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.

https://doi.org/10.5194/isprs-annals-X-G-2025-787-2025

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,

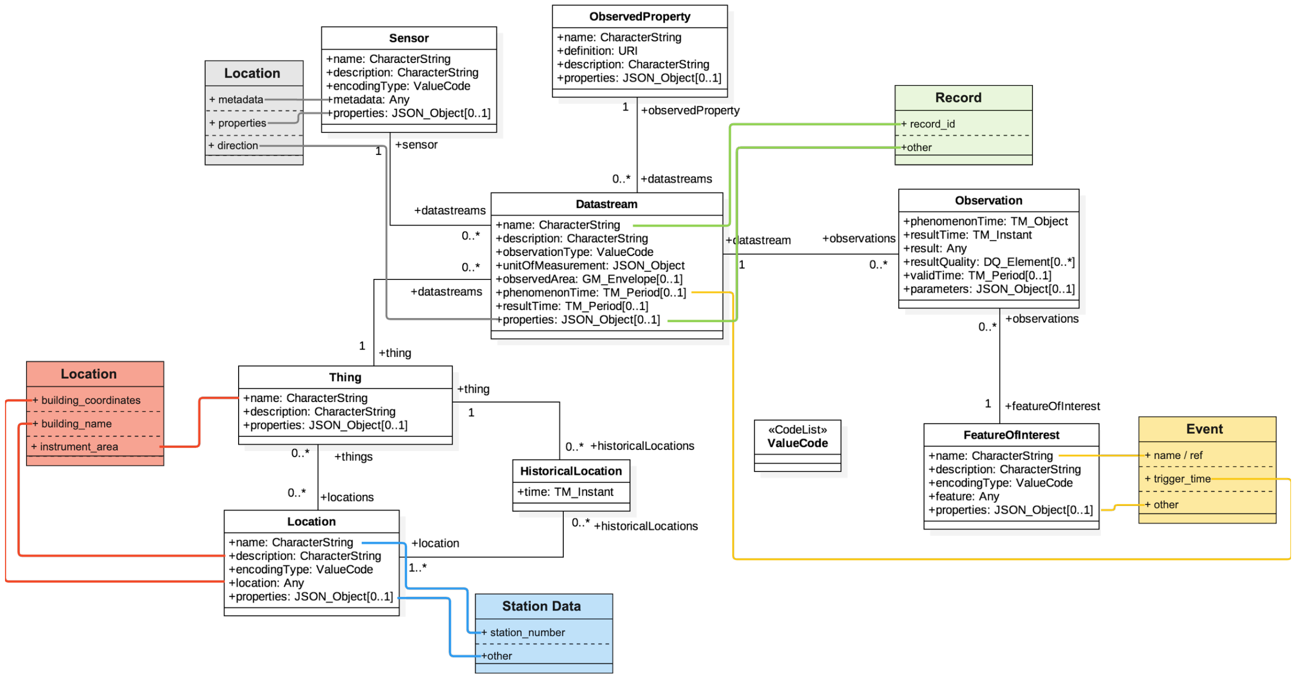

Figure 1. The OGC SensorThings data-model (white boxes) augmented to include the proposed mappings (colored boxes).

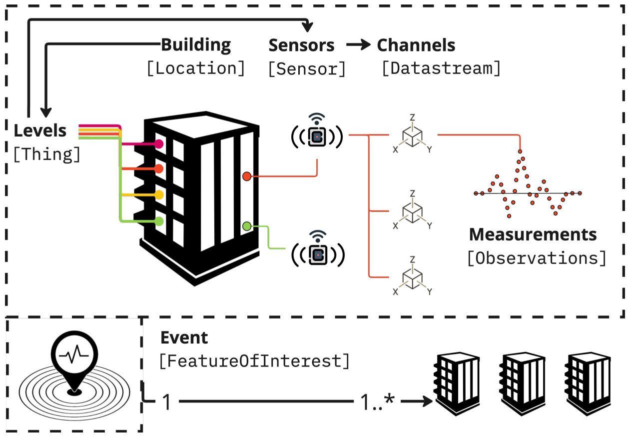

Figure 2. Relationship between real-world instrumentation set up and SensorThings schema. Fixed width text are the equivalent SensorThings entities.

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:

$$\ln Y = F_M + F_P + F_S + \epsilon\delta \qquad (1)$$

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

$$\ln(Z_{ij}) = F_{Y,ij} + F_{H,ij} + F_{T,ij} + \delta W^Z_{ij} \qquad (2)$$

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^Z_{ij}, 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

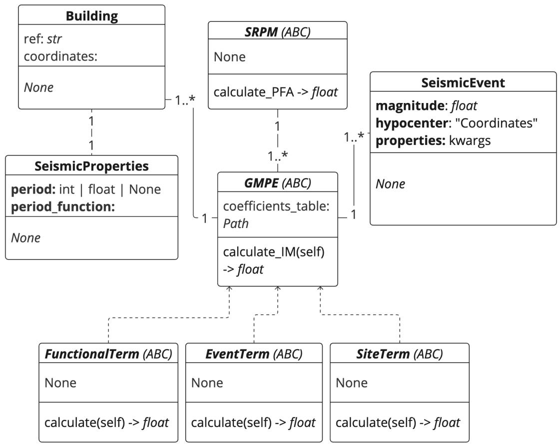

Figure 3. UML diagram showing relationship between Building, SeismicEvent, GMPE, its terms and the SRPM.

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

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

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:

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

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.

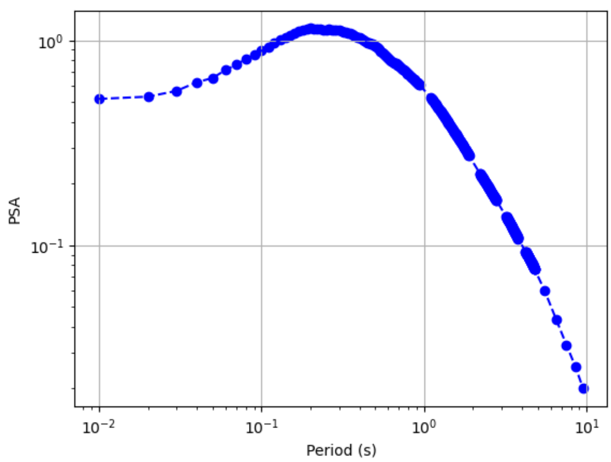

Figure 4. IM, Peak Spectral Acceleration (PSA) for a given range of periods return by the GMPE class for sites within 1km of event hypocenter.

SRPM and CBRR Implementation

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

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

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

$$\delta_{ij} = Z_{ij} - \bar{Z}_{ij} \tag{6}$$

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.

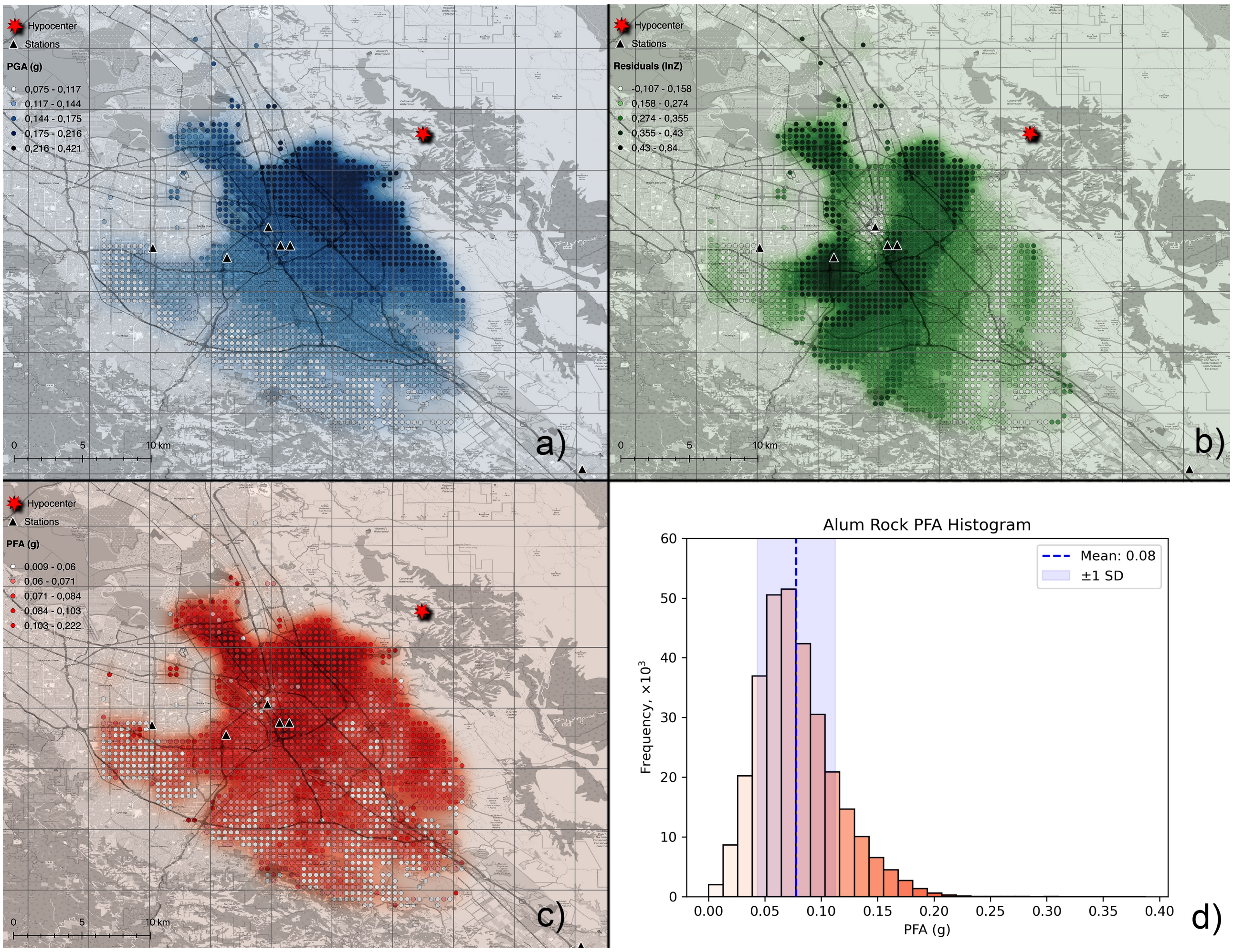

Figure 5. Heat-maps for a) PGA from the GMPE b) interpolated residuals from the CBRR c) PFA from the SRPM + CBRR and, d) histogram of PFA in study zone.

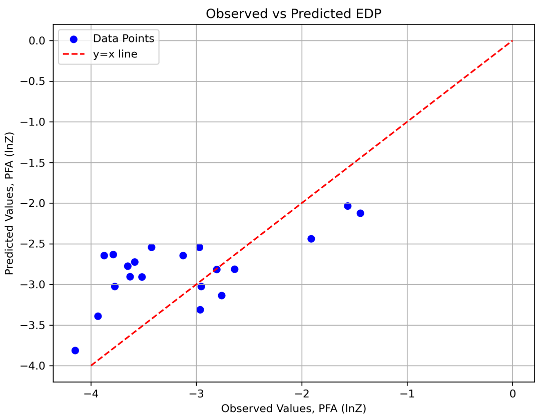

Figure 6. Observed EDP, lnZ against SRPM median predictions for monitored buildings.

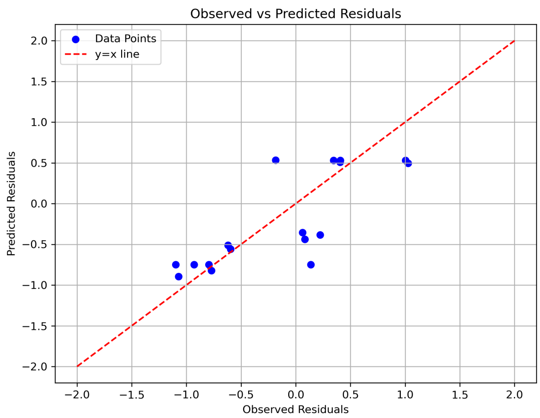

Figure 7. Observed residuals δ against predicted residuals.

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.