard Domain Model	Propose a model for each risk group in the spirit of Domain Driven Design; these models should fit in well conceptually but also making reference to existing software tools: how is domain knowledge represented in the software?			01 Mar 2025	12 Mar 2025
Multi Hazard Domain Model - Design and Implementation	50%	Journal Paper #1: The Multi-Hazard Domain Model				13 Mar 2025	02 Apr 2025
Wrap Up	20%	Journal Paper #1: The Multi-Hazard Domain Model			Will be quite busy with ISPRS and MultiCare Meetings.	03 Apr 2025	18 Apr 2025
MultiCare Task 7.2 Contribution	50%					11 Jan 2025	31 May 2025
MultiCare Task 9.1 Input	10%		Expected continued input on Task 9.1 (mostly interactions with WP8).			10 Feb 2025	12 Dec 2025
Journal Paper #2: Multi Hazard SDSS			A journal paper coming from my contribution to Multicare (Task 7.2).			31 May 2025	15 Oct 2025
Leverage Phase: Improvements to Risk Domain			Evaluation of the SDSS in what is a 'beta' state, and drawing up decisions (insight from earlier months) on how best to exploit the SDSS to make contributions to the 'risk' domain more broadly.			15 Oct 2025	15 Oct 2026
Finalization Year						15 Oct 2026	15 Oct 2027

Fig. 4. Research methodology, divided into three main phases as described in Section 8.

9 ROADMAP

9.1 Graduate School Planning and Progress

The progress and scheduled graduate school courses may be found in Table 1.

9.2 Publication Planning

A publication plan is presented in Table 2.

9.3 Project Planning

My PhD research is funded by the Horizon Europe `MultiCare` Project. In this project, I am directly involved in two work packages (WP), namely 1) WP7 - Spatial decision support framework and system for multi-hazard resilience analysis at the urban level, and 2) WP9 - Health monitoring of buildings for data-driven prediction and warning systems.

In WP7, my primary task is to integrate multi-hazard resilience plug-ins and algorithms, developed at the subsystem and building levels in earlier work packages, into a proprietary Digital Twin (DT) technology. This integration will focus on enabling dynamic data querying, ensuring system interoperability, and managing large-scale data efficiently, all while maintaining robust data governance practices. Leveraging the baseline model, the DT implementation will incorporate algorithms and plug-ins developed in WP5 and WP6, adapting and extending them to calculate and visualize the multi-hazard resilience index at both pre- and post-resilience improvement stages for buildings.

My role will also involve harmonizing the various datasets and algorithms to ensure interoperability within the DT framework. This requires integrating hazard and vulnerability metrics such as peak ground acceleration or flood susceptibility indices and visualizing outputs effectively through layered information systems. These outputs will be configurable to meet end-user needs, allowing data to be explored at scales ranging from individual buildings to districts or focus zones. My research objectives—focused on harmonization, open data, and compliance with OGC standards—are integral to achieving WP7's goals, addressing the complex challenges of integrating diverse data streams and tools into a cohesive, actionable system for urban resilience planning.

In WP9, my work focuses on developing a real-time data management framework to support the integration of sensor data from building demonstration sites defined in WP6. Building on the requirements for digital services outlined in WP4, this task involves identifying specific sensors to be deployed at these sites to meet the needs of the developed applications and digital services. A critical component of this work is the creation of a standardization framework based on open geospatial standards, such as OGC SensorThings, to ensure efficient integration and communication across heterogeneous real-time data sources.

This standardization effort will streamline the management and retrieval of sensor data, enabling seamless processing and predictive analyses within the digital services ecosystem. My implementation of OGC SensorThings is not only pivotal to achieving interoperability but also contributes to my broader objective of developing a 'vulnerability modeling language.' This language aims to standardize how vulnerability data is represented and analyzed, making it possible to leverage real-time data effectively for resilience and vulnerability assessments. By aligning real-time data management with open standards and harmonized frameworks, this work will directly support WP9's goal of enabling robust, scalable, and interoperable digital services.

Category	Course Code	Course Details	GS Credits	Year
	CSE1210	Probability Theory and Statistics (Q4)	5	2
Discipline	CS4505	Software Architecture (Q1)	5	3
	CSE1500	Web and Database Technology (Q2)	5	3
		Discipline, Total	**15**	
Research	LoJ	TA: Technical/Material Support/Correction	2	1 ✓
	LoJ	TA: assisting in laboratory course / tutorial	3	1 ✓
	LoJ	Writing the first conference paper	1	1 ✓
	R2.B2	Statistics for PhD Research	4	2
	R1.A2	The Informed Researcher - Information and Data Skills	1.5	2
	R1.C4	Advanced Problem Solving and Decision Making	1.5	2
	LoJ	Writing an international, peer-reviewed journal article	2	2
		Research, Total	**15**	
Transferable	T4.G1	PhD Startup Module	2	1 ✓
	T1.C1	Scientific Storytelling	2	2
	T3.B1	Coaching Individual Students	1.5	2
	T3.B1+	Coaching Individual Students + Project Groups	1.5	2
	T1.A7	Data visualisations - A practical approach	1	2
	T3.A2	Small Group Teaching and Lecturing	1	2
	T3.A3	Assessing Students and Master Thesis Projects	1	2
	T1.A9	Scientific Text Processing with LaTeX	1	3
	T1.B2	Writing a Dissertation	3	3
	T2.B2	Cross Cultural Communication Skills in Academia (online)	1	3
		Transferable, Total	**15**	

Tab. 1. Projected modules required to obtain 45 Graduate School (GS) credits. ✓: complete modules or activities.

Tentative Title	Authors	Journal or Conference	Status
An OGC SensorThings GIS Pipeline For Estimating Seismic Engineering Demand Parameters	Justin Schembri, Azarakhsh Rafiee, Peter van Oosterom	ISPRS, GeoSpatial Week 2025	Accepted
Multi-hazard Spatial Decision Support Systems: A domain model	To be determined	Environmental Modeling & Software	Year 2
An OGC compliant GIS information model for City Vulnerability Modeling to Natural Hazards	To be determined	To be determined	Year 2-3
A Harmonized Multi-hazard Vulnerability Database: A Pipeline for Statistical Leveraging of Global Vulnerability Data	To be determined	To be determined	Year 3
Resilience Improvements Under Uncertainty: A Vulnerability Centric Spatial Decision Support System	To be determined	To be determined	Year 3-4
Open Source Tools for Multi-risk Resilience Based Risk Assessments	To be determined	To be determined	Year 4

Tab. 2. Proposed publications and their details.

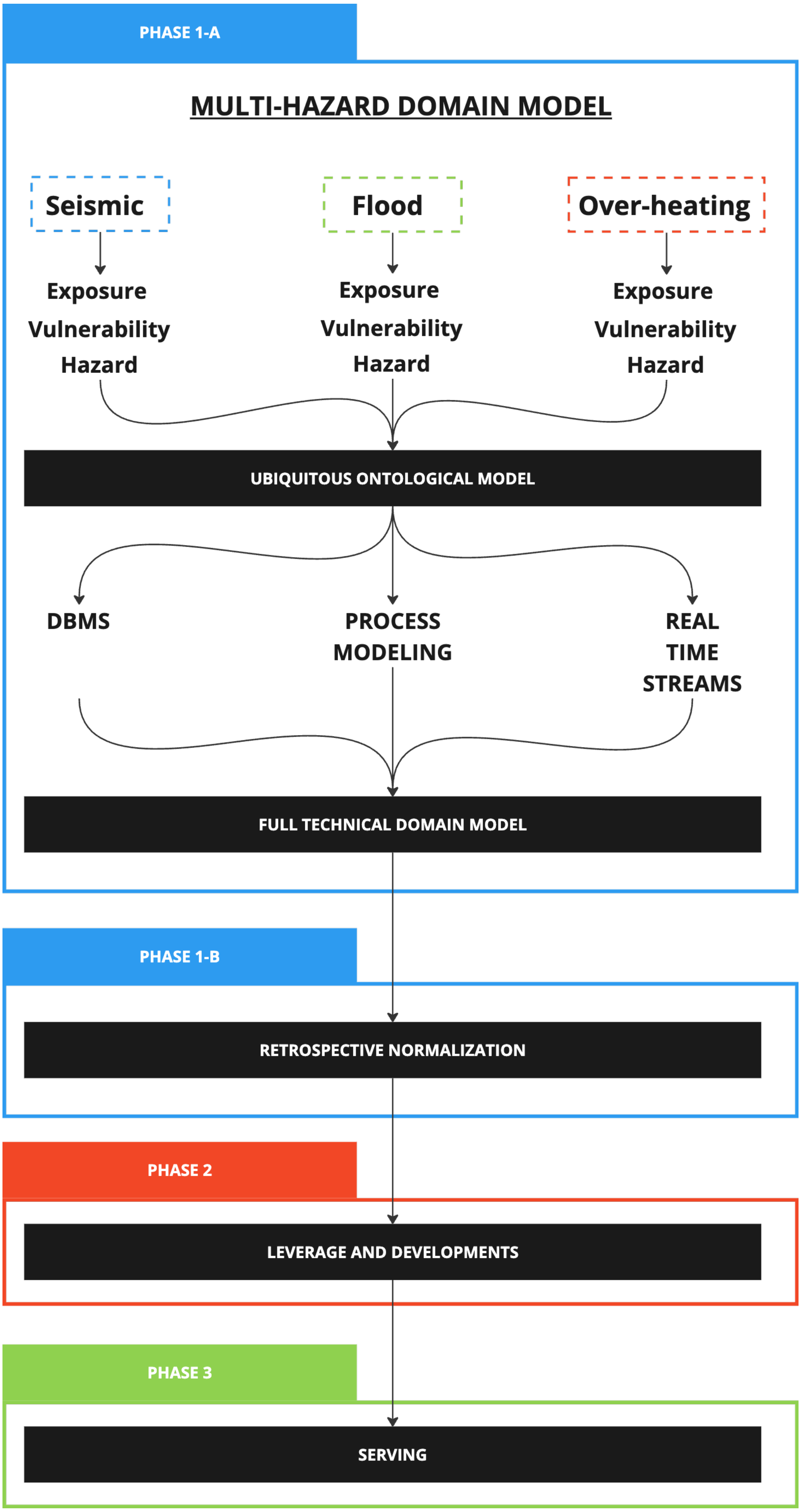

Fig. 5. Research methodology, divided into three main phases as described in Section 8.

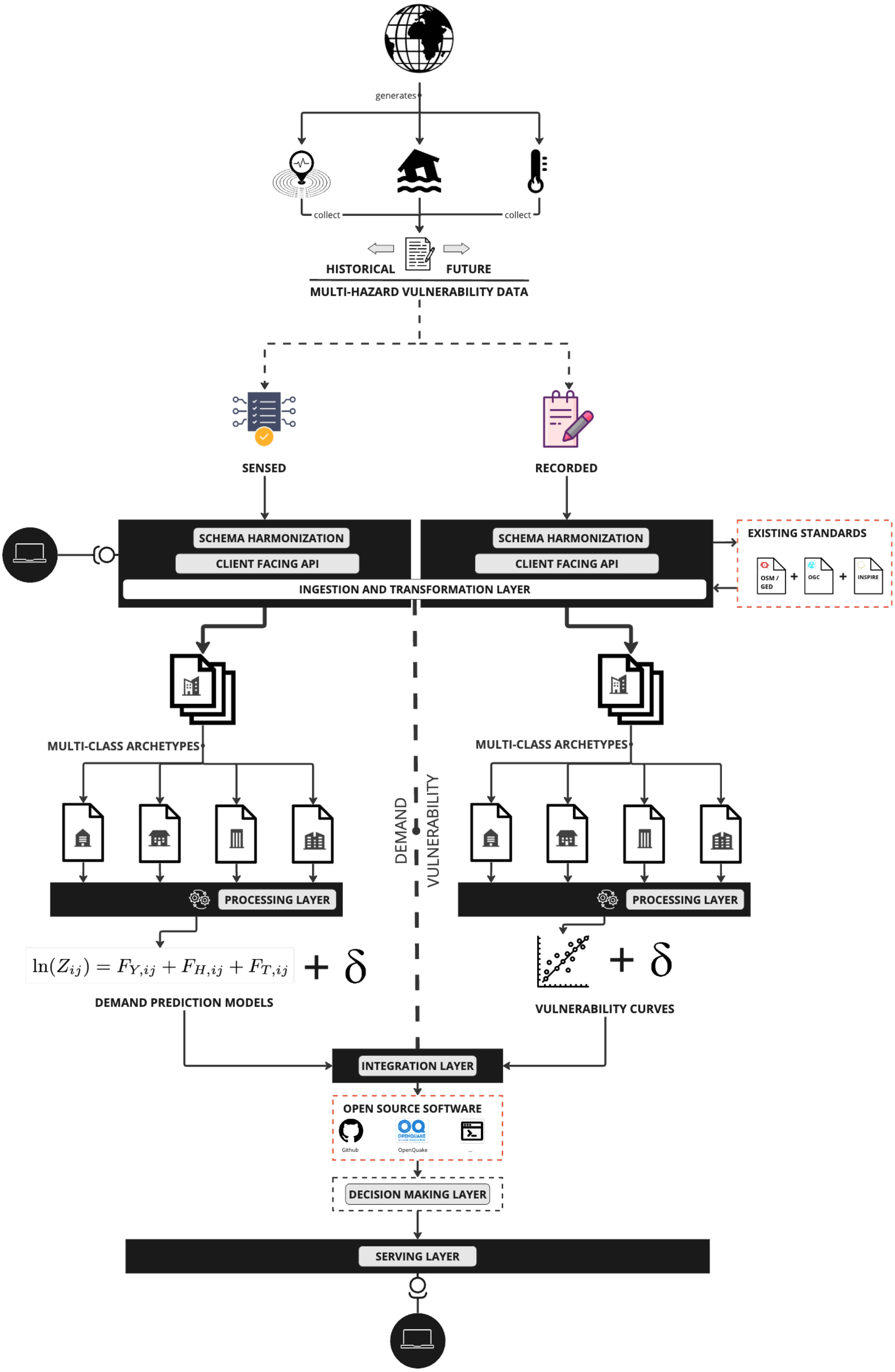

Fig. 6. Schematic pipeline for the digital tools supporting the vulnerability database.

References

[1] Open Geospatial Consotrium (OGC). *Climate and Disaster Resilience Pilot 2024.2: Call for Participation (CFP)*. 2024. URL: https://portal.ogc.org/files/108990 (visited on 01/09/2025).

[2] E. Abdelmalek-Lee et al. "Relational database for building strong motion recordings used for seismic impact assessments". In: *Earthquake Spectra* 39.2 (2023), pp. 1277–1297. DOI: 10.1177/87552930231169306.

[3] Mazen M. Abu-Abdullah et al. "A Flood Risk Management Program of Wadi Baysh Dam on the Downstream Area: An Integration of Hydrologic and Hydraulic Models, Jizan Region, KSA". In: *Sustainability* 12.3 (2020). ISSN: 2071-1050. URL: https://www.mdpi.com/2071-1050/12/3/1069.

[4] Nasrin Aghamohammadi et al. "Heat and Mental Health in Cities". In: *Urban Overheating: Heat Mitigation and the Impact on Health*. Ed. by Nasrin Aghamohammadi and Mat Santamouris. Singapore: Springer Nature Singapore, 2022, pp. 81–107. ISBN: 978-981-19-4707-0. DOI: 10.1007/978-981-19-4707-0_4. URL: https://doi.org/10.1007/978-981-19-4707-0_4.

[5] R. Ara Begum et al. "Point of Departure and Key Concepts". In: *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by H.-O. Pörtner et al. Cambridge, UK and New York, NY, USA: Cambridge University Press, 2022, pp. 121–196. DOI: 10.1017/9781009325844.003.

[6] Peng Bai et al. "Comparison of performance of twelve monthly water balance models in different climatic catchments of China". In: *Journal of Hydrology* 529 (2015), pp. 1030–1040. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2015.09.015. URL: https://www.sciencedirect.com/science/article/pii/S0022169415006988.

[7] Ye Bai, Zhenxing Zhang, and Weiguo Zhao. "Assessing the impact of climate change on flood events using HEC-HMS and CMIP5". In: *Water, Air, & Soil Pollution* 230.6 (2019), p. 119.

[8] Norhan Bayomi and John E Fernandez. "Quantification of heat vulnerability using system dynamics". In: *Frontiers in Built Environment* 9 (2023), p. 1025480.

[9] D. M. Boore et al. "NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes". In: *Earthquake Spectra* 30.3 (2014), pp. 1057–1085. DOI: 10.1193/070113EQS184M.

[10] David M Boore et al. "NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes". In: *Earthquake Spectra* 30.3 (2014), pp. 1057–1085.

[11] P. Buckle, G. Mars, and S. Smale. "New Approaches to Assessing Vulnerability and Resilience". In: *The Australian Journal of Emergency Management* 15.2 (2000), pp. 8–14. URL: https://search.informit.org/doi/10.3316/ielapa.369155833780624.

[12] M. Caruso, R. Pinho, F. Bianchi, et al. "Integrated economic and environmental building classification and optimal seismic vulnerability/energy efficiency retrofitting". In: *Bulletin of Earthquake Engineering* 19 (2021), pp. 3627–3670. DOI: 10.1007/s10518-021-01101-4.

[13] Eric Chaumillon et al. "Storm-induced marine flooding: Lessons from a multidisciplinary approach". In: *Earth-Science Reviews* 165 (2017), pp. 151–184. ISSN: 0012-8252. DOI: https://doi.org/10.1016/j.earscirev.2016.12.005. URL: https://www.sciencedirect.com/science/article/pii/S0012825216304597.

[14] C. Allin Cornell. "Engineering seismic risk analysis". In: *Bulletin of the Seismological Society of America* 58.5 (Oct. 1968), pp. 1583–1606. ISSN: 0037-1106. DOI: 10.1785/BSSA0580051583. eprint: https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/58/5/1583/5350226/bssa0580051583.pdf. URL: https://doi.org/10.1785/BSSA0580051583.

[15] John Douglas and Benjamin Edwards. "Recent and future developments in earthquake ground motion estimation". In: *Earth-Science Reviews* 160 (2016), pp. 203–219.

[16] B. El Kechebour and H. Zeloum. "Choice between Retrofitting and Reconstruction of Buildings in Reinforced Concrete after an Earthquake". In: *Advanced Materials Research* 1119 (2015), pp. 736–741. DOI: 10.4028/www.scientific.net/amr.1119.736.

[17] Tashina Esteves et al. "Multi-Scale Vulnerability Assessment for Adaptation Planning". In: *Current Science* 110.7 (2016). Accessed: 2024-11-08, pp. 1225–1239. URL: http://www.jstor.org/stable/24908012.

[18] FEMA. *FEMA Flood Map Service Center — Hazus 7.0*. 2024. URL: https://msc.fema.gov/portal/resources/hazus (visited on 01/14/2025).

[19] Neil S Grigg. “Comprehensive flood risk assessment: State of the practice”. In: *Hydrology* 10.2 (2023), p. 46.

[20] J. Huizinga, H. De Moel, and W. Szewczyk. *Global flood depth-damage functions: Methodology and the database with guidelines*. EUR 28552 EN JRC105688. Luxembourg: Publications Office of the European Union, 2017. ISBN: 978-92-79-67781-6. DOI: 10.2760/16510.

[21] Gisele Icyimpaye, Chérifa Abdelbaki, and Khaldoon A Mourad. “Hydrological and hydraulic model for flood forecasting in Rwanda”. In: *Modeling Earth Systems and Environment* 8.1 (2022), pp. 1179–1189.

[22] D. E. Kim, P. Gourbesville, and S. Y. Liong. “Overcoming data scarcity in flood hazard assessment using remote sensing and artificial neural network”. In: *Smart Water* 4.2 (2019). DOI: 10.1186/s40713-018-0014-5. URL: https://doi-org.tudelft.idm.oclc.org/10.1186/s40713-018-0014-5.

[23] Vijendra Kumar et al. “Comprehensive Overview of Flood Modeling Approaches: A Review of Recent Advances”. In: *Hydrology* 10.7 (2023). ISSN: 2306-5338. DOI: 10.3390/hydrology10070141. URL: https://www.mdpi.com/2306-5338/10/7/141.

[24] Evan R. Kuras et al. “Opportunities and Challenges for Personal Heat Exposure Research”. In: *Environmental Health Perspectives* 125.8 (2017), p. 085001. DOI: 10.1289/EHP556. URL: https://ehp.niehs.nih.gov/doi/abs/10.1289/EHP556.

[25] Shaofeng Liu et al. “Integration of decision support systems to improve decision support performance”. In: *Knowledge and Information Systems* 22 (2010), pp. 261–286.

[26] Kevin J Lomas and Stephen M Porritt. *Overheating in buildings: lessons from research*. 2017.

[27] M. B. Malgwi, S. Fuchs, and M. Keiler. “A generic physical vulnerability model for floods: review and concept for data-scarce regions”. In: *Natural Hazards and Earth System Sciences* 20 (2020), pp. 2067–2090. DOI: 10.5194/nhess-20-2067-2020. URL: https://doi.org/10.5194/nhess-20-2067-2020.

[28] L. Martins and V. Silva. *Global Vulnerability Model of the GEM Foundation*. 2023. DOI: 10.5281/zenodo.8391742. URL: https://github.com/gem/global_vulnerability_model/.

[29] Anna Mavrogianni et al. “Urban Overheating and Impact on Health: An Introduction”. In: *Urban Overheating: Heat Mitigation and the Impact on Health*. Ed. by Nasrin Aghamohammadi and Mat Santamouris. Singapore: Springer Nature Singapore, 2022, pp. 1–20. ISBN: 978-981-19-4707-0. DOI: 10.1007/978-981-19-4707-0_1. URL: https://doi.org/10.1007/978-981-19-4707-0_1.

[30] John Ryan McGreevy and Elisson Adrien. “Second impact syndrome: The influence of climate change and increased disaster frequency on livelihoods and adaptive capacity in rural Haiti”. In: *International Journal of Disaster Risk Reduction* 85 (2023), p. 103509. ISSN: 2212-4209. DOI: 10.1016/j.ijdrr.2022.103509.

[31] Bruno Merz et al. “Causes, impacts and patterns of disastrous river floods”. In: *Nature Reviews Earth Environment* 2.9 (Sept. 2021), pp. 592–609. ISSN: 2662-138X. DOI: 10.1038/s43017-021-00195-3. URL: https://doi.org/10.1038/s43017-021-00195-3.

[32] Caroline Michellier et al. “Evaluating population vulnerability to volcanic risk in a data scarcity context: The case of Goma city, Virunga volcanic province (DRCongo)”. In: *International Journal of Disaster Risk Reduction* 45 (2020), p. 101460. ISSN: 2212-4209. DOI: 10.1016/j.ijdrr.2019.101460.

[33] Ariane Middel and E. Scott Krayenhoff. “Micrometeorological determinants of pedestrian thermal exposure during record-breaking heat in Tempe, Arizona: Introducing the MaRTy observational platform”. In: *Science of The Total Environment* 687 (2019), pp. 137–151. ISSN: 0048-9697. DOI: https://doi.org/10.1016/j.scitotenv.2019.06.085. URL: https://www.sciencedirect.com/science/article/pii/S0048969719326531.

[34] Massimo Migliorini et al. “Data interoperability for disaster risk reduction in Europe”. In: *Disaster Prevention and Management: An International Journal* 28.6 (2019), pp. 804–816.

[35] Francesco Mulargia, Philip B Stark, and Robert J Geller. “Why is probabilistic seismic hazard analysis (PSHA) still used?” In: *Physics of the Earth and Planetary Interiors* 264 (2017), pp. 63–75.

[36] Norberto C. Nadal et al. “Building Damage due to Riverine and Coastal Floods”. In: *Journal of Water Resources Planning and Management* 136.3 (2010), pp. 327–336. DOI: 10.1061/(ASCE)WR.1943-5452.0000036. eprint: https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%29WR.1943-5452.0000036. URL: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0000036.

[37] Matthew J Nahlik et al. “Building thermal performance, extreme heat, and climate change”. In: *Journal of Infrastructure Systems* 23.3 (2017), p. 04016043.

[38] Hajar Nasiri, Mohd Johari Mohd Yusof, and Thamer Ahmad Mohammad Ali. "An overview to flood vulnerability assessment methods". In: *Sustainable Water Resources Management* 2.3 (2016), pp. 331–336. ISSN: 2363-5045. DOI: 10.1007/s40899-016-0051-x. URL: https://doi.org/10.1007/s40899-016-0051-x.

[39] Negin Nazarian et al. "Integrated assessment of urban overheating impacts on human life". In: *Earth's Future* 10 (2022), e2022EF002682. DOI: 10.1029/2022EF002682. URL: https://doi.org/10.1029/2022EF002682.

[40] Jeffrey Peter Newman et al. "Review of literature on decision support systems for natural hazard risk reduction: Current status and future research directions". In: *Environmental Modelling & Software* 96 (2017), pp. 378–409.

[41] Peter Damilola Ogunjinmi et al. "Rapid Post-Earthquake Structural Damage Assessment Using Convolutional Neural Networks and Transfer Learning". In: *Sensors* 22.9 (2022). ISSN: 1424-8220. DOI: 10.3390/s22093471. URL: https://www.mdpi.com/1424-8220/22/9/3471.

[42] M. Pagani et al. "OpenQuake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model". In: *Seismological Research Letters* 85.3 (2014), pp. 692–702. DOI: 10.1785/0220130087.

[43] A. Palla et al. "Pluvial flooding in urban areas: the role of surface drainage efficiency". In: *Journal of Flood Risk Management* 11.S2 (2018), S663–S676. DOI: https://doi.org/10.1111/jfr3.12246. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jfr3.12246.

[44] European Parliament and Council. *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (Inspire)*. Official Journal of the European Union. OJ L 108, 25.4.2007. 2007.

[45] Ryan Paulik et al. "RiskScape: A flexible multi-hazard risk modelling engine". In: *Natural Hazards* 119.2 (2023), pp. 1073–1090.

[46] Carolina M. Rodriguez and Marta D'Alessandro. "Indoor thermal comfort review: The tropics as the next frontier". In: *Urban Climate* 29 (2019), p. 100488. ISSN: 2212-0955. DOI: https://doi.org/10.1016/j.uclim.2019.100488. URL: https://www.sciencedirect.com/science/article/pii/S2212095519300744.

[47] T. Rossetto and A. Elnashai. "Derivation of vulnerability functions for European-type RC structures based on observational data". In: *Earthquake Engineering Structural Dynamics* 32.11 (2003). Received: 11 April 2002; Revised: 21 February 2003; Accepted: 24 February 2003, pp. 1647–1666. DOI: 10.1002/eqe.313.

[48] Tiziana Rossetto, Ioanna Ioannou, and David N. Grant. *Existing Empirical Fragility and Vulnerability Functions: Compendium and Guide for Selection*. Technical Report GEM.VULNSMOD.TR2015.01. Global Earthquake Model Foundation, 2015. DOI: 10.13117/GEM.VULNSMOD.TR2015.01. URL: https://www.researchgate.net/publication/282850853.

[49] M. C. de Ruiter et al. "Review Article: A comparison of flood and earthquake vulnerability assessment indicators". In: *Natural Hazards and Earth System Sciences* 17.7 (2017), pp. 1231–1251. DOI: 10.5194/nhess-17-1231-2017. URL: https://nhess.copernicus.org/articles/17/1231/2017/.

[50] Philip J Schneider and Barbara A Schauer. "HAZUS—Its development and its future". In: *Natural Hazards Review* 7.2 (2006), pp. 40–44.

[51] H. Sun et al. "Development of a Generalized Cross-Building Structural Response Reconstruction Model Using Strong Motion Data". In: *Journal of Structural Engineering* 148 (2022), p. 04022053. DOI: 10.1061/(ASCE)ST.1943-541X.0003336.

[52] MGM Van Der Heijden, BJE Blocken, and JLM Hensen. "Heat wave vulnerability classification of residential buildings". In: *conference; 7th Windsor Conference; 2012-04-12; 2012-04-15*. NCEUB. 2012, pp. 1–6.

[53] W. Viessman and G.L. Lewis. *Introduction to Hydrology*. Prentice Hall, 2003. ISBN: 9780673993373. URL: https://books.google.com.mt/books?id=pSOsQgAACAAJ.

[54] Mark D Wilkinson et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3 (2016). Erratum in: Sci Data. 2019 Mar 19;6(1):6. doi: 10.1038/s41597-019-0009-6, p. 160018. DOI: 10.1038/sdata.2016.18.

[55] Yong-Fei Zheng. "Plate tectonics in the twenty-first century". In: *Science China Earth Sciences* 66.1 (2023), pp. 1–40.

[56] Decheng Zhou et al. "Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives". In: *Remote Sensing* 11.1 (2019). ISSN: 2072-4292. DOI: 10.3390/rs11010048. URL: https://www.mdpi.com/2072-4292/11/1/48.

Appendix A: Proof of Writing

The paper annexed in this section was accepted for publication in the International Society for Remote Sensing and Photogrammetry Annals.

An OGC SensorThings GIS Pipeline For Estimating Seismic Engineering Demand Parameters

Justin Schembri [1], Azarakhsh Rafiee [2], Peter van Oosterom[2]

[1]Faculty of Architecture and the Built Environment, Delft University of Technology, The Netherlands - j.schembri@tudelft.nl
[2] Faculty of Architecture and the Built Environment, Delft University of Technology, The Netherlands

Keywords: Seismic Sensors, Structural Response, OGC SensorThings, Seismic Data Harmonization, Structural-response Prediction Model, Multi-source Seismic Observations

Abstract

Estimating the losses in the immediate aftermath of an earthquake is a key component of seismic response. Seismic rapid-loss estimates provide first responders with a prediction of where and what to prepare for. Improving the precision of quick loss estimates requires an estimate of how a buildings in the affected zone may have reacted to an event. Structural response prediction models are a novel approach to estimating building response from the observed displacement of instrumented buildings. Current SRPMs are built on relatively small databases but offer potential for expansion. There exists no robust building-specific database which could facilitate the construction of these models. As a reaction to this gap, this study applies, abstractly and concretely, the OGC SensorThings data model to building seismograph records. The harmonized records form part of a proposed abstract and concrete Structural Response Prediction Model to make estimates of building-response on other un-instrumented buildings. The utility of a abstracted observation data-model and pipeline is shown, with the potential for unifying existing data-sources. The work shall show that the OGC SensorThings integrates generally well, with some limitations, with the requirements of seismic observation record keeping.

1. Introduction

Earthquake hazard is a regional risk experienced by societies located near seismic faults. Estimating the risk posed by strong ground motion is an ongoing global effort. The key components in quantifying seismic risk are 1) hazard, i.e., the intensity of ground motion that a geological fault could plausibly generate, 2) the exposure, i.e., the amount, type, and distribution of buildings exposed to the hazard, and 3) the vulnerability of those exposed elements to ground motion. Risk assessments provide decision-making support required for long-term planning, guiding retrofit and preparation strategies (e.g., Probabilistic Seismic Hazard Assessments, PHSA, e.g., Baker et al., 2021). Rapid loss estimates, conversely, (e.g., Erdik et al., 2011) are near real-time post-event assessments of the potential damage to the built environment and inhabitants. Rapid loss estimates are of critical use to first responders, providing insight into the immediate needs and priorities of response.

Extensive historical observations of earthquake ground motion intensity (e.g., Peak Ground Acceleration, PGA, units: g), collected by free-field seismic stations, form the baseline datasets used in quantifying seismic hazard. Ground Motion Prediction Equations (GMPE, e.g., Boore et al. 2014) are predictive models, fitted by multi-stage regressions to historical records. Such models are conditioned by several variables, e.g., event magnitude, soil conditions, and site-to-source distance.

GMPEs estimate the geospatial distribution of ground motion caused by an event, but they do not explicitly describe the damage to buildings. The intensity measure (IM) of ground motion must then be translated to the building's expected response, which correlates with damage levels. A building's structural response, or Engineering Demand Parameter (EDP), to a given level of IM is dependent on several factors, e.g., its height and structural properties. Structural response may, similarly to ground motion, be captured by accelerometers installed within the building envelope at various levels. There has been significant effort in forming regional or global networks of seismograph records (e.g., Archuleta et al., 2006); however, such networks tend to be geared towards geological aspects rather than engineering aspects. To our knowledge, there is no "building-focused" database that could facilitate targeted analysis of building response and, furthermore, the construction of structural-response prediction models (SRPM). SRPMs are a fairly novel (Sun et al., 2022) proposal made by [author's name], on whose work we build, which uses historical data of EDP to fit coefficients and conditioning parameters. Furthermore, immediately following an event, buildings that are instrumented could provide insight into the accuracy and precision of the SRPM predictions. The Cross Building Reconstruction Response model (CBRR) proposed by Sun et al. (2022) measures the over- and underestimations of EDP observed at instrumented sites and spatially interpolates (e.g., via kriging) and assigns it to uninstrumented buildings, thus providing more accurate EDP predictions and, subsequently, better rapid-loss estimates.

Thus, a gap exists (e.g., Abdelmalek-Lee et al., 2023) in the harmonization of fragmented building-response records from global sources, hindering the construction of robust SRPMs. Presently, databases are geared toward geological aspects rather than the specific responses of buildings, underscoring the need for a structured approach to aggregate and utilizing building-focused observations in creating the foundational SRPM. Moreover, once a model is established, observations from subsequent seismic events must flow into a processing pipeline, such as the CBRR, to refine rapid-loss estimations. A standardized approach to data modeling would not only benefit the aggregation of existing data for building more robust SRPMs but also support the development of CBRR pipeline tools and software for consistent and open applications. This paper proposes the OGC SensorThings model as a candidate for such harmonization, as described below.

Firstly, both the baseline SRPM and the rapid-loss CBRR consist of components that depend heavily on geospatial data, often derived from geo-sources that may already comply with OGC standards. Standardizing SRPMs using historical data and applying the model after an event can therefore be facilitated by OGC-compliant formats, providing greater interoperability and accessibility. SensorThings, as a neutral, lightweight format focused on geospatial and IoT integration, aligns well with these goals. While seismic data generally contains extensive metadata, SRPMs primarily need only a few essential features—specifically, the maximum EDP experienced by a building. Thus, the raw observational data remains the remit of a seismological network, while the mapping of key processed data relevant to the SRPM and CBRR can be passed to SensorThings. Additionally, the IoT-centric design of SensorThings aligns well with the rapid-loss estimation processes, which require fast, automated processing.

In this study, we address the need for a standardized, building-focused approach to handling structural response data for earthquake risk assessment. We propose an abstract interface designed to ingest, transform, and map building-response observations and metadata into the OGC SensorThings framework. Additionally, we outline an abstract processing pipeline for leveraging these SensorThings objects to generate rapid estimates of EDP following seismic events and support real-time application in the CBRR framework. This standardization enables the geospatial SRPM and CBRR simulation pipeline to predict the structural responses for uninstrumented buildings after an earthquake. The study is organized as follows: Section 2 details the approach for mapping observations to the OGC SensorThings Data Model. Section 3 describes the prediction pipeline, illustrating how the abstraction of observation inputs and model components can enhance the methodology. Section 4 provides the concrete implementation of these proposals, with results and discussion presented in Section 5. Conclusions are presented in Section 6.

2. Building Response Mapping to SensorThings

Processed building accelerometer records may consist, broadly, of three components: 1) station data, 2) event data, and 3) waveform measurements. Station data provides information about the sensor's location and specifics (e.g. sampling rate), while event data includes details such as the event magnitude. Finally, the waveform measurements are corrected observations themselves.

While processed seismological records have no universally accepted domain model and encoding formats such as SEED (Ringler and Evans, 2015), SAC (Seismic Analysis Code, Helffrich et al., 2013), and ASDF (Adaptable Seismic Data Format, Krischer et al., 2016) are prevalent, and SEED is considered a de facto standard in some cases. Some formats are region-agnostic, while others were developed by regional seismological networks such as the COSMOS V1.2 (Archuleta et al., 2006). Some formats use binary encoding (e.g., SEED and ASDF), while others are human-readable ASCII formats (e.g., SAC, COSMOS V1.2). Some records separate station and event metadata from the waveform data, while others do not. The content across standards is, of course, relatively similar, and metadata tends to be extensive; the COSMOS V1.2 format allows for up to 100 lines of headers.

SensorThings, by contrast, is a relatively lightweight and neutral information-model. The application of generic models to domain specific records has potential drawbacks such as granularity loss, where multiple metadata elements which were separate in the original records are lumped together into a vague model attribute such as "properties". However, since the SRPM as introduced earlier does not require extensive metadata, we deem such losses acceptable.

The SensorThings schema consists of eight entities: Datastream, Thing, Location, Historical Location, Sensor, ObservedProperty, Observation, and FeatureOfInterest (Liang et al., 2024). Applying the schema to the real world (see Figure 2) instrumentation setups results in the following descriptive mapping: A building (Location) comprises multiple levels (Things) observed by one or more instruments (Sensors), each having multiple channels (Datastreams) observing acceleration or displacement (ObservedProperty), generating a waveform (Observations) for a given event (FeatureOfInterest). After reviewing the data in the standards described earlier, the following sub-categorization was established: 1) Event Data, 2) Location Data, 3) Record Information, 4) Sensor Metadata, 5) Station Data, 6) Waveform Observations. Some examples of data or metadata for each category included earthquake magnitude and depth (event data), record IDs, processing dates, and station numbers (record information), geographic coordinates of the station and sensor locations within the building (location data), and sample rate (sensor metadata).

A further granular examination of the data and metadata in the standards was used to construct a generic mapping protocol (seismic records to SensorThings) as tabulated in Table 1 and shown in Figure 1.

Seismograph Header	**SensorThings Entity**
Earthquake trigger time	Datastream. phenomenonTime
Earthquake name / reference	FeatureOfInterest. name
All other event-specific information	FeatureOfInterest. properties
Station or building name	Location. name
Station or building coordinates	Location. location
Instrument location	Thing. name
Station number or ID	Sensor. name
Non-metadata information	Sensor. properties
All record information	Datastream. properties
Sensor metadata	Sensor. metadata
Observation units	ObservedProperty
Observations result time	Observation

Table 1. Mapping Event, Location, Station, Record, and Sensor Data to SensorThings schema.

3. Generic Building Response Models

Observations mapped to SensorThings can support two key processes. The first involves leveraging historical observations to construct an SRPM by fitting a regression model. While beyond this study's scope, harmonization through SensorThings, as discussed earlier, could facilitate the expansion of data used in such a regression. The second, within-scope pipeline involves using SensorThings observations, and existing GMPEs,

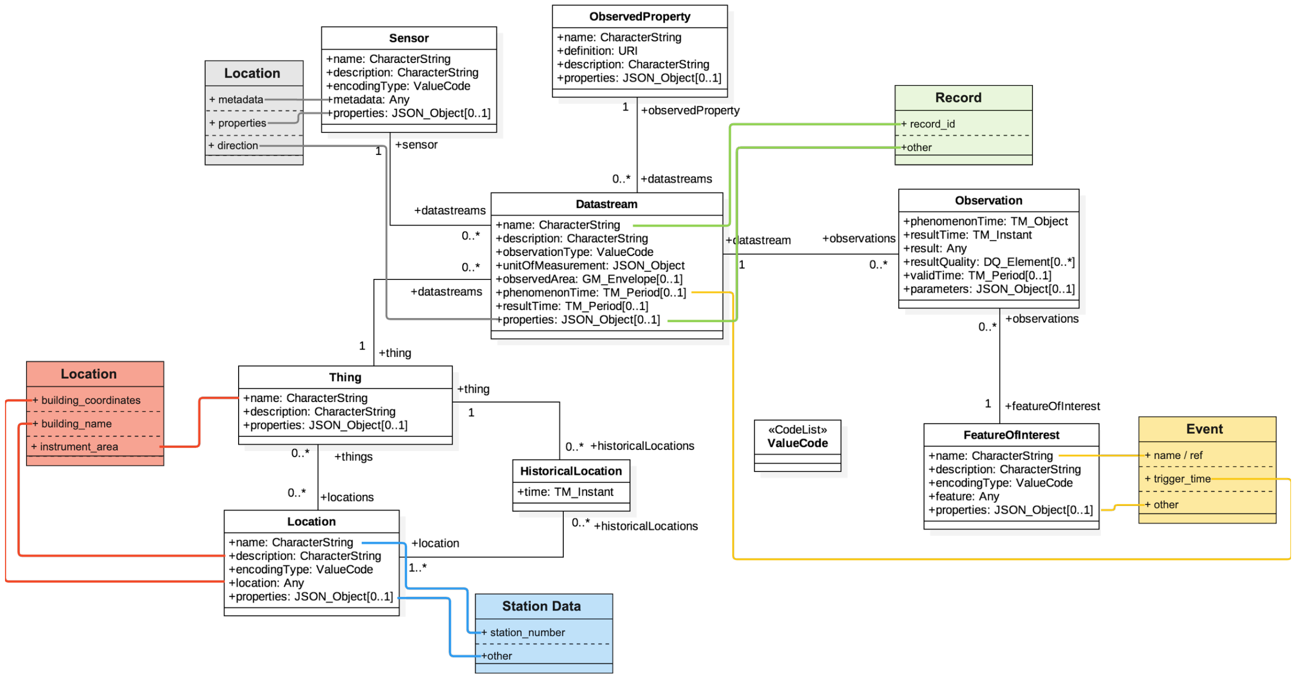

Figure 1. The OGC SensorThings data-model (white boxes) augmented to include the proposed mappings (colored boxes).

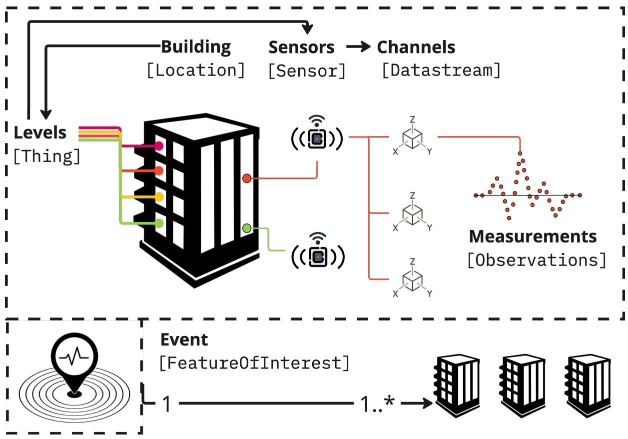

Figure 2. Relationship between real-world instrumentation set up and SensorThings schema. Fixed width text are the equivalent SensorThings entities.

SRPMs, and CBRRs to make predictions of EDP for uninstrumented buildings. This pipeline begins with the estimation of IM following an event. IM is given by, for example, Peak Spectral Acceleration (PSA, in units of g), and its intensity decays over distance. The IM function, F_Y, is calculated by a GMPE whose functional form is represented as:

$$\ln Y = F_M + F_P + F_S + \epsilon\delta \tag{1}$$

Where F_M, F_P and F_S are conditioning functions of event magnitude, event path (approximately distance) and site ground conditions. $\epsilon\delta$ are normalized model residuals.

Next, the IM is transformed to an EDP, such as Peak *Floor* Acceleration (PFA, units, g) through the SRPM, whose functional form Sun et al. (2022) is given by

$$\ln(Z_{ij}) = F_{Y,ij} + F_{H,ij} + F_{T,ij} + \delta W_{ij}^Z \tag{2}$$

Where, for event i and site j, $\ln(Z_{ij})$ is the EDP, Z, in natural log units, $F_{Y,ij}$, $F_{H,ij}$, and $F_{T,ij}$ are functions dependent on the IM, Y, a building's height H, and its fundamental period T, respectively. The fundamental period (units: *seconds*) is a property of a building describing its vibrational characteristics and correlates with its height and structure type. In the above equation, δW^Z is the difference between the SRPM prediction and the observed value for event i at site j. Observed values of EDP are collected via instrumentation for a small subset of buildings. The SRPM residuals, δW_{ij}^Z, are passed to the CBRR, which spatially interpolates the residual via the geostatistical kriging technique.

We build upon the work of Sun et al. (2022) by developing an abstract Python-based GIS pipeline that can reliably make rapid predictions of EDPs by using post-event records from instrumented buildings. The GIS pipeline proposed follows the Object-Oriented Programming paradigm (Wegner, 2003), thus components of the pipeline are described as "classes" or "objects," which have properties referred to as "attributes." Classes described in this section should be considered abstract base classes (ABCs), which enforce a number of internal functions (i.e., methods) and attributes, which any concrete implementation must adhere to.

Firstly, a GMPE ABC shall be responsible for calculating the IM (Y) at a given site i for a given building fundamental period. Several GMPE models have been developed. These possess common attributes as those described in (1), namely magnitude, distance to the site of interest from the event, and the period at which to estimate the IM. Magnitude is a property of the event, while the period is a property of a building, and the distance between them requires coordinates of both event and building.

Thus, the GMPE ABC must be provided with two classes, one representing the event and the other a building.

The representation of an event is achieved through a SeismicEvent base class. To satisfy the GMPE components, a SeismicEvent object must include a magnitude and epi- or hypocenter coordinates. Additional information about the event, such as depth, or fault type, can be provided. Some GMPEs require such additional information, but not universally; thus, other available parameters may be passed as optional keyword arguments to a SeismicEvent object.

The next object required by the GMPE is a Building ABC, which must, at a minimum, include coordinates and the fundamental period. The fundamental period may be estimated at varying levels of detail. In urban-level assessments, it is often calculated using simplified methodologies, which typically require only basic building features, such as overall height and structural system. Thus, if the fundamental period of a Building is not known, a function for calculating it can be provided, and the properties required by the function (such as height) can be stored in an attribute containing the building's SeismicProperties. Having satisfied the GMPE's requirements with the event and building classes, three additional ABCs are required for each of the functional terms, namely the EventTerm, PathTerm, and SiteTerm. These terms need only handle the arithmetic of each functional term.

With these foundational components established, the SRPM ABC then becomes responsible for calculating the EDP (Z). SRPMs are substantially more novel in comparison to GMPEs, and the formulation in (2) is, to our knowledge, the only available model. This formulation requires only building height and fundamental period, which it may inherit from the GMPE class. Thus, the SRPM needs only to implement a method to handle arithmetic. Future SRPMs may incorporate more than just building height and fundamental period. However, since the SRPM is building-centric, any additional parameters may be passed to the Building class' SeismicProperties container. It is unlikely that a formulation of an SRPM would not include IM, height, or fundamental period, as these are known to heavily correlate with EDP.

Once the SRPM generates EDP estimations for all sites, these outputs serve as essential inputs to the CBRR, which incorporates residuals based on observed EDP values. As outlined previously, SensorThings provides a standardized mapping regime for such observations. Therefore, the CBRR ABC will draw from both SRPM results and SensorThings objects corresponding to instrumented buildings. Key components from SensorThings include the building's coordinates (`Location.location`) and observed EDP values. When maximum EDP values are included in the metadata, they can be accessed through `Datastream.properties`; alternatively, they may be obtained by directly processing Observations. The CBRR implements a kriging process tailored to the data distribution, enabling accurate rapid-loss estimates for unmonitored structures.

Thus, the full pipeline, modelled in Figure 3 consisting of the the central GMPE, SRPM and CBRR classes executes the following processes:

- Calculate the IM for instrumented and un-instrumented buildings, using the GMPE,
- Calculate the median EDP for instrumented and un-instrumented buildings, using the SRPM,
- Normalize the observation records,
- Query the SensorThings object, extract the observed EDP at instrumented sites,
- Spatially interpolate the residuals at unmonitored sites,
- Add the residuals to the un-instrumented median predictions, return the total predicted EDP

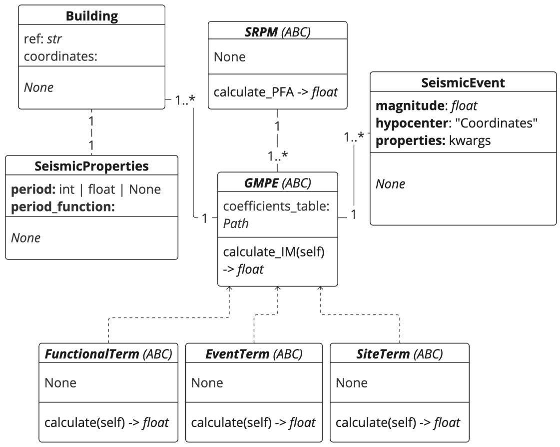

Figure 3. UML diagram showing relationship between Building, SeismicEvent, GMPE, its terms and the SRPM.

4. Case Study Implementation

A case study was developed using the widely used GMPE, BSSA13 (Boore et al. 2014). Additionally, implementations of the SRPM and CBRR from Sun et al. (2022) were utilized, which were fitted to historical building response records from the CSMED database (California Geological Survey and U.S. Geological Survey, 2005) for buildings in California. The CSMED records were also used in this case study.

The case study's source code, developed in the Python programming language, is openly available at `https://github.com/justinschembri/isprs`. The case study involved the following: 1) mapping the observation records to the SensorThings Model, 2) implementing concrete classes for the GMPE, SRPM, and CBRR, and 3) wrapping the process in a Predictor class that makes estimates of expected EDP for unknown buildings. We demonstrate the predictive functionality of the pipeline by generating EDPs for buildings in a past earthquake that occurred in California.

CSMED to SensorThings

The CSMED database provides response records in the COSMOS V1.2 format, at free-field stations and buildings. Records were downloaded for building stations ranging from 1984 to 2018. Each building's record contains a series of ASCII files, divided on a channel-by-channel basis (e.g., CHAN001.V2, CHAN002.V2, etc.). These channels correspond to an instrument at a given floor (e.g., 1st floor), and its direction (e.g., up, horizontal). Metadata for each record is spread over 45 lines. A single line generally contains multiple metadata items, with each constrained by its column position. Following the metadata, the observations are given as equally spaced float representations:

```
-.0001292 -.0001311 -.0001336 -.0001400 -.0001640
```

The COSMOS V1.2 schema provides the locations of metadata in specific lines and columns. For ASCII-based text parsing, it was efficient to represent the provided details in JSON format, following the protocol in Section 2. The JSON provides line and column numbers for each metadata point, as well as its equivalent SensorThings mapping:

```
    "lines": [
      {
        "line": 4,
        "column_start": 41,
        "column_end": 80,
        "short_description": eq_origin_time,
        "long_description": "Earthquake Origin
                             Time (GMT)",
        "sensorThings_mapping:
        "FeatureOfInterest.name"
      },
      ...
```

A LineParser class was developed to leverage JSON data to split, parse, and normalize directories of observations into SensorThings objects. This approach aligns with the original goal of expanding the database upon which the SRPM is built, which allows for a more comprehensive dataset. Furthermore, additional mapping and LineParser classes could be developed to handle multiple data formats, normalizing and mapping data from various databases to the common SensorThings model. The LineParser class returns a Dict of SensorThings objects:

```
# truncated for brevity
    (Datastream.phenomenonTime,
        (datetime.datetime(
        2007, 10, 31, 3, 4, 52,
        tzinfo=<UTC>),
        ...)
    ),
    ...
    (Thing.name, 1st Floor: Near Center)
```

The header metadata, in this particular case, includes enough information (the building's height and coordinates), to allow SensorThings objects to be passed directly to the Building instantiator. A Building also requires SeismicProperties. Since the period of the instrumented buildings is not part of the metadata, a function based on ASCE 7-10 (Equation 12.8-7, American Society of Civil Engineers, 2010) is provided to the instantiator. The function estimates the fundamental period of a building given its height and structural-system. The building's structural system was not part of the record metadata and was passed separately. Header metadata did, conveniently, include the peak EDP experienced by each channel. The highest value across a given record set was taken as the observed EDP, stored as a SensorThings object.

BSSA13 GMPE

The BSSA13 GMPE follows the functional form in (1) and is represented as

$$\begin{aligned}\ln Y = &F_E(M, mech) \\ &+ F_P(R_{JB}, M, region) \\ &+ F_S(V_{S30}, R_{JB}, M, region, z_1)\end{aligned} \quad (3)$$

Where Y is the median intensity measure; F_E is the event term dependent on M, magnitude, and $mech$, fault type; F_P is the path term dependent on R_{JB} distance, magnitude and region and F_S is the site term dependent on V_{S30}, shear wave velocity in the upper 30m of soil at the site and R_{JB}, M, $region$ and a constant. The three functional terms include coefficients which are period dependent, i.e., the value of the coefficient is dependent on the building's fundamental period: The event term, for example, is given as:

$$\begin{aligned}F_P = &[c_1 + c_2[M - M_{ref}] \ln(R/R_{ref}) \\ &+ (c_3 + \Delta c_3)(R - R_{ref})\end{aligned} \quad (4)$$

Where c_1, c_2, M_{ref}, R_{ref}, c_3, Δc_3 are period dependent model coefficients; M is magnitude, and R is derived from the distance R_{JB}.

A concrete BSSA13GMPE class and its functional terms (BSSA13PathTerm, BSSA13EventTerm, BSSA13SiteTerm) was implemented (`src/gmpe/bssa13.py`) through inheritance from the GMPE and FunctionalTerm ABCs. Each FunctionalTerm subclass, implemented a calculate method to handle the functional terms' arithmetic and calls a coefficient lookup helper function. The required dependent variables, M and $mech$ are inherited from the SeismicEvent, fundamental period from the Building, while R_{JB} was inferred from SeismicEvent and Building coordinate attributes.

A BSSA13GMPE instance is capable of calculating IM values across a continuous range of periods, T_{ij}, at a site j for event i (Figure 4). When passed a building, the discrete value of IM is produced which is used in the SRPM later.

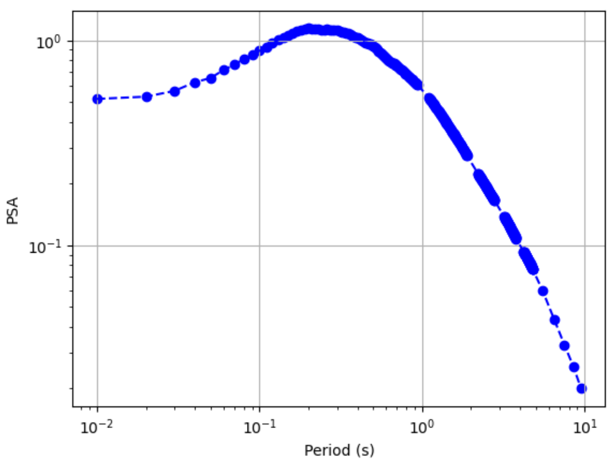

Figure 4. IM, Peak Spectral Acceleration (PSA) for a given range of periods return by the GMPE class for sites within 1km of event hypocenter.

SRPM and CBRR Implementation

The SRPM developed by Sun et al. (2022), is given as

$$\begin{aligned}\ln(Z_{ij}) = C_1 + C_2 \left(\ln(\hat{S}_{aT1})_{ij} + \eta_{E,i}\right) \\ + C_3(M_i - M_{\text{ref}}) \ln\left(\frac{H_j}{H_{\text{ref}}}\right)^{\frac{1}{2}} \\ + C_4(M_i - M_{\text{ref}}) \ln\left(\frac{T_j}{T_{\text{ref}}}\right) + \delta W_{ij}^Z \end{aligned} \quad (5)$$

Where Z_{ij} is the EDP, C_1, C_2, C_3, M_{ref}, H_{ref} and T_{ref} are model constants; the median IM, $(\hat{S}_{aT1})_{ij}$ at a given period T is the output calculated by `BSSA13GMPE.calculate()`; $\eta_{E,i}$ is the event term (approximately average difference between observed and predicted IM); M_i is the event magnitude, H_j is the building height, T_j is the building period.

A concrete SunSRPM class was implemented by inheriting from the SRPM ABC. This class includes a calculate_median_pfa method that returns the median EDP prediction. The method utilizes the Event and Building classes, along with the intensity measure (IM) inherited and calculated from the GMPE object, specifically through the `BSSA13GMPE.calculate()` method.

The CBRR is a class which extends the SRPM to predict the EDP for un-instrumented buildings. It achieves this by spatially interpolating the prediction residuals, δW_{ij}^Z derived from known sites. The interpolation is done through the geostatistical method of kriging. Kriging assumes that the closer an un-instrumented building is to an instrumented site with known residual, the more likely they are to have similar residuals. The residual at site j for event i may be calculated as the difference between the observed EDP from the SensorThings object against the median EDP prediction produced by the SRPM

$$\delta_{ij} = Z_{ij} - \bar{Z}_{ij} \quad (6)$$

Where δ_{ij} is the residual at site i for event j, Z_{ij} is the predicted value from the SRPM and $\bar{Z}_{ij}$ is the maximum EDP from the SensorThings object.

The CBRR implementation takes three objects, a list of monitored buildings, a list containing their respective residuals and a list of unmonitored predictions. The class implements a kriging algorithm and returns a list of residuals for the unmonitored sites. The individual components described in this section are all wrapped by the CBBRPredictor class (see `src/predictors.py`). The CBBRPredictor implements a `predict()` method which consists of the the following pipeline:

1. Map the observations in observations_path to SensorThings objects
2. Generate an internal list of instrumented Building objects from the SensorThings objects, and additional_metadata, if passed
3. Calculate the IM at instrumented and un-instrumented sites using the passed GMPE
4. Calculate the median EDP prediction using the passed SRPM,
5. Calculate the residuals at instrumented sites
6. Perform kriging to calculate residuals at un-instrumented sites
7. Add calculated residuals to the median EDP prediction
8. Return EDP prediction for all sites
9. Output as geodata

5. Results and Discussion

The pipeline was used to simulate the earthquake that occurred in 2007 at Alum Rock, California, near the city of San Jose. Geodata for buildings within a 30 km radius of the earthquake epicenter was sourced from OpenStreetMap. Only those buildings for which the source contained height data, approximately 305,000, were used in the simulation. The structural typology of the buildings was not available in the dataset and was assigned randomly to each building. Enhancing the prediction quality could be achieved through a more detailed assessment, which would involve assigning structure types based on additional data sources, although this falls outside the scope of the current work. The soil conditions, VS_{30}, required for the simulation was sourced from the U.S. Geological Survey (Thompson, 2018).

The CSMED records for this event included 41 instrumented buildings, five of which were within the 30 km study zone. The epicenter, magnitude, and fault type required by the model were obtained from the same database. The GMPE component of the predictor produced a shake-map of ground motion (PGA), as well as an estimate the IM experienced by each building at its specific period. We observe a maximum PGA of approximately 0.421g, exhibiting the expected strength decay conditioned by distance to the epicenter (see Figure 5a).

The SRPM, processed the IM estimations and made predictions of the median PFA for all buildings in the dataset, including those monitored buildings. The residuals at known sites were stored as an attribute of the CBRRPredictorClass. In the context of rapid loss estimates, the GMPE event-term, $\eta_{E,i}$ in (5) is not initially known and thus assumed to be zero. The event term may be added to the model as more information becomes available. The SRPM, partly due to the absence of this term made predictions which tended to generally underestimate (median residual, $\hat{\delta} = -0.184$, see Figure 6) the EDP values, but still suggests a linear prediction trend.

The CBRR fit a semivariogram based on the instrumented buildings and interpolated them geospatially using kriging (see Figure 5b). As a validation step, the 41 instrumented buildings were divided into approximately equal training and testing sets. The CBRR predictions of residuals (see Figure 7) is generally well-performing.

The concluding step of the pipeline adds the median predicted EDP from the SRPM and the spatially interpolated residuals from the CBRR, producing a "corrected" EDP prediction as shown in Figure 5c. The maximum PFA experienced by any building is around 0.22g. The distribution of PFA is typically log-normal (Figure 5d), with a mean of around 0.069g, this is up from the 0.057g median PFA predicted by the SRPM alone.

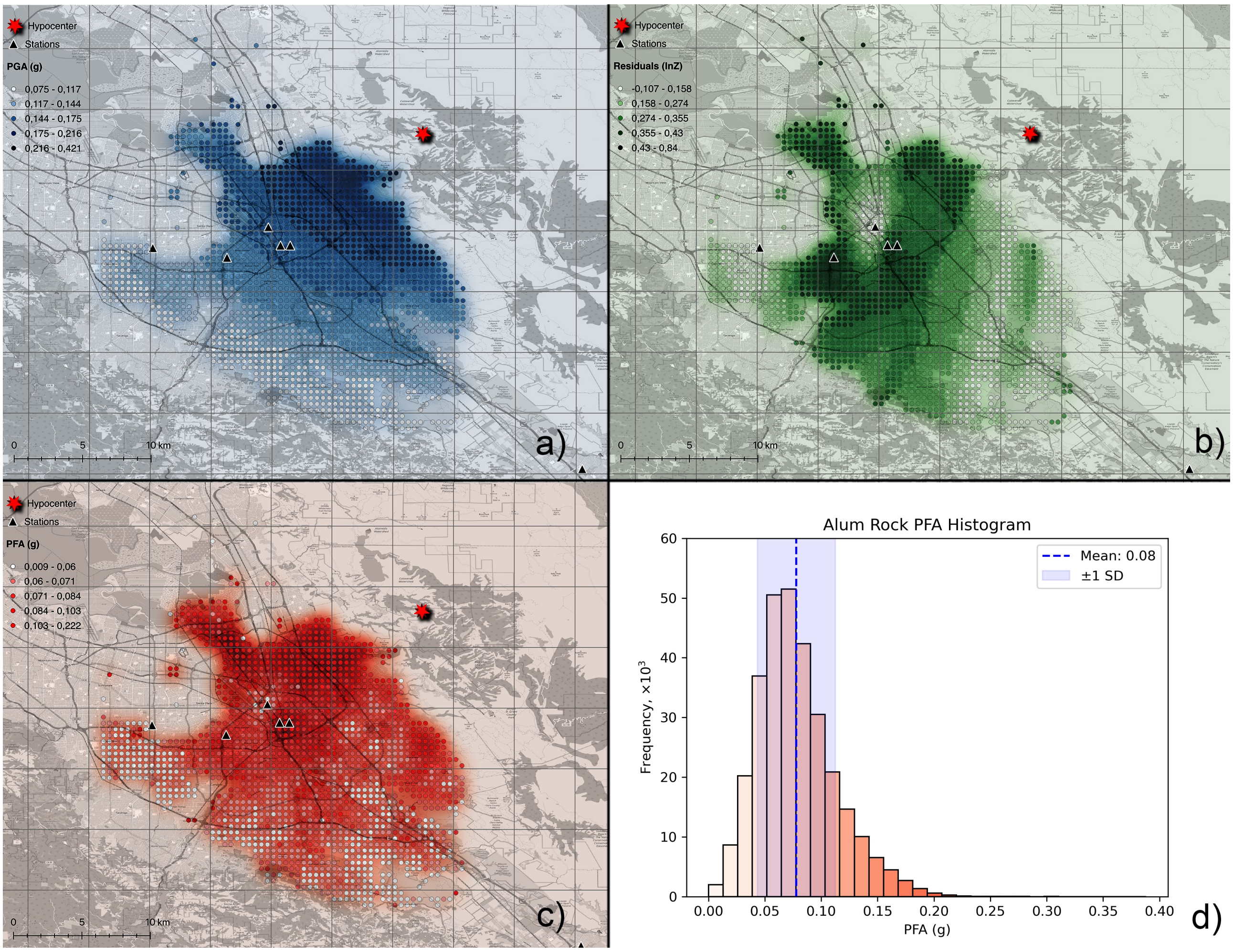

Figure 5. Heat-maps for a) PGA from the GMPE b) interpolated residuals from the CBRR c) PFA from the SRPM + CBRR and, d) histogram of PFA in study zone.

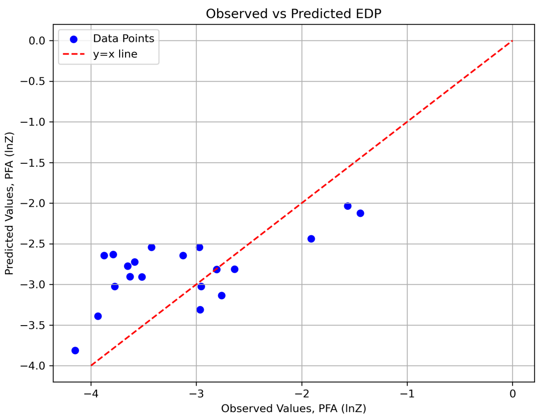

Figure 6. Observed EDP, lnZ against SRPM median predictions for monitored buildings.

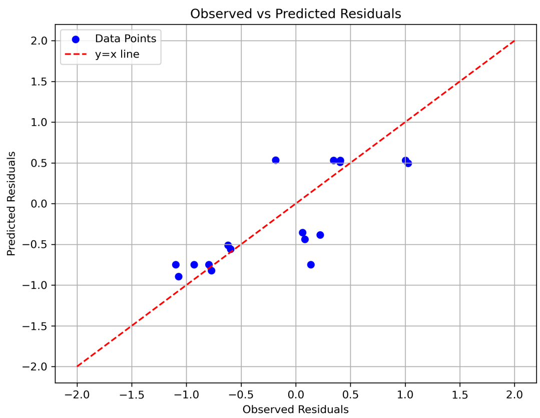

Figure 7. Observed residuals δ against predicted residuals.

The heat map reveals significant residual hot spots around known sites. In this case study, it was not feasible to compartmentalize the residuals based on other building characteristics such as height or structural type. The SRPM developed by Sun et al. (2022) similarly did not pursue such compartmentalization due to the initial database's relatively small size. To address this limitation, this study proposes adopting the OGC SensorThings model as a standardized data framework to facilitate the development of a more robust SRPM informed by larger, global building databases. The SensorThings model is lightweight and versatile, making it well-suited for integrating diverse datasets and enhancing the overall predictive capability of the SRPM. By expanding our repository to include a broader range of buildings, we can significantly improve the accuracy and applicability of seismic response predictions.

It is important to note that the GMPE component of the model may need regional adjustments, should the database be expanded to include global buildings. The proposed abstraction of the GMPE class, and indeed the entire pipeline, facilitates replacement of any of the integral components of the SRPM. It therefore becomes possible to conceive of larger global building-response database, an extended SRPM and with regionally conditioned GMPEs adjusting the model as required.

Finally there exists the potential to extend the concept whereby residuals are interpolated geospatially and thus enhancing or (correcting) predictions made by models for multiple hazards, such as over-heating or flood risk. For such a system interoperability becomes crucial, and we propose that OGC standards and SensorThings a suitable candidate.

6. Conclusions and Future Work

In this study, we presented an abstract building-response pipeline that leverages sensor readings, building properties, and ground motion prediction equations to estimate engineering demand parameters (EDP) for buildings. The readings from instrumented buildings served as benchmarks, allowing us to interpolate the differences between predicted and observed values geospatially using kriging. By harmonizing sensor data with the OGC SensorThings model, we envision compiling a larger global database of building-response data. This abstraction not only facilitates the integration of diverse datasets but also enhances the potential of our pipeline to develop more sophisticated models. With SensorThings as a common framework, the scope for extending this technique increases significantly, allowing for improved numerical methods in assessing the expected EDPs that buildings may experience during hazard scenarios.

While this work successfully demonstrated the core functionality of the pipeline, it was limited by the use of only one data source. Expanding to additional data sources may introduce unforeseen incompatibilities; however, this challenge presents an opportunity for further refinement and development. Additionally, the interpolation procedure used was relatively simplified, lacking subdivision of residuals by building type or height. Future work will focus on broadening the implementation to process larger datasets and developing an SRPM based on a more comprehensive database. Integrating and comparing our predictions with those generated by other numerical methods will also be a key aspect of this future research, thereby enhancing the robustness and applicability of our findings. Furthermore, there is also scope for incorporating the SRPM within Urban Digital Twins (UDTs). By embedding a building-response pipeline into a UDT framework, seismic hazard assessments can be continuously updated with real-time sensor data, improving the accuracy of rapid-loss estimates and long-term resilience planning.

Acknowledgments

This study has received funding from the European Union under the Horizon Europe Research & Innovation Programme (Grant Agreement no. 101123467 MULTICARE - Multi-hazard low-carbon resilient technologies and multi-scale digital services for a future-proof, sustainable & user-centred built environment).

References

Abdelmalek-Lee, E., Jain, T., Galicia Madero, S., Sun, H., Burton, H., Wallace, J., 2023. Relational database for building strong motion recordings used for seismic impact assessments. *Earthquake Spectra*, 39(2), 1277–1297. http://journals.sagepub.com/doi/10.1177/87552930231169306.

American Society of Civil Engineers, 2010. *Minimum Design Loads for Buildings and Other Structures*. 7th edn, ASCE/SEI. ASCE/SEI 7-10.

Archuleta, R. J., Steidl, J., Squibb, M., 2006. The COSMOS Virtual Data Center: A Web Portal for Strong Motion Data Dissemination. *Seismological Research Letters*, 77(6), 651–658. https://doi.org/10.1785/gssrl.77.6.651.

Baker, J., Bradley, B., Stafford, P., 2021. *Seismic Hazard and Risk Analysis*. 1 edn, Cambridge University Press.

Boore, D. M., Stewart, J. P., Seyhan, E., Atkinson, G. M., 2014. NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes. *Earthquake Spectra*, 30(3), 1057–1085. http://journals.sagepub.com/doi/10.1193/070113EQS184M.

California Geological Survey, U.S. Geological Survey, 2005. Center for engineering strong motion data (cesmd). Users who utilized the data center's services to access strong-motion products are recommended to cite CESMD as: California Geological Survey and U.S. Geological Survey (2005). Center for Engineering Strong Motion Data (CESMD). U.S. Geological Survey.

Erdik, M., Şeşetyan, K., Demircioğlu, M. B., Hancılar, U., Zülfikar, C., 2011. Rapid earthquake loss assessment after damaging earthquakes. *Soil Dynamics and Earthquake Engineering*, 31(2), 247–266.

Helffrich, G., Wookey, J., Bastow, I., 2013. *The Seismic Analysis Code: A Primer and User's Guide*. Cambridge University Press. Google-Books-ID: LfyaAAAAQBAJ.

Krischer, L., Smith, J., Lei, W., Lefebvre, M., Ruan, Y., de Andrade, E. S., Podhorszki, N., Bozdağ, E., Tromp, J., 2016. An Adaptable Seismic Data Format. *Geophysical Journal International*, 207(2), 1003–1011. https://doi.org/10.1093/gji/ggw319.

Liang, S. H., Tania, K., Hylke, v. d. S., 2024. Ogc sensorthings api part 1: Sensing version 1.1" ogc® implementation standard (2021). original-date: 2016-05-24T12:04:05Z.

Ringler, A. T., Evans, J. R., 2015. A Quick SEED Tutorial. *Seismological Research Letters*, 86(6), 1717–1725. https://doi.org/10.1785/0220150043.

Sun, H., Burton, H. V., Stewart, J. P., Wallace, J. W., 2022. Development of a Generalized Cross-Building Structural Response Reconstruction Model Using Strong Motion Data. *Journal of Structural Engineering*, 148(6), 04022053. https://ascelibrary.org/doi/10.1061/

Thompson, E. M., 2018. An Updated Vs30 Map for California with Geologic and Topographic Constraints.

Wegner, P., 2003. Object-oriented programming (OOP). *Encyclopedia of Computer Science*, John Wiley and Sons Ltd., GBR, 1279–1284.

Appendix B: Personal Evaluation

This year, I dedicated a significant effort in closing knowledge gaps, particularly with regards to my capabilities in the digital technology domain. I am pleased to report that this was quite successful: I began my PhD concerned that I would be unable to keep up with the the technology I was meant to be applying in my work; now, while gaps still exist, the foundation I have built is robust and I can keep up with new content without the initial "newbie" problems. Through my supervision of masters students' labs, assignment writing and grading, I feel confident in my ability and look forward to the upcoming years. Below follows a brief outline of my progress this year:

Discipline Skills

I have become familiar, and it in some cases competent with the digital technologies required for the execution of my work, namely:

1. Python as the main programming language, strong competency,
2. Database Management Systems (DBMS), PostgreSQL, PostGIS,
3. Version Control: Git, GitHub
4. Web Development; familiarity with:
   (a) HTML,
   (b) JS,
   (c) CSS,
5. Web Standards: mostly OGC, specifically OGC SensorThings and OGC API.
6. Testing, namely python tests (unittest and pytest),
7. Web Technology:
   (a) Hosting a linux server,
   (b) Apache Tomcat / Apache Webserver
   (c) FROST Server
8. Software architecture generally:
   (a) APIs,
   (b) Software stacks generally (conceptually),
   (c) Business / front end layers (conceptually),

Items 5 and 7 were skills formed through my role as a lab assistant in the GEO1007 (Geoweb Technology) course. On the side of risk and resilience studies, I've become more competent in the aspects of seismic risk engineering. There is room for growth in the other hazard domains which I intend to tackle at the beginning of my second year.

Research Skills

1. Research Management: Through this first year, I've studied the main concepts of building and designing research. Specifically, research which follows the Design Science paradigm.

2. Academic Thinking: It has been an effort to shift from a 'deliverable' based way of working from my previous professional career, to the 'research' based thinking. I have slowed down, but there is room for improvement.

3. Learning-on-the-job: This has been one of the most effective ways for my gaining of experience and ability. Working on the geo-web course hands on was a great experience; as was helping with running the labs. I'm very excited about mentoring. My first conference paper (as a PhD research) was also very good experience. I do notice I need to learn how to scope out my reading better and intend to really focus on that this year.

Transferable Skills

During this first year, I completed the PhD StartUp module. A complete plan of my upcoming courses was presented in Table 1 earlier.

Appendix C: Data Management Plan

Plan Overview

A Data Management Plan created using DMPonline

Title: Spatial Decision Support Systems for Multi-Hazard Resilient Cities

Creator: Justin Schembri

Principal Investigator: Justin Schembri

Data Manager: Justin Schembri

Affiliation: Delft University of Technology

Template: TU Delft Data Management Plan template (2021)

Project abstract:

As weather-driven hazards intensify, cities face increased risk from multiple hazards; in parallel the building stock ages and densifies, leading to populations both exposed and vulnerable. Vulnerability is, to a greater extent, the only mitigable risk component and it is defined as the capacity of a building to perform well under varying level of stress induced by natural hazard. Urban vulnerability is modeled in several ways, though all models have in common an attempt to correlate an expected level of damage to the building stock when impacted by a hazard. At the urban level, vulnerability modeling invariably involves assumptions, aggregations and uncertainty; the impacts of such assumptions are very difficult to addressed. With vulnerability calculations playing such a critical role in mitigation decision-making, having a reliable and robust methods of calculating risks is critical. Advances in building simulations, machine learning (ML), artificial intelligence (AI) underpin much of the present approach to vulnerability modeling. It is, however, difficult to validate such techniques against real observations for various reasons, the most significant of which is an apparent shortage of detailed data.Such an apparent scarcity of empirical vulnerability data, this study shall argue, is more of an issue of 1) interoperability and 2) openness, rather than true lack. In fact, throughout the risk domain, one observes a lack of standardization of the basic data and processes; a reality more apparent in the relatively novel multi-hazard domain.

To address this, this work shall investigate and \textbf{develop a standards based, information and processing infrastructure
which leverages heterogeneous vulnerability data sources and services in one environment, supporting multi-hazard decision support systems.} The first half of this research investigates is design-science centric and seeks to develop (digital) technical and infrastructural solutions which enable the building and maintenance of a harmonized, open-access, multi-hazard empirical vulnerability database. The key aspect here is designing a system for long-term support, rather than one-off data preparation approaches which are common in research. After resolving data incompatibilities, a pilot database will be populated available historical records associated with vulnerability or performance of the built environment.

The second half of the research leverages the harmonized data. Statistical and machine-learning methods will be used to develop an initial suite of cross-building, multi-hazard fragility and /or vulnerability functions, conditioned on relevant predictor variables. A sensitivity analysis will assess statistical dependencies and sensitivities to explicitly define uncertainty within building performance models. Additionally, methods will be developed to enrich the vulnerability functions when more detailed local information is available. A hybrid

method for incorporating numerical approaches will complement long-return-period hazards (e.g., seismic) in the vulnerability functions.Geostatistical techniques will enhance the models, enabling real-time sensor data to update and refine the baseline through Bayesian techniques, particularly for low-return-period hazards like building overheating. The final contribution is a Spatial Decision Support System, integrated with existing open-source tools (e.g., OpenQuake, EnergyPlus), to assist decision-makers in economic planning under information uncertainty. This system will clarify model uncertainties, quantify the benefits of additional data collection in terms of uncertainty reduction, and help assess whether reducing uncertainty justifies the associated costs.

ID: 165413

Start date: 15-10-2023

End date: 15-10-2027

Last modified: 01-12-2024

Spatial Decision Support Systems for Multi-Hazard Resilient Cities

0. Administrative questions

1. Name of data management support staff consulted during the preparation of this plan.

My faculty's (Architecture) data steward shall be asked to review this DMP.

2. Date of consultation with support staff.

I. Data description and collection or re-use of existing data

3. Provide a general description of the type of data you will be working with, including any re-used data:

Type of data	File format(s)	How will data be collected (for re-used data: source and terms of use)?	Purpose of processing	Storage location	Who will have access to the data
Hazard maps	GeoJson, GDB, Shapefile, GeoTIFF, GML, WKT, WKB	Open data	Risk assessments	Local machine / TUD VM	Principal researcher
Energy urban data	GeoJson, GDB, Shapefile, GeoTIFF, GML, WKT, WKB	Open data	Risk assessment	Local machine / TUD VM	Principal researcher
Exposure models	GeoJson, GDB, Shapefile, GeoTIFF, GML, WKT, WKB	Open data	Risk assessment	Local machine / TUD VM	Principal researcher
Hazard data	Geospatial	GeoJson, GDB, Shapefile, GeoTIFF, CSV, EPW	Compilation of existing hazards data	Local machine / TUD VM	Principal researcher

4. How much data storage will you require during the project lifetime?

- < 250 GB

II. Documentation and data quality

5. What documentation will accompany data?

- Data will be deposited in a data repository at the end of the project (see section V) and data discoverability and re-usability will be ensured by adhering to the repository's metadata standards
- README file or other documentation explaining how data is organised
- Methodology of data collection

III. Storage and backup during research process

6. Where will the data (and code, if applicable) be stored and backed-up during the project lifetime?

- Git(lab)/subversion repository at TU Delft
- Project Storage at TU Delft

IV. Legal and ethical requirements, codes of conduct

7. Does your research involve human subjects or 3rd party datasets collected from human participants?

- No

8A. Will you work with personal data? (information about an identified or identifiable natural person)

If you are not sure which option to select, first ask your Faculty Data Steward for advice. You can also check with the privacy website . If you would like to contact the privacy team: privacy-tud@tudelft.nl, please bring your DMP.

- No

8B. Will you work with any other types of confidential or classified data or code as listed below? (tick all that apply)

If you are not sure which option to select, ask your Faculty Data Steward for advice.

- No, I will not work with any confidential or classified data/code

9. How will ownership of the data and intellectual property rights to the data be managed?

For projects involving commercially-sensitive research or research involving third parties, seek advice of your Faculty Contract Manager when answering this question. If this is not the case, you can use the example below.

This is an open-source project and thus IP will be published under the Apache or GNU license.

V. Data sharing and long-term preservation

26. What data will be publicly shared?

- All validated non-positive results
- All data (and code) underlying published articles / reports / theses
- All data (and code) produced in the project

28. How will you share your research data (and code)?

- I will upload the data to another data repository (please provide details below)
- All data will be uploaded to 4TU.ResearchData

I expect to use Github too.

30. How much of your data will be shared in a research data repository?

- < 100 GB

31. When will the data (or code) be shared?

- As soon as corresponding results (papers, theses, reports) are published

32. Under what licence will be the data/code released?

- Apache

VI. Data management responsibilities and resources

33. Is TU Delft the lead institution for this project?

- Yes, leading the collaboration - please provide details of the type of collaboration and the involved parties below

34. If you leave TU Delft (or are unavailable), who is going to be responsible for the data resulting from this project?

I intend to host / store the data on a TUD managed VM / server, so the data would be accessible there.

35. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Quite a lot of my energy is going into ensuring the data will be FAIR; I don't expect to go past 1TB of data.