INSSEPT: An open-source relational database of seismic performance estimation to aid with early design of buildings

Article *in* Earthquake Spectra · June 2020

DOI: 10.1177/8755293020919857

Mohsen Zaker Esteghamati
Virginia Polytechnic Institute and State University

Jeonghyun Lee
University of California, Los Angeles

Madeleine Flint
Virginia Polytechnic Institute and State University

INSSEPT: An open-source relational database of seismic performance estimation to aid with early design of buildings

Mohsen Zaker Esteghamati[1], Jeonghyun Lee[1,2], Matthew Musetich[1,3], and Madeleine M. Flint[1]

[1] Department of Civil and Environmental Engineering, Virginia Polytechnic Institute, and State University, Blacksburg, Virginia
[2] Department of Civil and Environmental Engineering, University of California at Los Angeles, Los Angeles, California
[3] Department of Civil and Environmental Engineering, University of Notre Dame, Notre Dame, Illinois

Abstract. Performance-based earthquake engineering (PBEE) assessments are data-, effort and time-intensive, usually requiring a detailed structural model and limiting their integration with early design. Decades of research have produced an abundance of PBEE assessments for different structural systems and building taxonomies. The results of these PBEE studies can be assimilated to approximately represent the seismic design space for new structures and to identify possibly optimal systems with low effort. This paper introduces an open-source relational database, **In**ventory of **S**eismic **S**tructural **E**valuations, **P**erformance Functions and **T**axonomies for Buildings (INSSEPT) that contains PBEE assessment of 222 buildings from literature and is freely available to the public in a natural hazard repository. INSSEPT is organized to provide a curated building taxonomy and PBEE data to readily serve as a resource for early design or PBEE-derived regional seismic risk analysis.

Keywords: Relational database· Early design· Fragility assessment· Seismic performance inventory· Performance-based earthquake engineering

1 Introduction

A desire to increase community resilience motivates mitigating the tremendous economic, social, and environmental losses caused by natural hazards. A first step in achieving community resilience is the incorporation of resilience assessment in building design, especially during early design when many important decisions such as the selection of a soil foundation-structure-envelope (SFSE) system, have yet to be made. The lack of constraints in early design supports consideration of a broad range of SFSE alternatives, but this design phase is also not standardized, imprecise, and iterative. The lack of details at the preliminary design stage makes it unfeasible to perform high-fidelity and computationally expensive simulations. Therefore, a reasonably accurate and less intensive approach to incorporate resilience assessment is needed.

The "*Resilient and Sustainable Buildings* (RSB)" initiative at Virginia Tech is supported by a multi-disciplinary research group that aims to fill this gap through a modular framework. The proposed framework comprises three modules: (1) generating site-specific SFSEs (2) probabilistic assessment of performance and operation and (3) ranking and optimization of candidate systems [1]. It uses stakeholder preferences and a basic building characterization (e.g., location, number of stories and dimensions) to identify optimal SFSE systems in terms of resiliency and sustainability metrics. The first step in the proposed RSB methodology relies on a performance inventory of the generic SFSE system as possible candidate systems, which is the focus of this paper.

A performance inventory can pave the path for the application of simplified models in early design using existing results. If properly organized, it can provide a wealth of data for further examination using machine learning tools and facilitates regional or sectoral performance based earthquake engineering (PBEE) assessment by aggregating analyses of different individual buildings. Historically, performance inventories are developed as part of regional risk tools using different methodologies, such as empirical, analytical, hybrid and expert-based[2]. Nevertheless, the development of building performance inventories is an arduous task, as it requires computationally demanding analysis (in case of analytical methodologies[3]), post-earthquake data availability (in case of empirical methodologies[4])) and time and human resources.

Recent efforts have developed global or local databases to quantify vulnerability to natural hazards. The Applied Technology Council developed a performance inventory as part of the ATC-13 project and implemented it as the HAZUS Multi-Hazard (MH) loss assessment tool. The database is specifically compiled for the United States and contains loss information for buildings, lifelines, and transportation. The HAZUS methodology for seismic loss comprises six independent modules to allow different levels of fidelity. These modules include earth science hazards, infrastructure inventory, direct damage prediction, induced (indirect) damage, direct loss, and indirect loss. At the core of the methodology is damage estimation through the direct damage module. In HAZUS, damage is predicted using fragility curves and capacity curves of 36 types of steel, concrete, wood and masonry buildings (support for user-defined curves is also provided)[5].

Since HAZUS is tailored to US locations, similar tools have been developed for other geographical locations, e.g., SELENA [6], LNECLOSS [7]. However, most of these alternative tools use a methodology similar to that of HAZUS, and their differences stem from the GIS-based modules for hazard and building inventories. Perhaps the most comprehensive effort is made by the global earthquake model (GEM) initiative. GEM's vulnerability database contains more than a thousand fragility/vulnerability, capacity, and damage-to-loss curves from varied sources [8]. The database is organized around five main categories of general information, data, geographical location, modeling information, and quality rating. Each category then contains several attributes. For example, general information uses three attributes of category (types of structure), type of assessment (fragility function, vulnerability function, etc.) and documentation (authors, type of publication) [9]. The fragility functions are obtained from an empirical database by Rossetto et al. [10], an analytical database by D'ayala and Meslem and the European Syner-G project [11] and are implemented in the OpenQuake open-source tool [12].

Databases such as those developed for HAZUS and GEM relied on substantial new research efforts to create comprehensive performance inventories[8]. Building a performance inventory without additional (new) structural analysis requires integration of data from PBEE studies scattered across sources such as conferences, journal papers, or technical notes. These existing evaluations use varying methods and tools and the final results significantly depend on these analysis assumptions and choices. The breadth of PBEE methods and scope requires a careful ontology to organize features that are flexible for categorization of information of all selected studies and expansion of database to capture future studies. Further application of these existing evaluations consequently requires a comprehensive database that compiles and organizes the results; details the methodologies used in each study to ensure consistent applicability; and curates metadata.

This paper introduces the Inventory of Seismic Structural Evaluations, Performance Functions and Taxonomies for Buildings (INSSEPT), which has been created as part of the RSB initiative. INSSEPT is an open-source relational database to support the rapid derivation of PBEE assessment for a large number of systems with a minimum amount of time and effort. While effort has been made to develop a schema extensible to all structure types, version 1.0.3 of the database is structured around the published results of PBEE assessments of 222 mid-rise buildings, and is hosted by the DesignSafe Cyberinfrastructure [13]. The curated performance data use the Pacific Earthquake Engineering Research (PEER) Center's PBEE methodology [14,15] to relate site seismic hazard in terms of intensity measure (IM) to buildings' engineering demand parameters (EDP) and damage states (DS). Performance metrics were extracted in terms of damage fragilities and probabilistic demand models. INSSEPT access is provided for DesignSafe registered users through SQL query commands in an accompanying Jupyter notebook. An interactive tool is provided to extract and visualize data based on location, number of stories and type of needed results [16].

This paper first reviews the relevant aspects of PBEE that provide the content of INSSEPT, including the typical data obtained from a PBEE assessment. It next explains the implementation structure of INSSEPT as a relational database, its database schema, and the taxonomy used in recording entries. INSSEPT was developed by focusing on structural engineering use cases, rather than the intricacies of relational databases, and as a result, the design prioritized the institutional background and intuition of civil engineers. Querying is discussed to facilitate extraction of data from the database, and several examples are presented. Lastly, conclusions regarding the significance of this effort and its possible future extension to other types of buildings and hazards are provided.

2 Performance-based earthquake engineering data

PBEE assessments are commonly performed by subjecting a numerical **model**[1] of a building to different ground motion (GM) excitation records. The numerical model should be able to cover the wide spectrum of structure's behavior from elastic to inelastic under the applied GM records, and considers the true mechanical and dynamic aspects of actual structure such as damping, presence of second-order effects, and degradation of stiffness and strength. The structure's response is then measured in terms of **engineering demand parameters** (EDP) (such as peak floor acceleration and drift values), and related to ground motion shaking intensity levels through **intensity measures** (IM) (such as peak ground acceleration (PGA) or spectral acceleration (S_a). The probabilistic relationship between IM and EDP forms the basis of PBEE assessment and is referred to as probabilistic seismic demand analysis (PSDA).

Different **procedures** are available to perform PSDA; the focus of the current version of INSSEPT is on two popular approaches of "*incremental dynamic analysis* (IDA)" (Figure 1.a) and "*cloud analysis*" (Figure 1.b). Cloud analysis subjects the numerical model to unscaled (or minimally scaled) GM records and develops a regression model between logarithmic values of EDP and

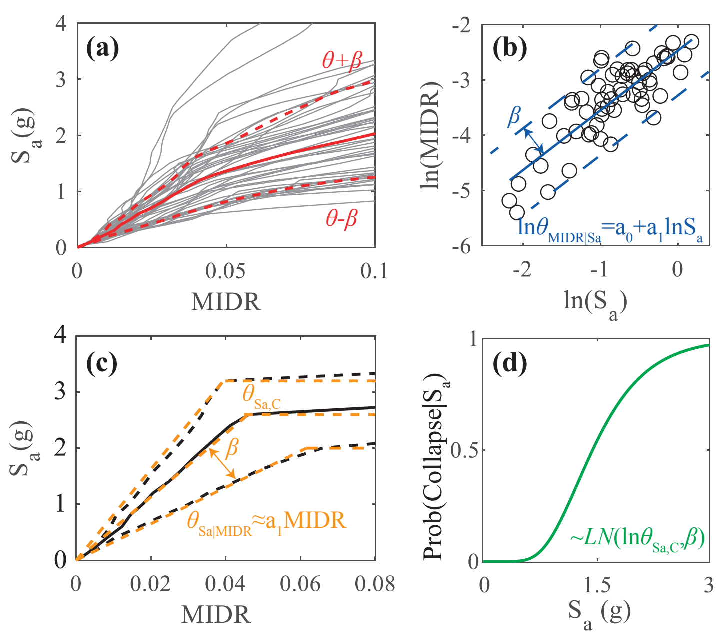


Fig. 1: Examples of results and stored parameters; figure replicated from accompanying manuscript. (a) Incremental Dynamic Analysis (IDA) empirical quantile curves. (b) Cloud analysis and log-log regression. (c) Simplified IDA representation with reduced data storage. (d) Collapse fragility curve.

[1] Bold items are categories of data collected in INSSEPT

IM to determine the median probability of exceeding a certain EDP level ($\theta_{EDP|IM}$); uncertainty is accounted for using the regression error (β). In contrast, IDA applies increasingly scaled GM records and estimates the probability of exceeding a particular EDP level either empirically (by directly counting the number of records that forces the structure to exceed the given threshold) or by fitting conditional distribution parameters.**GM scaling** treatment and **selection** are significantly different between cloud and IDA methods. While cloud analysis uses unscaled site-specific GM suites, the IDA method mostly uses standardized GM sets which are then repeatedly scaled to different levels of the selected IM. Nevertheless, both IDA and cloud methods aim to provide the probability of exceeding a particular EDP level, conditioned on IM level, which is referred to as "**seismic fragility**"(Figure 1.d). The accuracy of the estimated seismic fragility largely depends on the precisionof numerical model.

An EDP level for a fragility is selected in a way that can describe structure's performance level, following standardized definitions in literature, such as immediate occupancy or collapse performance levels. Collapse is of particular interest for structural engineers, historically due to the fact that modern seismic building codes are developed to prevent structures from collapse under rare strong earthquakes. As a result, PBEE literature provides several methods for collapse determination, e.g., using pre-defined EDP thresholds, or non-convergence of a valid and stable numerical model due to excessive displacement demands that revoke equilibrium, or a certain reduction percentage of IDA curves' initial slopes. As shown in Figure 1.d, the seismic fragilities are often modeled using a lognormal distribution with given median ($\theta_{IM,C}$) and lognormal standard deviation (β), although collapse fragilities can be modeled using other distributions, such as logistic or binomial.

While the previous discussion depicts the varying types of data involved in PBEE, additional effort is needed to summarize conventional PBEE data. The raw data often do not provide insight, occupy larger storage, and reduce machine learning and data mining performance [17]. In this regard, while cloud analysis methods **results** and seismic fragilities can be summarized using regression coefficients and distribution **parameters**, respectively, most of published IDA results do not provide any summary statistics as shown in Figure 1.a. In this regard, as part of first module of the RSB initiative, Tahir developed a digitization tool to summarize published IDA results in terms of pre-collapse slope and IM value at collapse, where IDA of midrise buildings can be represented using a bilinear model following ATC 19's simplified behavior of short or long-period structures [18]. Figure 1.c illustrates this concept for an arbitrary IDA result; a complete description of this methodology is provided in [19]. Tahir's database of collected IDA information is also included in INSSEPT.

3 Methodology

3.1 A primer on relational databases

INSSEPT was developed as an open-source, structured, relational database (RDB) using MySQL. While spreadsheets (such as Excel) store data as a grid of rows and columns, databases use tables and fields to organize data more efficiently in terms of searchability and size. The general structure of the database is a series of interrelated tables, which are linked to one another through shared fields called keys. Tables consist of a combination of keys and attributes. A primary key is a unique identifier for each tuple in a table, whereas a foreign key acts as a link between tables, referencing the primary key of another table to formally define a relationship between tables. An attribute is the description of table entities, and takes a type such as character, integer, or Boolean. The combination of tables, attributes, and relationships between tables is referred to as the schema, or structure, of the database. Therefore, compared to conventional spreadsheets, RDBs seamlessly integrate multiple data sources; reduce data redundancy; improve resource and user efficiency; and facilitate easy data retrieval.

Figure 2 contrasts the conventional spreadsheet approach with an alternative RDB to demonstrate the benefits of adopting RDBs for the next generation of data sets in PBEE applications. In

Figure 2.a, a typical spreadsheet treats each row as a unique data entry corresponding to one particular building, and the values across column contain, e.g., information about building composition, structural modeling, site class, and median collapse spectral acceleration. While it is straightforward to present the PBEE data in this manner for a small data set, several issues emerge for bigger data sets. First, spreadsheets cannot provide data structuring beyond individual fields. For example, information about building geometry and structure is presented alongside information about numerical modeling and site seismic hazard, and hence all different layers of information are presented in columns. However, by dividing the layers of information into several fields such as building taxonomy (stories, floor area, etc.), material (e.g. steel, wood) and lateral-resisting systems (material, type, and configuration), structural modeling attributes (P-, damping), and site description (site seismic class, location), a holistic picture of the systems is provided. While subcategories might be added as additional spreadsheet columns, this trivial solution becomes impractical with a large number of fields. Even for a smaller number of fields, RDB improves readability of data compared to a conventional spreadsheet. Second, spreadsheets can contain a large number of repetitive data, and it can be difficult to distinguish between entries that are duplication with limited variation (e.g., the same building analyzed assuming 2% and 5% damping) and entries that more coincidentally share common attributes. RDBs' keys provide an effective tool to reduce data redundancy and increase resource efficiency.

Structure Name	Stories	Category	Occupancy	Damping Model	Element Type	Location	Site Class
G2007-Com-C4-PER-A	4	Commercial	Office	Rayleigh	Concentrated plasticity	Los Angeles, CA	D
L2011-Com-RC8-S-Ductile	8	Commercial	Office	Rayleigh	Concentrated plasticity	Los Angeles, CA	D
C2009-Com-RC9-Rix	9	Residential	Residential	Rayleigh	Fiber, FE	Memphis, TN	D
J2011-Res-S3	3	Residential	Residential	NULL	Concentrated plasticity	Tehran, Iran	D
J2014-Com-RC3-C2	3	Commercial	Office	NULL	Concentrated plasticity	Avellino, Italy	B

(a) Conventional spreadsheet

BUILDING

BldgID	StructNm	Stories	CatOccID	StrctModID	SiteID
1	G2007-Com-C4-PER-A	4	2	3	1
2	L2011-Com-RC8-S-Ductile	8	2	3	1
3	C2009-Res-RC9-Rix	9	1	1	3
4	J2011-Res-S3	3	1	2	2
5	J2014-Com-RC3-C2	3	2	2	4

SITE

SiteID	Location	Site Class
1	Los Angeles, CA	D
2	Tehran, Iran	D
3	Memphis, TN	D
4	Avellino, Italy	B

STRCTMOD

StrctModID	Damping Model	Element Type
1	Rayleigh	Fiber, FE
2	NULL	Concentrated plasticity
3	Rayleigh	Concentrated plasticity

CATOCC

CatOccID	Category	Occupancy
1	Residential	Residential
2	Commercial	Office

Legend
Primary Key
Foreign Key

(b) Relational database

Fig. 2: Translation of PBEE data from spreadsheet grid to RDB schema

Figure 2.b shows the alternative presentation of spreadsheet data through a simple RDB. While the table contains multiple groups of repeated fields (i.e. lateral material and class, site class and location, etc.), the RDB uses several foreign keys as links between building identifiers and the other related information presented in all inter-related tables. For example, A `MatSysID` of 2 indicates that the given building has a concrete frame lateral system. Therefore, rather than entering the information for a building's materials and system each time a new building is incorporated, the information only needs to be added once for each unique `MatSysID`.

As Figure 2 illustrates, there are two main types of relationships between different 210 tables: "*one-to-many*" and "many-to-many". In a one-to-many relationship, the parent record can reference multiple records in a related table. For example, the relationship between the `BUILDING` and `SITE` tables is a one-to-many type. For example, a building will only exist at one site class (i.e. `SiteClass` attribute), while a site class may be common between multiple buildings. On the other hand, in a many-to-many relationship, several records are linked simultaneously. For example, `AUTHLIST` and `SOURCE` have a many-to-many relationship, where a paper could have multiple authors, and an author could have written multiple papers. These relationships are accounted for with junction tables. As shown in Figure 3, `AUTHLIST` acts as a junction table, where the junction table references both the primary key for the `INDV` and `SOURCE` tables. Any given combination of `INDV` and `SOURCE` foreign keys is unique and is then assigned its own primary key.

While not pursued in the development of INSSEPT, a key advantage of RDBs is their ability to integrate different types of data from different sources seamlessly. Several RDBs (from different servers) can be linked with minimum effort to update and maintain information for a broader framework. For example, an RDB might contain information from building taxonomy and seismic performance, whereas another RDB (hosted at a different server), could provide information on the structure's performance under hurricane. The data from both RDBs can be extracted and formatted as a new RDB for multi-hazard performance assessment.

3.2 Schema overview

The INSSEPT schema was developed following an extensive literature review of published PBEE results, existing related schemas [20,21,22], and core tenets of RDB design [23,24] . Several iterations of classification and assessment of candidate fields for concise and efficient representation were required. In this regard, INSSEPT is structured following three normalization rules. First, data should be reduced to its simplest form. For example, a field such as '`AuthorName`' should be presented as '`FirstName`' and '`LastName`', as opposed to a single field containing both first name and last name, and noting that 'Given' and 'Family' names cannot be practically inferred from the available bibliographic data. The second rule requires that attributes within a table be dependent on the primary key. For example, In the `SOURCE` table, the values of `DOI, Citation, and Year`, depends on Source table's primary key, `SourceID`. The third normalization rule requires that attributes are independent of one another. For example, in the `BUILDING` table, there are attributes for floor area (i.e. `FlrArea`) and number of stories (i.e. `Stories`). It would be redundant to include another attribute for total area, as it can be deduced by simply multiplying the values of the two aforementioned attributes.

Figure 3 shows the complete INSSEPT schema, which prioritizes intuitiveness and practicality. In this regard, INSSEPT is centered on the `BUILDING` table because this table higher relevance to early design comparing to other information such as references or the type of numerical model.

INSSEPT's tables can be categorized into four broad categories: bibliographic, building taxonomy, PBEE assessment, and results. Bibliographic tables consist of `SOURCE` (reference/citation), `LOG` (history of database updates), `AUTHLIST` (authorship), `INDV` (individuals contributing to sources or the database), and `AVAIL` (the type of information in the reference). Building taxonomy comprises `BLDGMAN` (a linking table between the building and its design manuals/building codes), `MANUAL` (e.g., material specifications), `CATOCC` (categorized use and occupancy), `MATSYS` (primary building material and structural system), `MATCLASS` (material), and `SITE` (location). PBEE assessment tables cover

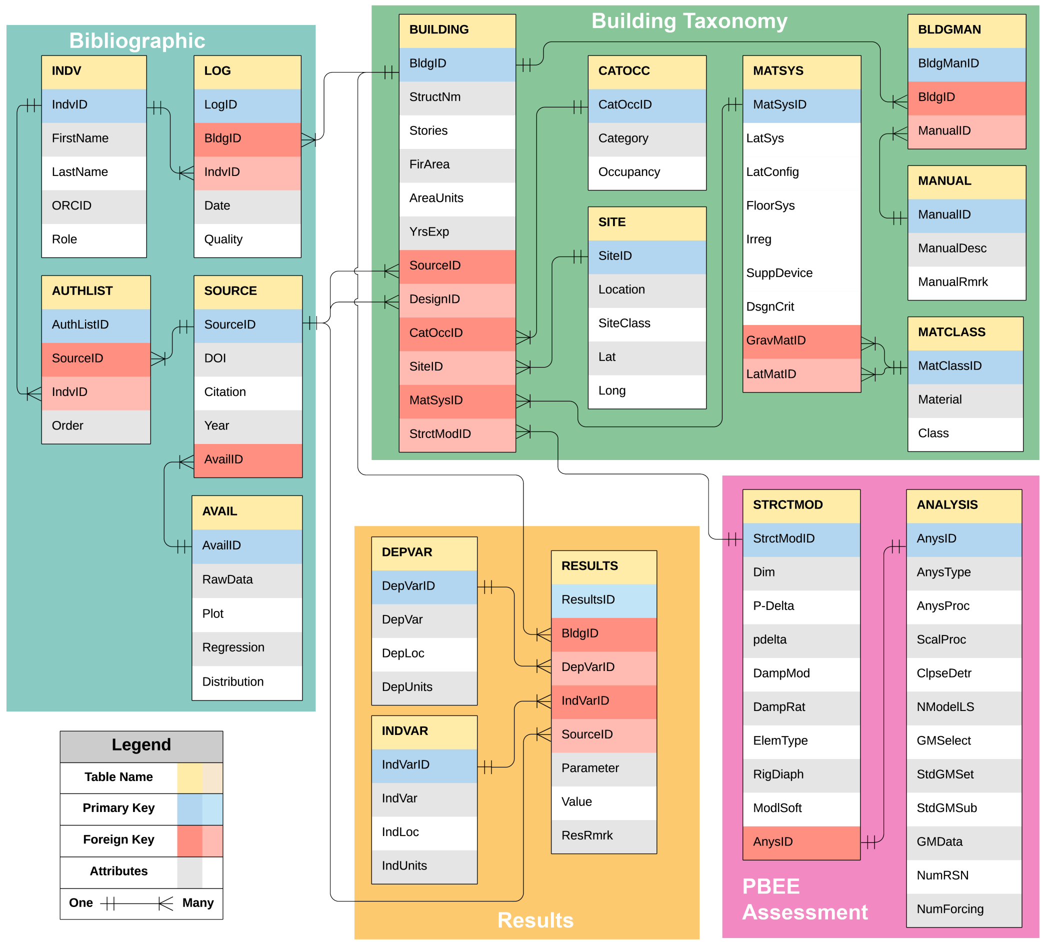


Fig. 3: INSSEPT schema and grouping of tables by the type of contained data; building taxonomy can be considered metadata, whereas PBEE assessment is the core performance data. Additional information on INSSEPT tables and fields is provided in the manual document [25] which is accessible at the DesignSafe repository

STRCTMOD (numerical structural model) and ANALYSIS (type of PBEE assessment, e.g., IDA). Lastly, Results tables include DEPVAR (type of dependent variable), INDVAR (type of independent variables of analysis), and RESULTS (numerical values of assessment).

An entry in the BUILDING table represents the PBEE evaluation of a unique structural system design (i.e., component sizing, geometry, and connections). The attributes housed in this table are those that are non-repetitive across structures and attached to a specific building/structure. That is, while the value of attributes such as Stories may not be unique across all buildings studied, each building has only one value of this attribute. The repetitive attributes are grouped into inter-related tables and are linked to the BUILDING table using foreign keys. Some studies consider multiple variants of a structure with the same geometry and loading; in such cases, the result of each analysis is separately recorded and indexed using AnysID key.

The schema of INSSEPT could have been more complicated due to the variability of PBEE assessment methods, scopes, and terminology. However, it was decided to direct database development efforts towards standardizing allowable entries in data collection through quality control and assurance, rather than relying on a large number of many-to-many relationships.

3.3 Description of INSSEPT main tables

INSSEPT consists of sixteen tables. The description of the four most important tables and corresponding fields are briefly discussed as follows:

1. **BUILDING** serves as the core table of INSSEPT and contains general taxonomy of the buildings and foreign keys to other tables such as reference documents (SOURCE), building site information (SITE) and structural modeling information (STRCMOD). In this table, StructNm is the structure name (identifier); Stories and FlrArea AreaUnits capture the unit used for FlrArea field such ft^2 or m^2; YrsExp defines the number of years that structure is assumed exposed to the environment; MatSysID is a foreign key to the material used in the structure; CatOccID is the foreign key to category and occupancy table; SiteID is the foreign key to building's site table; SourceID is the foreign key to references table and DesignID is the foreign key to structural design source in case it is different from the source that contains the PBEE assessment.
2. **MATSYS** presents information on material and types of the building's structural lateral290 resisting and gravity systems. MatSysID is the identifier key for the combination of the used material and systems; LatSys is the type of lateral-resisting system;LatConfig captures spatial configuration of the lateral system (such as perimeter);FloorSys is the type of floor system used in the building; Irreg provides any type of irregularity of the building (such as horizontal or torsional); SuppDevice describes any supplemental seismic devices, such as dampers or base isolation, DsgnCrit lists all the special criteria that used to design the building (e.g., a specific beam to column strength ratio); LatMatID and GravMatID are foreign keys to the material-class MATCLASS table for lateral and gravity systems.
3. **STRCTMOD** contains information on features of the building's numerical structural model. StrctModID is an identifier key for the model; Dim indicates whether model is two- or three-dimensional; P-Delta and p-delta capture whether large and small second-order geometric effects are captured in the numerical model; DampMod and DampRat records damping model (such as Rayleigh) and damping ratio of the model; the describes which elements are used to model building components, and is categorized as linear, concentrated plasticity, distributed plasticity and continuum FE;RigDiaph indicates whether floors are assumed to act as rigid diaphragms;ModlSoft lists all software used in the analysis, such as OpenSees, Code Aster or Zeus-NL.
4. **ANALYSIS** provides information on the type of PBEE assessment conducted on the structure. AnysType describes the general type of structural analysis and can be categorized as nonlinear dynamic, linear dynamic, nonlinear static; AnysProc records the procedure used to link ground motions to structural response, including methods such as incremental dynamic, cloud or multiple stripe; ClpseDetr indicates how collapse is determined in the numerical model (e.g. using pre-defined limit states); NmodelLS field tracks all non-modeled limit states not directly captured in the numerical model but included in the performance assessment, (e.g., results are post317 processed for connection fracture, which is linked to a damage state); GMData captures the database/source of GM records such as PEER NGA-WEST or European Strong motions databases; GMSelect is an explanatory field to describe the GM selection, e.g., based on casual parameters or other methods; ScalProc discusses scaling of records and covers methods such as scaling to conditional or uniform hazard spectra; StdGMSet and StdGMSub captures the name of any applicable standard GM sets and their subsets such as FEMA P-695 far-field; NumRSN is the number of unique record sequence numbers, using the terminology of NGA-West2[26] (i.e.,different recording stations for the same earthquake event have different RSN values); whereas NumForcing represents the number of GM variants used as input accelerations to the numerical model (i.e., two forcings per RSN if NS and EW directions are used).

The organization of the RESULTS table is discussed subsequently, as it is best interpreted in the context of an illustrative query. Documentation with description of the remaining tables is provided in DesignSafe[16]. This documentation [25] also provides a complete list of applicable data types for each field with several examples and a full bibliographic list of the reviewed papers.

3.4 Extraction of PBEE results and incorporation in INSSEPT

Data collection for INSSEPT was conducted as part of a summer research experience for undergraduates (REU) supported by the National Science Foundation. Five undergraduate students underwent training led by the RSB initiative faculty and graduate students to gain familiarity with PBEE concepts, RDBs, and goal-oriented literature review. The effort made to compile INSSEPT can be viewed as two separate parts: (1) efforts aimed to measure and assure the quality of the database (quality assurance) and (a) systematic control of the database during and after compiling to ensure its quality (quality control). For quality assurance, attention was given to ontology (relationships between types of data) development by adjusting and revising proposed fields and optimizing INSSEPT schema to be clear and intuitive for structural engineers. From a quality control perspective, the collected information was cross checked by the RSB team through an iterative process, where several passes were made to remove any errors in the collected data.

Papers were identified in a collaborative setting using scientific search engines such as Google Scholar, Science Direct, and Engineering Village. The collected papers were then reviewed as candidates for inclusion in INSSEPT, and were deemed eligible based on data completeness (providing sufficient performance results, availability of building description, etc.) and freshness (no earlier than 2005). Additionally, papers were preferred that would increase the diversity of the database and were originated from reputable sources, such as the ASCE library. The current version of the database includes the results of 39 studies from more than a hundred authors. The PBEE assessments of 222 analytical models of different buildings are assimilated (1211 results values), where 144, 30 and 24 of the building models have moment frame, braced frame and shear walls lateral-resisting systems, respectively. The scope of these studies covered a variety of locations including the US (48 in California, 40 elsewhere), Europe (25;6 Italy) and Asia (14 Iran, 4 India, 16 China).

A major issue in PBEE data collection was the lack of a completely standardized terminology, where different authors might refer to the same concept through different terms. One author might refer to a type of nonlinear modeling approach as a "plastic hinge" model, where another author would prefer "concentrated plasticity". Such descriptions were included verbatim in preliminary database development and subsequently standardized to reduce possible confusion, as well as facilitate the easier application of machine learning-based algorithms. However, the performance inventory database should allow for some natural variability of PBEE assessments, which led to the creation of attributes specifically designed to capture the unique features of each study. For example, `DesignCrit` records any additional information regarding structural design that is not captured by the standardized fields., e.g., the use of "connection stiffeners".

While effort was made to select papers with sufficient data to fill all the necessary attributes, the absence of required data is inevitable, particularly for PBEE assessments where a consensus on the scope and detail of output results are yet to be established. A `Null` value was used in cases where data was missing. It should be noted that `Null` cases are different from `None` or `0`, which refer to fields that do not specifically share that attribute. For example, if a structure does not have any irregularity, the `Irreg` attribute will be `None`, whereas if the paper does not provide any information on irregularity, and the absence or presence of an irregularity cannot be inferred from the building description, a `Null` value is assigned.

A `LOG` table is developed to maintain a transparent quality control procedure. This table records each time an update was made to a building, giving information on the modified building, individuals' initials, date, and the type of modification (Quality attribute).

3.5 User interaction through query

Querying combines data that has been disaggregated into multiple tables and fields to form a comprehensive display of results while offering the user the ability to filter results based on selected parameters. If the creation of an RDB can be compared to dismantling an expansive set of data as shown in Figure 2, querying is analogous to reconstructing the dismantled data to form a bespoke table that caters to the user's needs. This section focuses on topics that are related to most likely use cases of INSSEPT; interested readers are referred to numerous resources that cover SQL querying in detail (such as [23,24,27]).

Figure 4's sample query creates a table of buildings taller than 3 stories and located on soil site class D. Querying commands in MySQL are based on the `SELECT` statement, which allows the user to select desired attributes from the RDB. All attributes in MySQL are denoted following the convention `TableName.AttributeName`. For instance, `BUILDING.YrsExp` represents the attribute `YrsExp` (Years of Exposure) in the `BUILDING` table. In the case where the user wishes to select all attributes of a table, the attribute name can be replaced with an asterisk (e.g. `SITE.*`). The `SELECT` statement contains a default FROM clause, which specifies the main table to which auxiliary data and tables will be appended.

The `SELECT` statement alone cannot fully accomplish the two purposes of combination and filtering discussed above, and is usually accompanied by two clauses that execute the respective tasks:

1. The combination of tables in RDB is formally referred to as performing a "join". Multiple types of joins are possible, but inner joins are most suitable matching primary and foreign keys. If the foreign key of one table (e.g. `BUILDING.SiteID`) calls for additional information, the information is retrieved by locating the primary key in a corresponding table (e.g. `SITE.SiteID`). The `INNER JOIN` clause specifies the table that will be appended; `ON` contains the matching condition (e.g. `BUILDING.SiteID = SITE.SiteID` in Figure 4). Multiple tables can be appended to the main table with the use of additional `INNER JOIN` clauses.
2. Filtering of data is accomplished using the WHERE clause, which can contain multiple conditions provided that they pertain to attributes in the main table or appended tables. Standard relational operators (=, <=, >, etc.) and logical operators (AND, OR, NOT) are used to specify and compound each condition. In our case, `BUILDING.Stories > 3` and `SITE.SiteClass = "D"` are combined with the `AND` operator to only select buildings that are more than three stories and are located at site class D.

A portion of the resulting output is displayed in Figure 5, where the `CATOCC` and `SITE` tables have been appended to the main BUILDING table. Selected attributes in the `BUILDING` table and all attributes of the appended tables have been displayed, as specified by the querying script. Another form of querying involves many-to-many relationships. As discussed in Section 3, many-to-many relationships in RDB are accounted for using junction tables that combine the foreign keys of the respective tables. Figure 6 shows one such query example for `RESULTS` table.

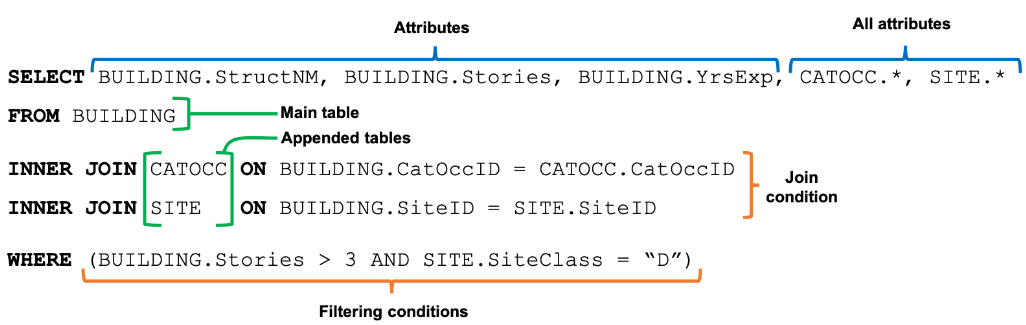


Fig. 4: Structure of a sample SQL query applicable to INSSEPT

	StructNM	Stories	YrsExp	CatOccID	Category	Occupancy	SiteID	Location	SiteClass	Lat	Long
0	A2011-NS-S7-BRB	7	0	1	None	None	4	CA	D	NaN	NaN
1	A2011-NS-S7-BRBw/moment	7	0	1	None	None	4	CA	D	NaN	NaN
2	A2011-NS-S7-SMRF	7	0	1	None	None	4	CA	D	NaN	NaN
3	A2011-NS-S7-SMRFw/moment	7	0	1	None	None	4	CA	D	NaN	NaN
4	U2004-NS-S6-CONVENTIONAL	6	0	1	None	None	5	Los Angeles, CA	D	NaN	NaN
5	U2004-NS-S6-BRB	6	0	1	None	None	5	Los Angeles, CA	D	NaN	NaN
6	J2011-Res-S7	7	0	2	Residential	Residential	3	Tehran, Iran	D	NaN	NaN
7	J2011-Res-S15	15	0	2	Residential	Residential	3	Tehran, Iran	D	NaN	NaN
8	V2013-Res-S4	4	0	2	Residential	Residential	10	None	D	NaN	NaN
9	V2013-Res-S5	5	0	2	Residential	Residential	10	None	D	NaN	NaN
10	V2013-Res-S6	6	0	2	Residential	Residential	10	None	D	NaN	NaN

Fig. 5: INSSEPT output for sample query shown in Figure 4

The `RESULTS` table is a three-way junction table that connects the `BUILDING, INDVAR` (independent variable), and `DEPVAR` (dependent variable) tables. Therefore, querying in a many-to-many relationship is a multi-step process, with each step involving the join of one table to the junction table. The first join is initiated by appending the junction table (`RESULTS`) to the main table (e.g. `BUILDING`). All subsequent joins are made by appending the desired table (`INDVAR, DEPVAR`) to the junction table rather than the main table. Figure 7 shows the

INSSEPT output. It should be noted that the `RESULTS` table accommodates all types of numerical data through the combination of independent (`INDVAR`) and dependent (`DEPVAR`) tables. The rationale behind this separation is to increase the flexibility of types of the results that INSSEPT can host, which include, e.g., PSDA regression parameters, fragility curves, and numerical model period. For example, structural first-mode period is single value and can be stored by assigning "period" to `DEPVAR.DepVar` and its value in `RESULTS.value`. A regression formula might consists of `DEPVAR.DepVar = transient interstory drift ratio, .DepLoc= maximum (across stories)`, and `DEPVAR..DepUnits = in/in`. The associated `INDVAR.IndVar` might be spectral acceleration, with `.IndLoc = structure first-mode elastic` and `.IndUnits = g`. Regression coefficients (slope and intercept) would be recorded as separate entries in `RESULTS.parameters` and `RESULTS.values`. This RDB structure is significantly more flexible than can be accommodated based on predefined combinations (e.g. IDR-Sa(T_1)). Additional queries and a complete description of these tables' descriptions are provided in supplementary documentation[25].

```
SELECT BUILDING.StructNm,  RESULTS.Parameter, RESULTS.Value,
INDVAR.IndVar, INDVAR.IndLoc, INDVAR.IndUnits,
DEPVAR.DepVar, DEPVAR.DepLoc, DEPVAR.DepUnits
FROM BUILDING
INNER JOIN RESULTS ON BUILDING.BldgID = RESULTS.BldgID
INNER JOIN DEPVAR ON RESULTS.DepVarID = DEPVAR.DepVarID
INNER JOIN INDVAR ON RESULTS.IndVarID = INDVAR.IndVarID
ORDER BY BUILDING.StructNm, DEPVAR.DepVar;
```

Main table

Junction table appended to main table

Additionals table appended to junction table

Fig. 6: Querying using junction table

4 Illustrative examples

This section presents several possible applications of INSSEPT. While not comprehensive, the examples demonstrate the advantages of compiling PBEE data to exploit the state of knowledge in hazard impact modeling.

	StructNm	Parameter	Value	IndVar	IndLoc	IndUnits	DepVar	DepLoc	DepUnits
0	A2010-Res-S10	Dispersion	0.297	PGA	None	g	Collapse Prevention (CP)	Global	None
1	A2010-Res-S10	Median	1.200	PGA	None	g	Collapse Prevention (CP)	Global	None
2	A2010-Res-S10	Dispersion	0.197	PGA	None	g	Collapse Prevention (CP)	Global	None
3	A2010-Res-S10	Median	0.600	PGA	None	g	Collapse Prevention (CP)	Global	None
4	A2010-Res-S10	Median	0.360	PGA	None	g	Immediate Occupancy (IO)	Global	None
5	A2010-Res-S10	Dispersion	0.080	PGA	None	g	Immediate Occupancy (IO)	Global	None
6	A2010-Res-S10	Median	0.300	PGA	None	g	Immediate Occupancy (IO)	Global	None
7	A2010-Res-S10	Dispersion	0.128	PGA	None	g	Immediate Occupancy (IO)	Global	None
8	A2010-Res-S10	None	1.440	None	None	None	Period	1	s
9	A2010-Res-S10	None	1.550	None	None	None	Period	1	s

Fig. 7: Output of sample Query shown in Figure 6

4.1 Comparing different lateral system alternatives for a specific location

Figure 8 shows different lateral resisting systems for a mid-rise buildings (3 to 8 stories) located in California that are compared in terms of their median collapse capacity. Code 1.1 query syntax is used and 28 results are extracted which were subsequently categorized as steel bucking restrained braced frame (BRB), steel braced frame, steel moment-resisting frame (MRF), concrete MRF and concrete non-ductile frame.

While non-ductile frames are not an alternative considered for the high seismic region of California, they provide a basis to compare seismically designed modern structures to older construction. Several observations can be made from the results shown in Figure 8. For the same number of stories, concrete MRFs show significantly larger median collapse capacity than non-ductile concrete frames, in line with expectations. Within concrete MRFs, taller frames have smaller median collapse capacity, which could be because these structures are designed for a lower seismic force comparing to shorter ones following recent building codes For Peer Review provisions [28]. INSSEPT provides opportunities to further examination by querying structural period and checking the ratio of median collapse fragilities in terms of spectral acceleration to the designed spectral acceleration from code's spectrum. Lastly, while the limited number of studies prevent general conclusions, it can loosely be interpreted that buckling restrained frames shows better collapse performance than other included systems and should be considered in possible SFSE systems for this site. This is the main use case that INSSEPT is developed for and readers are encouraged to refer to Flint et al. [29] for a more detailed example of this application. However, it should also be noted that as PBEE results are strictly dependent on the assessment assumptions, which can vary widely, users are advised not to solely rely on the result values from the query. Instead, care should be taken to evaluate all pertinent information (such as design, modeling and analysis data presented in the fields such as `DesignCrit, ElemType, AnysProc`) before any interpretation is made.

Code 1.1: SQL query to extract median collapse fragility for mid-rise structures located in California

```
SELECT BUILDING.StructNm, RESULTS.Parameter, RESULTS.Value,
INDVAR.IndVar, INDVAR.IndUnits, MATSYS.LatSys,DEPVAR.DepVar,
 BUILDING.Stories, SITE.Location, SITE.SiteClass
FROM BUILDING
INNER JOIN MATSYS ON BUILDING.MatSysID=MATSYS.MatSysID
INNER JOIN SITE ON BUILDING.SiteID = SITE.SiteID
INNER JOIN RESULTS ON BUILDING.BldgID = RESULTS.BldgID
INNER JOIN DEPVAR ON RESULTS.DepVarID = DEPVAR.DepVarID
INNER JOIN INDVAR ON RESULTS.IndVarID = INDVAR.IndVarID
WHERE ((SITE.Location LIKE "CA%" OR SITE.Location LIKE "LOS%")AND BUILDING.Stories > 2 AND
    BUILDING.Stories <9 AND DepVar LIKE "Col%" AND parameter="Median");
```

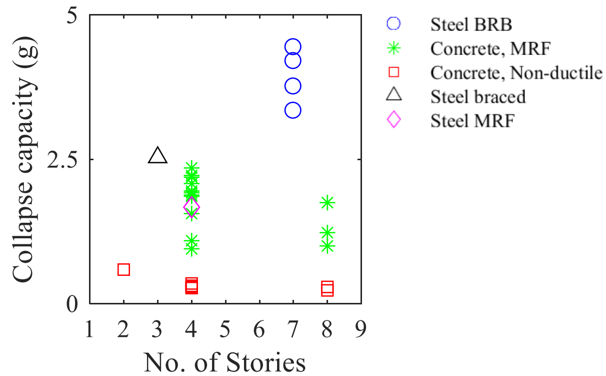


Fig. 8: Sample collapse median for midrise buildings in California

4.2 Investigating seismic vulnerability for central and eastern US

Figure 9 shows the median fragility values at near-collapse (corresponding to collapse prevention and complete damage states in FEMA 273 and HAZUS guidelines) for the Central and Eastern US building stock provided in INSSEPT using Code 1.2 syntax. First, it is observed that the results have more dispersion than the previous example and a clear distinction between different lateral systems cannot be made. With the exception of unreinforced masonry, buildings without lateral-resisting system (i.e. those which explicitly state that there is only gravity framing and/or are not seismically-designed) have higher period (lower stiffness). In addition, on average steel braced and MRF show better seismic performance than other systems. The results are not surprising as most of the structures in this region are designed for relatively low lateral wind forces and do not comply with more rigorous requirements of the seismic regions. Therefore, the lack of seismic design leads to structures yielding inconsistent PBEE results. The results of URM high performance is contrary to engineering judgement, and is likely attributable to design assumptions and modeling choices made in the original papers. Indeed, this observation demonstrates the need for rigorous evaluation prior to making any interpretation of results. In addition, this example could be perceived as a possible application of future performance inventories, where a large number of structures in a particular region could be leveraged for regional loss and portfolio assessments. An important issue in regional seismic loss assessments is to select fragility data that can reasonably represent the range of buildings in the region from developed fragilities inventories [30], which can be conveniently accomplished using performance inventories

Code 1.2: SQL query to extract median fragility values for structures in Central Eastern US

```
SELECT BUILDING.StructNm, RESULTS.Parameter, RESULTS.Value,INDVAR.IndVar, INDVAR.IndUnits,
    MATSYS.LatSys,DEPVAR.DepVar, BUILDING.Stories, SITE.Location, SITE.SiteClass
FROM BUILDING
INNER JOIN MATSYS ON BUILDING.MatSysID=MATSYS.MatSysID
INNER JOIN SITE ON BUILDING.SiteID = SITE.SiteID
INNER JOIN RESULTS ON BUILDING.BldgID = RESULTS.BldgID
INNER JOIN DEPVAR ON RESULTS.DepVarID = DEPVAR.DepVarID
INNER JOIN INDVAR ON RESULTS.IndVarID = INDVAR.IndVarID
WHERE ((SITE.Location LIKE "Cent%" OR SITE.Location LIKE "Mem%") AND (DepVar LIKE "Col%"
    OR DepVar LIKE "com%" ) AND parameter="Median");
```

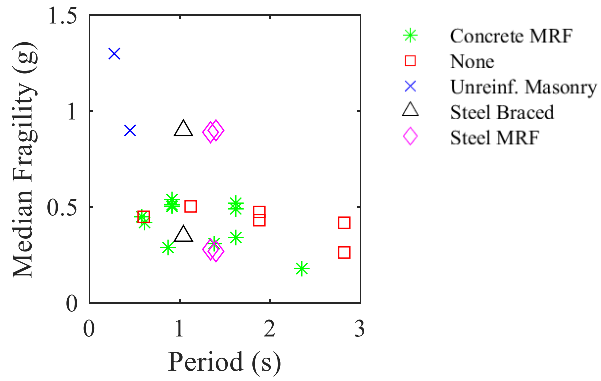


Fig. 9: Sample near-collapse fragility results for buildings in the Central Eastern US

4.3 Leveraging additional performance data to assess different structural systems for a particular site

While median fragility collapse capacity data reflects a structure's expected behavior under extreme seismic loads, understanding a range of behavior from elastic to collapse supports more informed decision making. SQL code syntax 1.3 is used to access cloud analysis data for a 4-6 story mid-rise structures in the central and eastern US. The results are saved as a CSV file 501 using OUTPUT command in SQL[1] .

The results of this query capture linear slope (a1) and intercept (a0) of EDP-IM relationship in 503 log-log space, as well the model standard error () (as shown in Figure 1.b), which can readily be used to derive probability of exceeding a certain drift level. For example, for a 4-story steel MRF (K2007-NS-S4) the query returns $a_0 = -2.89$, $a_1 = 1.10$ and $= 0.25$, whereas for a 6-story X braced (cross-braced; E2009-NS-S6-xbraced) it returns $a_0 = -3.52$, $a_1 = 0.88$ and $= 0.25$. Using this information, the conditional probability of exceeding a certain drift level given ground 508 motion intensity (i.e. fragility curve) can be computed. For instance, with the MRF:

$$P(MIDR > midr|S_a = s_a) = 1 - \Phi(\frac{\ln midr - (a_0 + a_1 \ln s_a)}{\beta}) = 1 - \Phi(\frac{\ln midr - (2.89 + 1.1 \ln S_a)}{0.25}) \tag{1}$$

Figure 10 compares the fragility curves of the MR (K2007-NS-S4) and X-braced (E2009-NS-S6-xbraced) frames. The fragilities are normalized to the corresponding Sa demand at the design basis event (DBE) hazard level. At lower drift levels (i.e. 0.5% drift level, Fig.10a and 1% drift, Fig. 10b) the systems show similar performance, whereas at higher drift levels (i.e. 3% drift, Fig.10c) the braced frame outperforms the moment-resisting system. Alternately, the distribution of drifts given Sa could be computed for a design check— as was performed in Flint et al. [29] to identify feasible systems in early design, as indicated in Fig. 10.d. Both systems have very low probabilities of exceeding 2% drift at the design basis event (DBE).

Code 1.3: SQL query to extract median fragility values for structures in Central Eastern US

```
SELECT BUILDING.StructNm, RESULTS.Parameter, RESULTS.Value,INDVAR.IndVar, INDVAR.IndUnits,
    MATSYS.LatSys,DEPVAR.DepVar, BUILDING.Stories, SITE.Location, SITE.SiteClass,
    MATSYS.LatSys
INTO OUTFILE 'output.csv'
FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'
LINES TERMINATED BY '\n'
```

[1] In the accompanying Jupyter notebook [16], the CSV is instead directly generated using Python's Panda package.

```
FROM BUILDING
INNER JOIN MATSYS ON BUILDING.MatSysID=MATSYS.MatSysID
INNER JOIN SITE ON BUILDING.SiteID = SITE.SiteID
INNER JOIN RESULTS ON BUILDING.BldgID = RESULTS.BldgID
INNER JOIN DEPVAR ON RESULTS.DepVarID = DEPVAR.DepVarID
INNER JOIN INDVAR ON RESULTS.IndVarID = INDVAR.IndVarID
WHERE ((SITE.Location LIKE "%East%") AND (DepVar="MIDR" ) AND (parameter LIKE "Slope-log%"
    OR parameter LIKE "intercept-log%" OR parameter LIKE "Disp%")AND BUILDING.Stories > 3
    AND BUILDING.Stories <7);
```

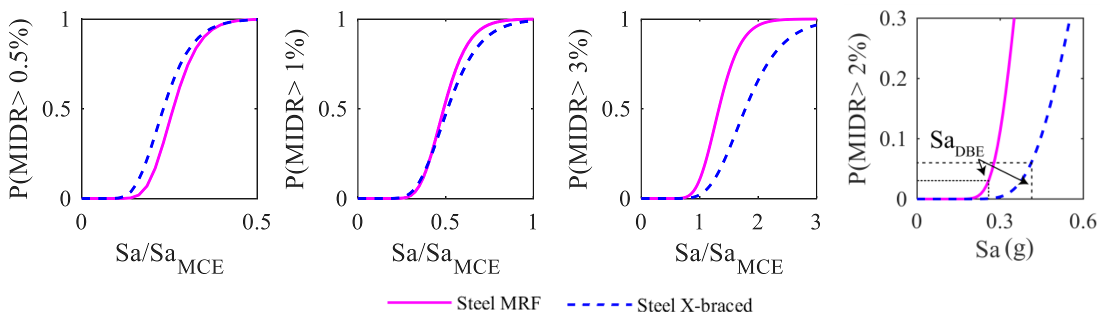


Fig. 10: Comparison of steel MRF and X-braced system at different drift levels of (a)0.5%, (b) 1% and (c) 3%. Figure d shows the normalization Sa levels at DBE for both systems

4.4 Identifying trends in modeling

Figure 11 shows the histogram of analysis procedure (cloud versus IDA) and element type concentrated versus distributed plasticity models) for the compiled references using Code 1.4 syntax. Overall, IDA and concentrated plasticity models are more prevalent in the database, with the number of studies peaking from 2010 to 2012. IDA studies appear to be declining after 2012, which could be attributed to the publication of major PBEE projects such as FEMA P695. A trend is not as easily identifiable for nonlinear modeling approach, however, it is expected to find more papers using concentrated plasticity elements to model mid-rise buildings, whereas other types of structures such as gravity dams or bridge piers are commonly modeled through distributed plasticity and continuum finite element models. It should be emphasized that this example is provided to demonstrate how a meta-analysis can be performed. Such analysis paves the road for machine learning algorithms applications. For example, a clustering technique might be used to assess what type of ground motion selection procedure is used for different parts of US, or which modeling techniques are popular for steel moment-resisting frames.

Code 1.4: SQL query to compare structural element types and analysis methods

```
SELECT SOURCE.*, BUILDING.StructNM, STRCTMOD.ElemType
FROM BUILDING
INNER JOIN SOURCE ON BUILDING.SourceID = SOURCE.SourceID
INNER JOIN STRCTMOD ON BUILDING.StrctModID = STRCTMOD.StrctModID
WHERE STRCTMOD.ElemType LIKE "%plastic%";

SELECT SOURCE.Citation, SOURCE.Year,BUILDING.StructNM, ANALYSIS.AnysProc
FROM BUILDING
INNER JOIN SOURCE ON BUILDING.SourceID = SOURCE.SourceID
INNER JOIN STRCTMOD ON BUILDING.StrctModID = STRCTMOD.StrctModID
INNER JOIN ANALYSIS ON STRCTMOD.AnysID = ANALYSIS.AnysID
```

```
WHERE ANALYSIS.AnysProc LIKE "Cloud%" OR ANALYSIS.AnysProc LIKE "IDA%";
```

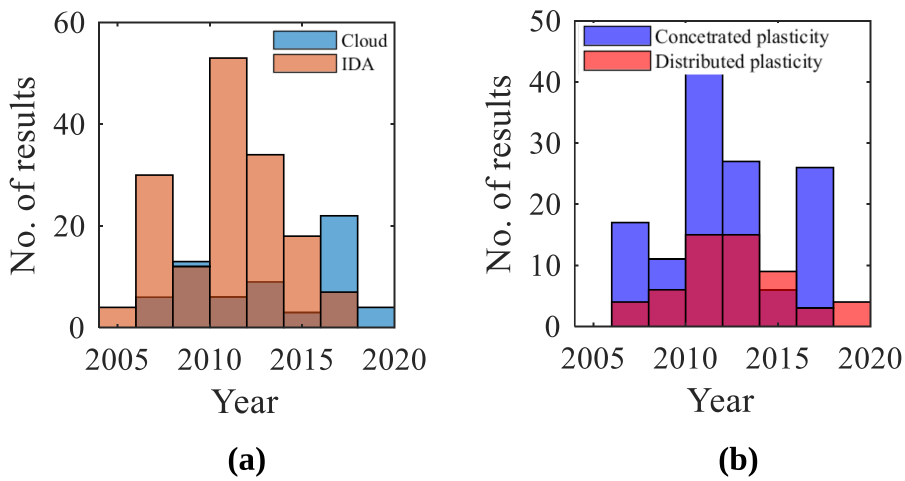


Fig. 11: Histograms of (a) different analysis procedure and (b) nonlinear modeling

5 Conclusion

This paper summarizes the efforts and conceptual framework behind developing a seismic performance inventory, INSSEPT, from existing PBEE literature. The key concept of INSSEPT lies in leveraging collective past efforts of PBEE community to arrive at a more risk informed future. INSSEPT's intuitive organization is founded on structural engineering principles while being consistent with relational databases norms. INSSEPT is expected to be applicable for early design of structures through comparison of different lateral systems, as well as regional seismic assessment by providing standardized fragility data for a specific location. A glimpse of INSSEPT's possible applications is provided through some examples.

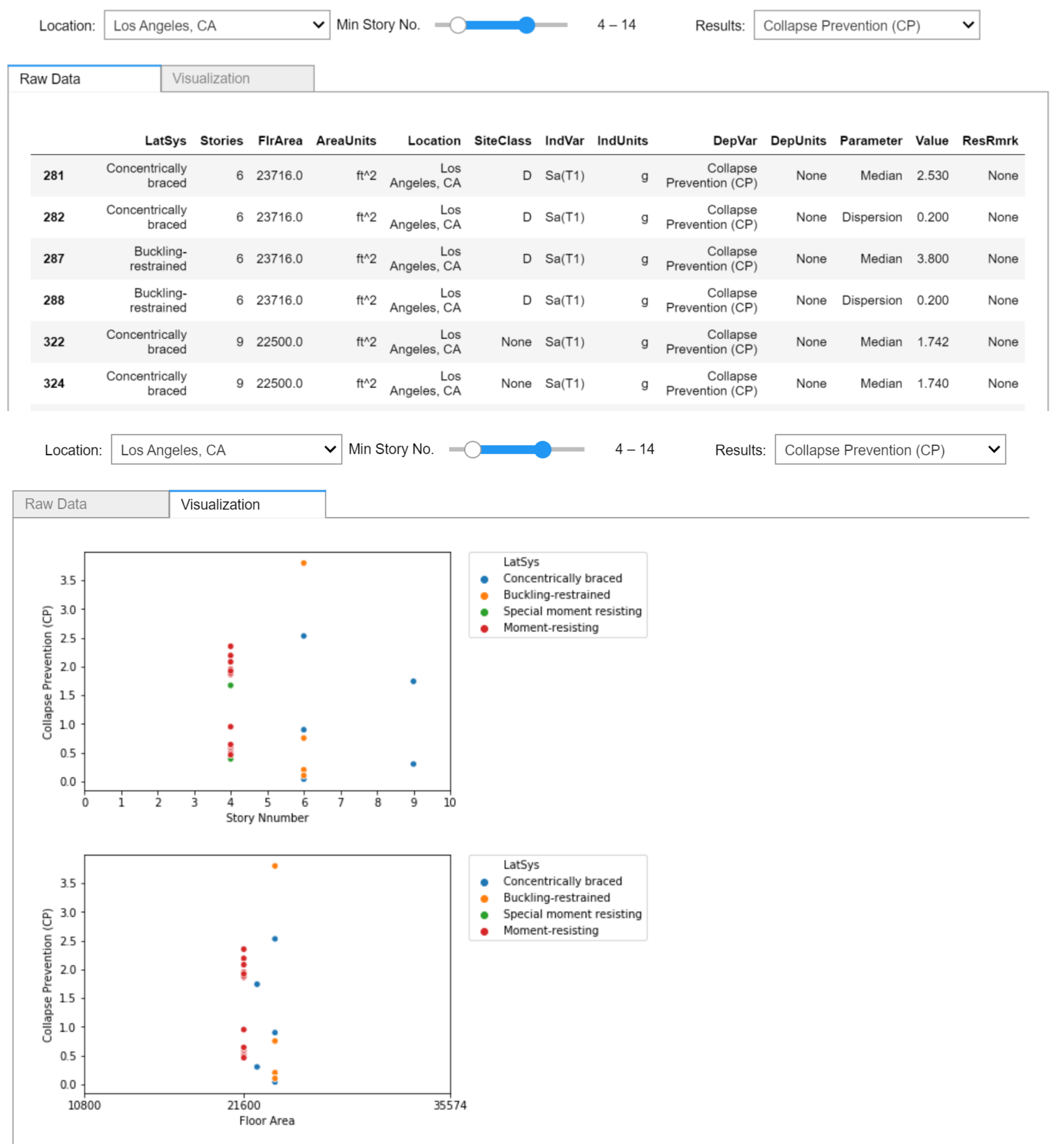
It should be noted that while the current version of INSSEPT is static and does not allow additional data to be added by other users, the schema is made public to facilitate future versions that can be maintained and updated through collaborative efforts. In addition, the scope of INSSEPT can easily be extended to other types of structures and natural hazards. Perhaps the more important takeaway of this paper is the opportunities that are possible with increased standardization and organization of data in PBEE and the natural hazard engineering community. In this regard, INSSEPT is a small effort for a future vision where structural engineering researchers and practicing engineers have broad access to high quality datasets and can benefit from advances in machine learning and other computational frameworks, perhaps reminding the readers of INSSEPT's homonym, "incept".

Acknowledgement

This research is supported by the US National Science Foundation, Division of Civil, Mechanical, and Material Innovation, through Resilient Sustainable Buildings Award 1455466. Diep Ton, Natalia Zhukova, Patricia Asiatico, Elaine-Joy Consolacion, Hasib Tahir and Hyeeun Kong are also thanked for their contributions to data collection and revision. The team at DesignSafe is also thanked for their valuable input into the database schema and assistance with hosting the database.

APPENDIX 1: INSSEPT interactive tool

An interactive Jupyter notebook is provided as part of supporting documents of INSSEPT. This notebook aids user to quickly extract data based on location, number of floors and results type, without the need to perform any SQL query. A screenshot of the tool is shown below.



	LatSys	Stories	FlrArea	AreaUnits	Location	SiteClass	IndVar	IndUnits	DepVar	DepUnits	Parameter	Value	ResRmrk
281	Concentrically braced	6	23716.0	ft^2	Los Angeles, CA	D	Sa(T1)	g	Collapse Prevention (CP)	None	Median	2.530	None
282	Concentrically braced	6	23716.0	ft^2	Los Angeles, CA	D	Sa(T1)	g	Collapse Prevention (CP)	None	Dispersion	0.200	None
287	Buckling-restrained	6	23716.0	ft^2	Los Angeles, CA	D	Sa(T1)	g	Collapse Prevention (CP)	None	Median	3.800	None
288	Buckling-restrained	6	23716.0	ft^2	Los Angeles, CA	D	Sa(T1)	g	Collapse Prevention (CP)	None	Dispersion	0.200	None
322	Concentrically braced	9	22500.0	ft^2	Los Angeles, CA	None	Sa(T1)	g	Collapse Prevention (CP)	None	Median	1.742	None
324	Concentrically braced	9	22500.0	ft^2	Los Angeles, CA	None	Sa(T1)	g	Collapse Prevention (CP)	None	Median	1.740	None

References

1. M. Flint, L. Dhulipala, Y. Shahtaheri, H. Tahir, T. Ladipo, MR. Eatherton, JL. Irish, CG. Olgun, G. Reichard, A. Rodriguez-Marek, et al. Developing a decision framework for multi-hazard design of resilient, sustainable buildings. In *1st Int. Conf. on Natural Hazards and Infrastructure (ICONHIC2016)*, 2016.
2. G.M. Calvi, R. Pinho, G. Magenes, J.J Bommer, L.F Restrepo-Vélez, and H. Crowley. Development of seismic vulnerability assessment methodologies over the past 30 years. *ISET journal of Earthquake Technology*, 43(3):75–104, 2006.
3. V. Silva, H. Crowley, H. Varum, R. Pinho, and Ro. Sousa. Evaluation of analytical methodologies used to derive vulnerability functions. *Earthquake Engineering & Structural Dynamics*, 43(2):181–204, 2014.
4. M. Colombi, B. Borzi, H. Crowley, M. Onida, F. Meroni, and R. Pinho. Deriving vulnerability curves using italian earthquake damage data. *Bulletin of Earthquake Engineering*, 6(3):485–504, 2008.
5. C.A. Kircher, R.V. Whitman, and W.T. Holmes. Hazus earthquake loss estimation methods. *Natural Hazards Review*, 7(2):45–59, 2006.
6. S. Molina, D.H. Lang, and C.D. Lindholm. Selena–an open-source tool for seismic risk and loss assessment using a logic tree computation procedure. *Computers & Geosciences*, 36(3):257–269, 2010.
7. AC. Costa, ML. Sousa, A. Carvalho, and E. Coelho. Seismic loss scenarios based on hazard disaggregation. application to the metropolitan region of lisbon, portugal. In *Assessing and managing earthquake risk*, pages 449–462. Springer, 2008.
8. C. Yepes-Estrada, V. Silva, T. Rossetto, D. D'Ayala, I. Ioannou, A. Meslem, and H. Crowley. The global earthquake model physical vulnerability database. *Earthquake Spectra*, 32(4):2567–2585, 2016.
9. C. Yepes-Estrada, V. Silva, and H. Crowley. Gem vulnerability database for the openquake-platform. In *Proceedings of 2nd European conference on earthquake engineering and seismology, Istanbul*, pages 25–29, 2014.
10. T. Rossetto, I. Ioannou, and DN. Grant. Existing empirical fragility and vulnerability relationships: compendium and guide for selection. *GEM Foundation, Pavia*, 2013.
11. K. Pitilakis, H.n Crowley, and AM. Kaynia. Syner-g: typology definition and fragility functions for physical elements at seismic risk. *Geotechnical, Geological and Earthquake Engineering*, 27, 2014.
12. M. Pagani, D. Monelli, G. Weatherill, L. Danciu, He. Crowley, V. Silva, P. Henshaw, L. Butler, M. Nastasi, L. Panzeri, et al. Openquake engine: An open hazard (and risk) software for the global earthquake model. *Seismological Research Letters*, 85(3):692–702, 2014.
13. E.M. Rathje, C. Dawson, J.E. Padgett, JP Pinelli, D. Stanzione, A. Adair, P. Arduino, S.J Brandenberg, T. Cockerill, C. Dey, et al. Designsafe: new cyberinfrastructure for natural hazards engineering. *Natural Hazards Review*, 18(3):06017001, 2017.
14. J. Moehle and G.G. Deierlein. A framework methodology for performance-based earthquake engineering. In *13th world conference on earthquake engineering*, volume 679, 2004.
15. K.A. Porter. An overview of peer's performance-based earthquake engineering methodology. In *Proceedings of ninth international conference on applications of statistics and probability in civil engineering*, 2003.
16. M. Zaker Esteghamati, J. Lee, M. Musetich, M.M. Flint, and M. Sharifi Mood. Inventory of seismic structural evaluations, performance functions and taxonomies for buildings (INSSEPT). `http://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2469`, 2019.
17. B.J. Ko, S. Wang, T. He, and D. Conway-Jones. On data summarization for machine learning in multi-organization federations. In *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 63–68. IEEE, 2019.
18. ATC. Structural response modification factors (atc-19), 1995.
19. H. Tahir. *Development of fragility curve database for multi-hazard performance based design*. PhD thesis, Virginia Tech, 2016.
20. JP Stewart, SL Kramer, DY Kwak, MW Greenfield, RE Kayen, K Tokimatsu, JD Bray, CZ Beyzaei, M Cubrinovski, T Sekiguchi, et al. Peer-ngl project: Open source global database and model development for the next-generation of liquefaction assessment procedures. *Soil Dynamics and Earthquake Engineering*, 91:317–328, 2016.
21. BuildingSMART. Ifc overview summary. `http://www.buildingsmart-tech.org/specifications/ifc-overview`.
22. gbxml schema. `hhttps://github.com/GreenBuildingXML/gbXML_Schemas`.

23. P. DuBois. *MySQL cookbook.* " O'Reilly Media, Inc.", 2006.
24. R. J.T. Dyer. *MySQL in a Nutshell: A Desktop Quick Reference.* " O'Reilly Media, Inc.", 2008.
25. Mohsen Zaker Esteghamati and Madeleine M. Flint. Inventory of seismic structural evaluations, performance functions and taxonomies for buildings (INSSEPT) manual. `http://www.doi.org/10.13140/RG.2.2.19855.33444`, 2019.
26. T. Ancheta, Y. Bozorgnia, R. Darragh, WJ. Silva, B. Chiou, JP. Stewart, DM. Boore, R. Graves, NA. Abrahamson, KW. Campbell, et al. Peer nga-west2 database: A database of ground motions recorded in shallow crustal earthquakes in active tectonic regions. In *Proceedings, 15th World Conference on Earthquake Engineering*, 2012.
27. B. Schwartz, P. Zaitsev, and V. Tkachenko. *High performance MySQL: optimization, backups, and replication.* " O'Reilly Media, Inc.", 2012.
28. M. Zaker Esteghamati, M. Banazadeh, and Q. Huang. The effect of design drift limit on the seismic performance of rc dual high-rise buildings. *The Structural Design of Tall and Special Buildings*, 27(8):e1464, 2018.
29. M Flint, Zaker Esteghamati M., Y. Shahtaheri, L. Dhulipala, E. Gil, H. Tahir, MR. Eatherton, J. de la Garza, JL. Irish, R. Leon, CG. Olgun, G. Reichard, A. Rodriguez-Marek, and C. Zobel. A performance-based decision framework for early design of multi-hazard resilient, sustainable buildings. *In press.*
30. J.S. Steelman and J.F. Hajjar. Influence of inelastic seismic response modeling on regional loss estimation. *Engineering Structures*, 31(12):2976–2987, 2009.