

Rapporti tecnici

INGV

**Management, visualization and
comparison of multiple hazards and
risk using free software: the ByMuR tool**

397



Direttore Responsabile

Silvia MATTONI

Editorial Board

Luigi CUCCI - Editor in Chief (INGV-RM1)

Raffaele AZZARO (INGV-CT)

Mario CASTELLANO (INGV-NA)

Viviana CASTELLI (INGV-BO)

Rosa Anna CORSARO (INGV-CT)

Mauro DI VITO (INGV-NA)

Marcello LIOTTA (INGV-PA)

Mario MATTIA (INGV-CT)

Milena MORETTI (INGV-ONT)

Nicola PAGLIUCA (INGV-RM1)

Umberto SCIACCA (INGV-RM2)

Alessandro SETTIMI

Salvatore STRAMONDO (INGV-ONT)

Andrea TERTULLIANI (INGV-RM1)

Aldo WINKLER (INGV-RM2)

Segreteria di Redazione

Francesca Di Stefano - Referente

Rossella Celi

Tel. +39 06 51860068

redazionecen@ingv.it

in collaborazione con:

Barbara Angioni (RM1)

REGISTRAZIONE AL TRIBUNALE DI ROMA N.173 | 2014, 23 LUGLIO

© 2014 INGV Istituto Nazionale di Geofisica e Vulcanologia

Rappresentante legale: Carlo DOGLIONI

Sede: Via di Vigna Murata, 605 | Roma



Rapporti tecnici INGV

MANAGEMENT, VISUALIZATION AND COMPARISON OF MULTIPLE HAZARDS AND RISK USING FREE SOFTWARE: THE BYMUR TOOL

Paolo Perfetti¹, Roberto Tonini², Jacopo Selva¹, Licia Faenza³, Anita Grezio¹, Laura Sandri¹

¹INGV (Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna)

²INGV (Istituto Nazionale di Geofisica e Vulcanologia, Sezione Sismologia e Tettonofisica)

³INGV (Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti)

397

How to cite: Perfetti P. et al., (2018). Management, visualization and comparison of multiple hazards and risk using free software: the ByMuR tool. Rapp. Tec. INGV, 397: 1-34.

Contents

Introduction	7
1. Overview	8
2. Data initialization and management.....	9
2.1 Database connection.....	10
2.2 Database creation	10
2.3 Grid definition and loading.....	10
2.4 Inventory loading.....	11
2.5 Datasets loading.....	11
2.6 Ensemble models	12
3. Visualization.....	13
3.1 Hazard	13
3.2 Inventory	14
3.3 Fragility	15
3.4 Loss	15
3.5 Risk	15
3.5.1 Multiple areas aggregated risk.....	15
3.5.2 Risks comparison.....	17
4. Conclusion.....	17
Acknowledgements.....	18
References	18
Appendices	21
A. Code and structure.....	23
B. Installation	25
C. Database internals	26
D. XML file formats	27
D.1 Hazard XML format.....	28
D.2 Inventory XML format.....	28
D.3 Fragility XML format.....	28
D.4 Loss XML format.....	28
D.5 Risk XML format.....	30
D.6 Dataset files validation	30
Glossary	31

Introduction

Nowadays, **risk** assessment is an essential tool for risk management and risk mitigation policies since it provides an evaluation of the risk associated to a recognized threat that can cause harm or damage to humans, their structures or the environment. Risk results from the combination of three factors (UNDRO, 1982):

$$\text{Risk} = \text{Hazard} \times \text{Exposure} \times \text{Vulnerability} \quad (1)$$

In this very general framework, the **hazard** is a measure of the likelihood of the expected size of a dangerous phenomenon in a given target area and in a given exposure time, the **exposure** measures the assets that are potentially at risk (both people and structures) and the **vulnerability** represents the degree at which the assets may be damaged due to the hazard. Of course, this represents a simplification of the reality, which does not account for any uncertainty. Modern risk assessment is based on Probabilistic Risk Analyses (PRA), making use of probabilistic hazard analysis, fragility and loss models. Such probabilistic factors are typically combined through Pacific Earthquake Engineering Research (PEER) formula [Cornell and Krawinkle, 2000; Der Kiureghian, 2005], which has been originally proposed for seismic risk and it has been recently proposed as a possible base for multi-risk assessments [Marzocchi et al., 2012; Selva, 2013; Mignan et al., 2014; Liu et al., 2015].

In this framework, the project called ByMuR (<http://bymur.bo.ingv.it>), funded by the Italian Ministry of Education, Universities and Research (Ministero dell'Istruzione, dell'Università e della Ricerca – MIUR) and lasted 4 years from November 2010 to November 2014, has been accomplished with the goal of providing a quantitative and objective method for quantifying the multi-risk for a given area. The approach is based on Bayesian inference that allows to account for both aleatory and epistemic uncertainties along all the phases of the computation, from hazard to risk (for hazards: [Selva and Sandri, 2013; Faenza et al., 2017; Sandri et al., 2016; Grezio et al., 2015; Selva et al., 2015]; for vulnerability and risk: [Selva et al., 2013; Selva et al., in prep.]; for the uncertainty treatment: [Marzocchi et al., 2015]). More specifically, the project was focused to: (i) provide a quantitative and objective general method for a comprehensive long-term multi-risk analysis in a given area, accounting for inter-model epistemic uncertainty through Bayesian methodologies, and (ii) apply the methodology to seismic, volcanic and tsunami risks in Naples (Italy).

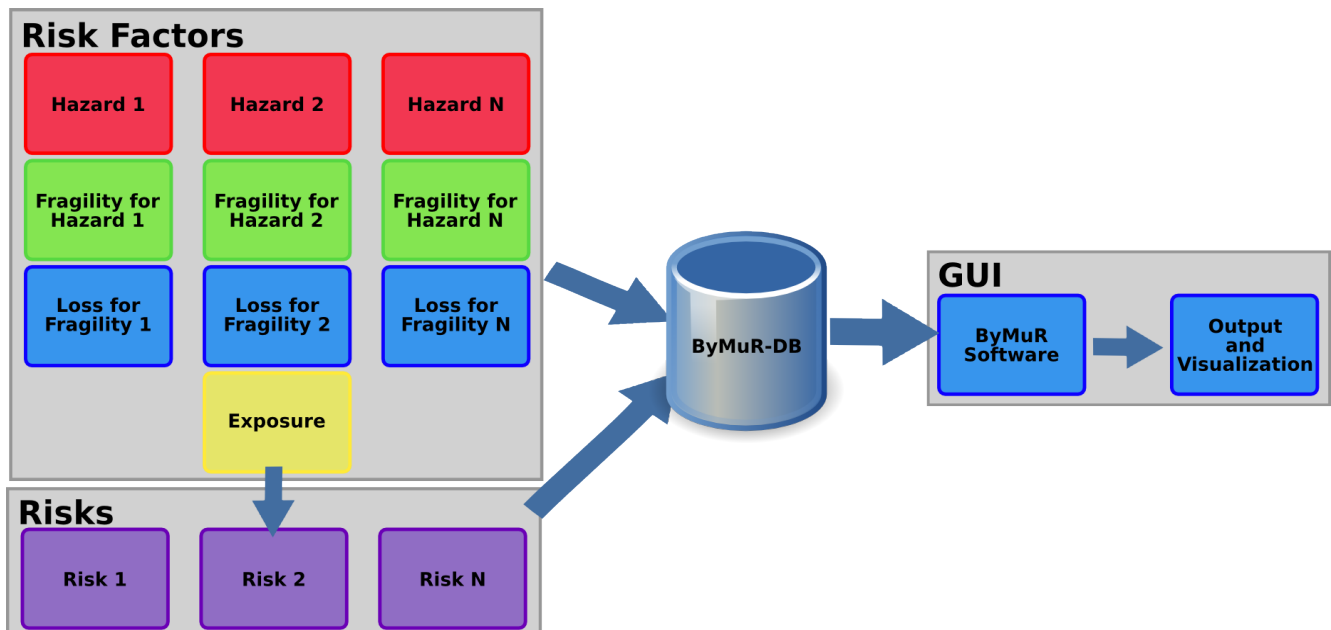


Figure 1. General scheme of all the inputs to ByMuR software for the case study of Naples. The tool handles both single and multi risk factors through a database and users can visualize or combine them to perform risk/multi-risk analysis for a selected target area.

One of the main products/results provided by the project is the homonym ByMuR software [Perfetti and Tonini, 2018], an open source tool aiming to manage, visualize and compare PRA results (e.g. risk curves, risk indexes) for different hazards, as well as, all the components of any PRA (e.g. hazard, fragilities and exposure), within a multi-risk perspective. For each component and for PRA results, epistemic uncertainty is handled through the mean and several percentiles of the community distribution (e.g., SSHAC 1997) or of the ensemble model [Marzocchi and Jordan, 2014]. The ByMuR software handles separately the probabilistic hazard assessments of different kind of hazardous phenomena, the relative fragility and loss models, exposure data, as well as the corresponding probabilistic risk results. The software manages pre-computed results, which are input through standard formats and allows several post-processing of such data in order to produce *ensemble* hazard models and spatially aggregated results.

In this report we present the implementation of the ByMuR software, by describing its features and by illustrating how it could serve the scope of supporting a multi-hazard and multi-risk analysis, as well as, for any hazard assessment. The results computed for the municipality of Naples are demonstrative of the general methodology and, at the same time, are here used to illustrate the features and usability of the ByMuR software. The case study is focused on the long-term (5, 10 and 50 years) multi-risk assessment due to volcanic (tephra fall and pyroclastic flows), seismic and tsunami hazardous phenomena.

1. Overview

Interactive mode of ByMuR (available at <https://github.com/perfett/ByMuR> under the GNU Affero General Public License version 3), is based on a main application window divided in four functional panels:

- 1) the *control panel*, top left, which selects the kind of the phenomenon and the relative hazard models to plot. Through this panel the user can interactively customize the maps and plots showing the hazard assessments by setting important parameters which are: (i) the **exposure time** (i.e., the time window considered as reference for all the assessments and for all the elements that are exposed to the hazard in the risk evaluation), (ii) the **mean return period threshold** (i.e., the reciprocal of mean recurrence rate, often indicated with λ) and (iii) the **intensity measure threshold** (a value of hazard's intensity in its units, e.g., in g units, for the Peak Ground Acceleration in seismic hazard). The *Easting* and *Northing* combo box also selects the UTM coordinates of a specific point of the required map;
- 2) the *data panel*, bottom left, which is a read-only panel that shows the inventory and geographic informations of the selected areas, when available;
- 3) the *curve panel*, bottom right, which groups all tabs containing curves and assets plots. It allows switching among hazard, fragility, inventory, loss and risk graphs, zooming and panning the selected results and comparing several risks when they are available (see below chapter 3 for more details);
- 4) the *map panel*, top right, which contains the hazard and the probability map corresponding to the mean return period and the intensity measure thresholds selected in 1), respectively. More specifically, the hazard map shows the hazard intensity corresponding to the selected mean return period (which, in turn, corresponds to a specific probability of exceedance in the selected exposure time). The probability map reports the exceedance probability corresponding to the selected intensity measure threshold in the selected exposure time. Clicking on both maps it is possible to select a specific target point or, alternatively, highlight an area to which refer the plots in 3).

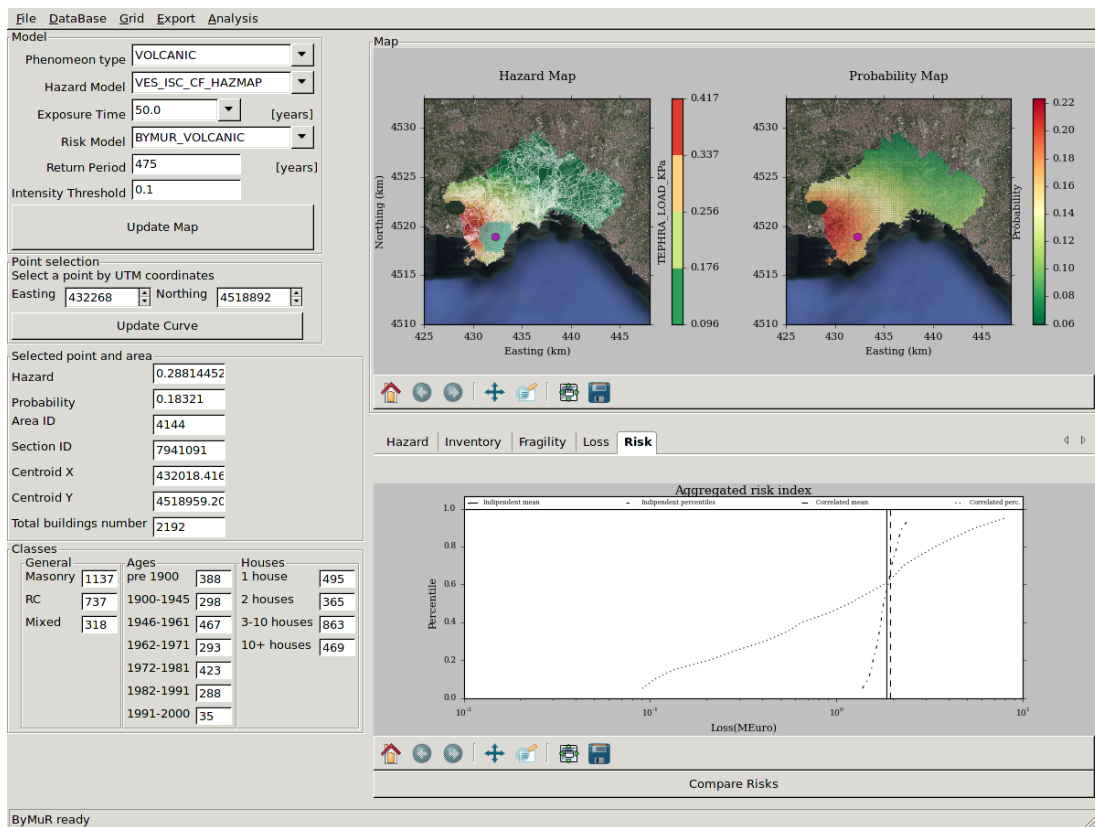


Figure 2. ByMuR software's main window.

In order to use the ByMuR software all the results have to be prepared following specific formats (see Appendix D) and the software will then automatically store and manage them using a specific relational database (ByMuR-DB, see Appendix C). When a database is connected, the left column and the control panel are activated showing the hazard models loaded in the database. The different input fields in both control and data panels should then be filled in subsequently from top to down, since the drop-down menus contain the corresponding entries, for each parameter, which depend on the ones previously selected: for instance, after choosing a particular *Phenomenon type*, the list of items in *Hazard model* and *Exposure time* will be populated accordingly. The remaining text boxes are filled in with default values and they are free float fields.

When the *Update map* button is pressed, the dataset corresponding to the hazard model selection is retrieved from the database and visualized in the map and the curve panels, according to the values set in the control panel. Clicking on a specific point of the map the corresponding curves will be plotted in the curves panel.

2. Data initialization and management

ByMuR plotting operations rely on specific queries to a relational database management system (RDBMS) containing all relevant data, hazard models, inventory, fragility models, loss models, risk results, and grid points. On the other hand, several file formats are used to import and export the datasets from and to the database. It is not possible to use multiple databases at the same time, so all the data relevant to a specific analysis should be preloaded in a unique database.

To exchange data between researchers and groups, XML files are preferred over other database dumps or copies, as the XML markup language ensures abstraction from every applicative and configuration details due to the used database.

Complexities derived from the utilization of an RDBMS, as the knowledge of SQL syntax to execute database queries, can be ignored. This is possible thanks to the procedures that the ByMuR GUI makes available in the Database menu, like the creation or the deletion of data repositories and the loading of a new datasets. Even if the main feature of ByMuR is the interactive visualization of data and results, some

additional functions of data processing have been implemented to improve the analysis and support the interpretation of the results. This chapter will focus on the data management procedures, while unfolding the rationale behind. More detailed explanation of the database structure and its internals are described in Appendix C and the file exchange formats full specifications are discussed in Appendix D.

2.1 Database connection

Database connection is available in the menu *File* and the dialogue will prompt for the connection details. It is possible to keep several databases to reflect different stages of the development process (e.g. production, testing, ...) or different research groups datasets and then select which one ByMuR should connect at every execution of the program.

Once the “Ok” button is pressed, if the connection is successfully established, then the available hazard models and the associated exposure times are loaded in the corresponding drop-down menus. In case of connection error, an exception will be raised.

Visualization operations do not require write permission on the database and thus it is possible to configure different privilege levels to access stored data, creating a read only account if needed.

2.2 Database creation

In the menu *Database*, the entry *Create Database (DB)* is available to proceed to the creation of a new database and a dialogue will be prompted asking for the database details (see Figure 3).

MySQL server installation and configuration will not be covered in this report but it is worth noticing that, to avoid errors, the database user must exist and the needed permissions must be granted before attempting database creation through the ByMuR menu.

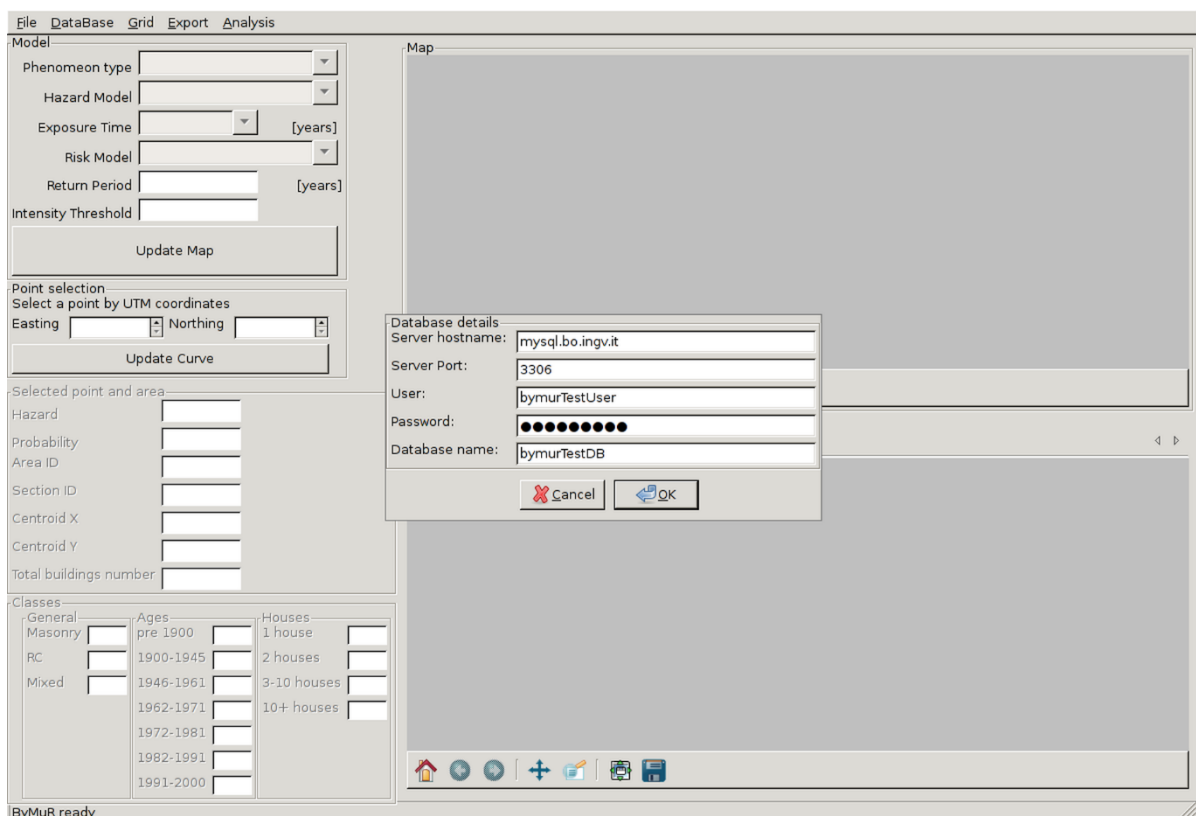


Figure 3. ByMuR Create Database dialogue.

2.3 Grid definition and loading

The first information to load after the database creation is a common grid map over which all hazard results are computed. The memorization of any hazard dataset depends on the previous existence of the reference grid points. Grid loading can be done via the *Grid - Load grid file* menu entry or using the mandatory flag ‘-g’ when in batch mode and later the grid will be referenced using its uppercased filename.

The file format used to exchange grids is very simple: a text file (with tab-separated columns and the LF, line feed, special character as line ending) containing easting, northing and (possibly) elevation with respect the sea level expressed in WGS84 UTM coordinates. Future development will enhance grid specification flexibility adding the possibility to include WGS84 zone specification and to use different spatial reference systems.

Grid points are decomposed and singularly stored in the database, to avoid useless replication and to easily collect all data relative to a single location.

2.4 Inventory loading

Before enabling any of the functionality derived from the fragility, loss and risk results, it is needed to load the detailed territory inventory adopted for the corresponding risk analysis. This inventory, along with the spatial grid for hazards, defines the common ground for the different components of each risk (hazard, fragility and loss), as well as, for the different risks. The data model for the inventory includes the definition of the areas, each identified by a section ID and containing specific information about its building inventory, that is, buildings number, age, and size. Since the building classes to be used for the inventory depend on the risk type, it is required also that the taxonomy is consistent between inventory, fragility and loss model.

As for the grid, inventory areas definitions had to be present in the database before other data referencing them can be loaded. By now only one inventory is supported in database and it can be loaded via the *Database -Add data to DB* (see Figure 4).

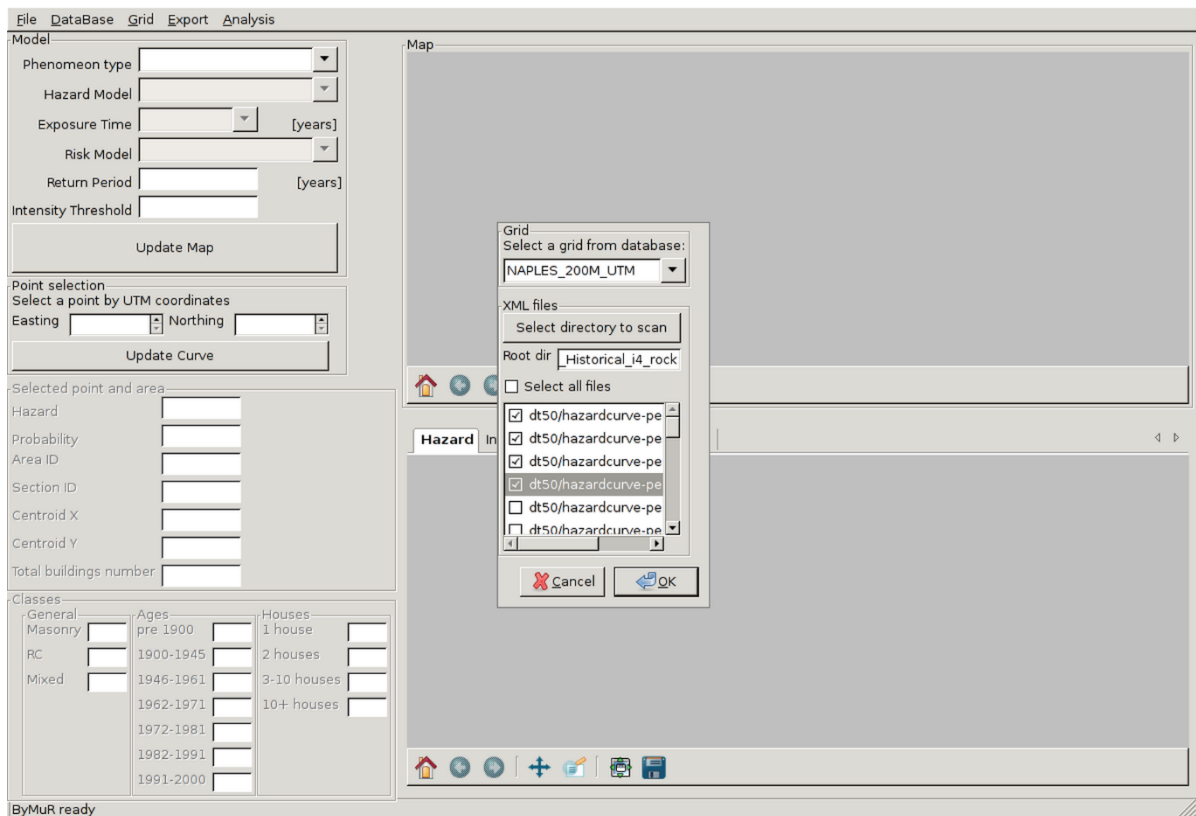


Figure 4. Add Data window: all XML file in selected directory will be listed and different inventory and hazards files can be loaded at once.

2.5 Datasets loading

A newly created database contains just the basic structure and the list of supported hazard phenomena families as defined in the inventory (*Seismic, Volcanic* and *Tsunami* in our application). As seen in the previous sections, reference grid and inventory have to be loaded before any dataset would be imported, otherwise the import procedure will raise an exception and fail.

Hazard, fragility, loss and risk data can be loaded through menu *Database - Add Data to DB*.

To proceed loading datasets, an available grid has to be selected in the corresponding text box. It is

worth noticing that the hazard datasets must be computed precisely on the grid points, not only “in the region” or within the shape of the grid. This is relevant because, to render the maps, only grid points data is searched in the database; so if a curve has no point associated then the curve will be never shown. Note that the data model for the hazard has been adopted also for the volcanic hazard assessments produced for the project INGV-DPC-V1 (2012-2013).

After selecting the grid name, the root directory of the dataset repository must be indicated using the dedicated dialog (Figure 4): this directory will be recursively scanned searching for XML files which will be added to checklist box. Listed path names are relative to selected root directory and no check is done at this stage on internal file format. When the ‘Ok’ button is pressed, the import procedure individually scans all files to distinguish between the supported XML formats and to correctly import data in the database. Thus it is possible to import file of different kinds of data in a single import procedure execution. New imported data referring to the same hazard/fragility/loss/risk model name and grid point will update the previously loaded ones, without any duplication.

2.6 Ensemble models

The ByMuR program offers the possibility to produce hazard models that combines two or more of the available homogeneous hazard models. Such newly created ensemble hazard models are then stored in the database and can be later visualized in the maps.

Two or more hazard models, referring to the same hazardous event, can be combined through weighted statistical mixing, that is, calculated using a weighted sampling of the probability distributions of the selected hazard models (e.g., Ray and Lindsay, [2005]). The obtained model entails a new model that can be considered an ensemble of the starting models [see Marzocchi and Jordan, 2014; Marzocchi et al., 2015; Selva et al., 2016]. This approach has been recently used for probabilistic volcanic hazard assessments [Selva et al., 2014; Tonini et al., 2015; DPC-INGV-V1, 2012-2013] and it is used here as general method to combine hazard models of the same origin. Selecting the item *Create Ensemble* in the drop-down menu *Analysis*, a new frame is opened and it is possible to choose which hazard models, among the loaded ones, will be used to calculate the ensemble model.

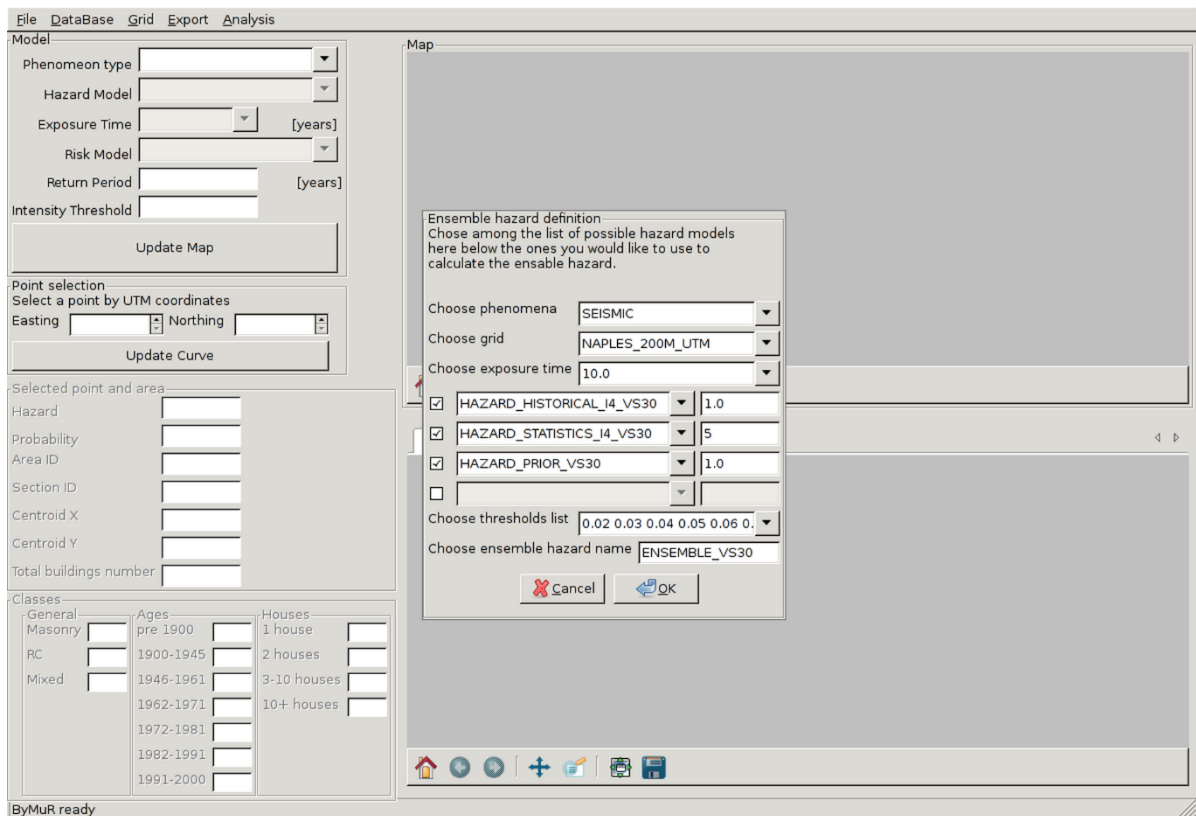


Figure 5. Ensemble function: weights can be assigned with any value, since the software automatically normalize them to the total sum to 1.

A weight can be assigned to each model (i.e., the higher weight to the most reliable model) and a time window shared between the selected models has to be selected (see Figure 5). As soon as it is created, the new ensemble model is automatically stored in the previously connected ByMuR database and made available to be loaded and visualized as any other hazard model.

3. Visualization

After connecting to the database and to begin visualize datasets in ByMuR, necessary parameters has to be selected:

- 1) Phenomenon.
- 2) Hazard Model.
- 3) Exposure Time.
- 4) Risk model (only if it is available for given hazard model and exposure time).
- 5) Return period.
- 6) Intensity thresholds.

When *Update Map* button is pressed, the dataset is retrieved from ByMuR-DB and the hazard and probability maps are visualized in the upper right panel of ByMuR's main window. User can zoom, pan and examine the maps using the navigation bar located under the plot axis (Figure 6). More details on hazard and probability maps are given in the next subsection.

In the bottom right panel can be visualized five kinds of plot: hazard curves, inventory classes, fragility curves, loss curves and risk indexes and curves (see Figure 7).

Selection on the map can be done in two ways:

- single point/area selection by simply clicking on (or near) a sample point on maps; ByMuR will plot all curves defined for that point contained in ByMuR-DB and belonging to the selected hazard model. If a risk model is defined, also fragility, loss, risk and inventory data for the selected area will be plotted;
- multiple areas selection by clicking and dragging the mouse on a map. Resulting selection will contains all the areas intersecting rectangular selection done by the user. In this case only risk index plot will be shown, aggregating data available in ByMuR-DB for the highlighted areas, and all other tabs will be disabled.

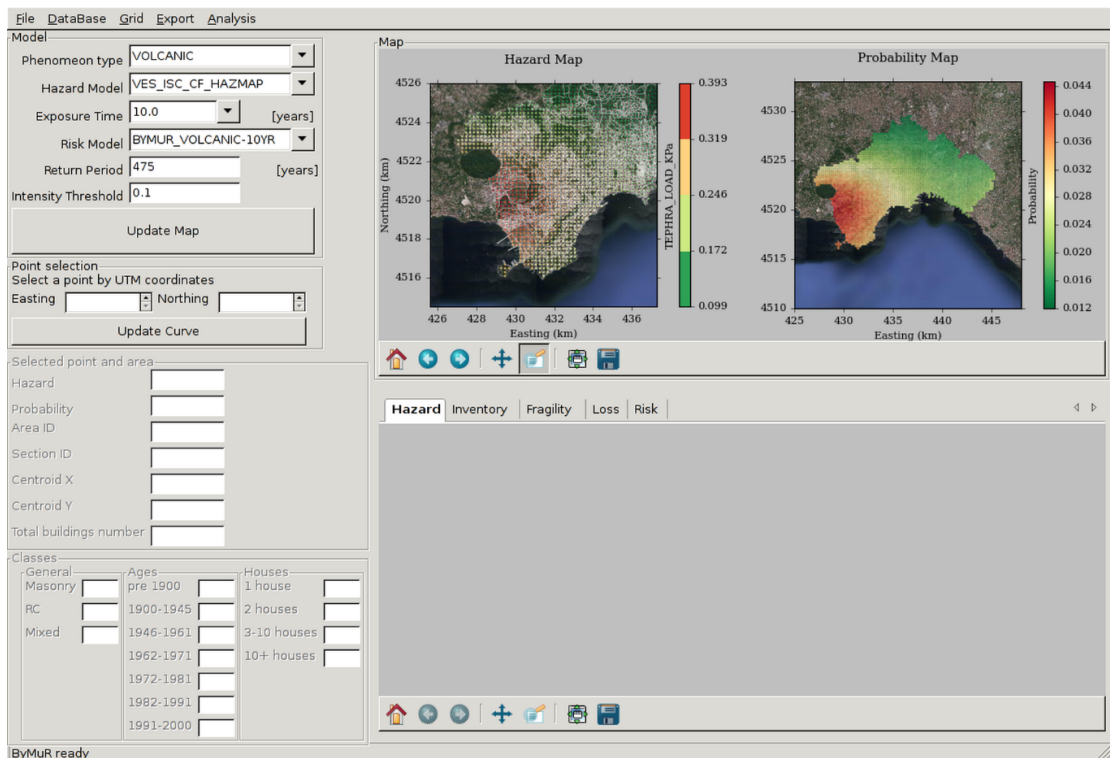
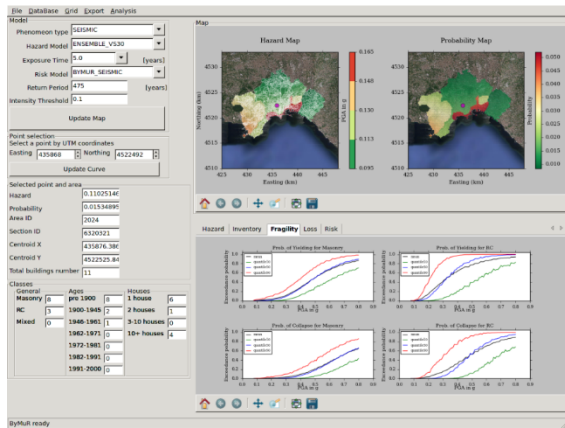
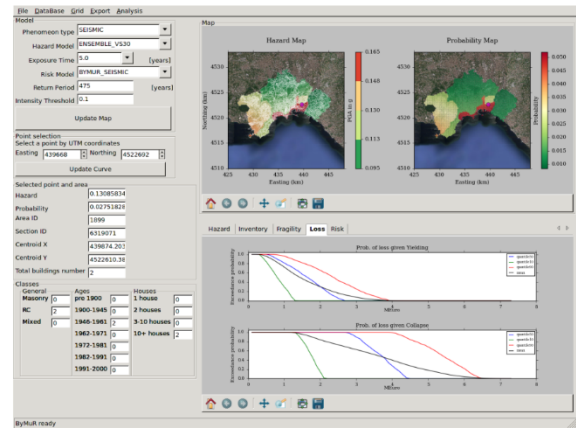


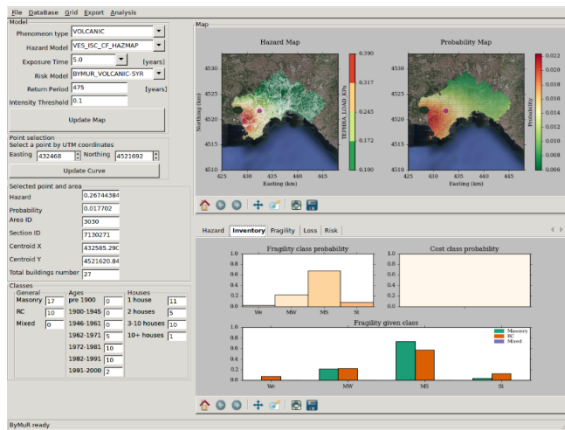
Figure 6. Hazard and probability map visualization: user can zoom, pan and examine the maps using the navigation bar located under the plot axis.



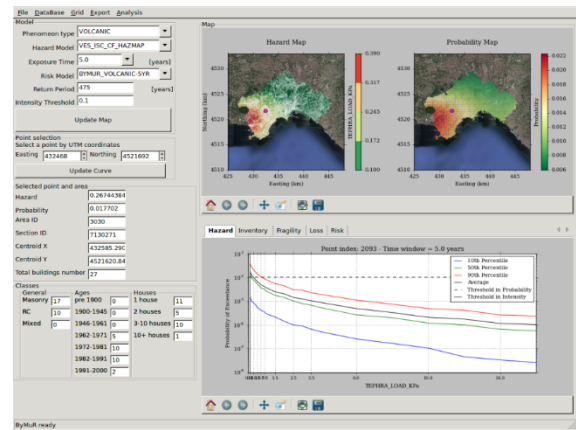
a)



b)



c)



d)

Figure 7. a) Fragility curves, b) Loss curves, c) Inventory classes, d) Hazard curves.

3.1 Hazard

In the Curve Panel, the first tab depicts the *Hazard Curve*: this curve represents the probability of exceedance in the exposure time of different values of the hazard intensity measure. The units and range of values depend on the hazardous phenomenon, as stored in the database. Percentiles (10th, 50th and 90th), describing the epistemic uncertainty related to the hazard quantification, are plotted along with the mean probability, providing a quantitative estimation of the uncertainty associated to the probabilistic assessment.

Hazard and probability maps shown in the upper panel are obtained by cutting, for each point in the computational grid, the average hazard curves by a horizontal (a fixed threshold in probability, expressed through the mean return time in the *Return Period* textbox of the control panel) or a vertical (a fixed threshold in the hazard intensity, set in the *Intensity Threshold* textbox) line, respectively. A very similar approach to visualize hazard assessment and its uncertainty was adopted for volcanic hazard in the PyBetVH software [Tonini et. al., 2015].

In the left column, in the Data Panel, reference point numerical values are shown including UTM coordinates, exceedance probability, area inventory ID, section ID and area centroid.

3.2 Inventory

Inventory data is central in the procedure of determining the risk of a specific area. In the Data Panel numerical values are listed and it is possible to know with great details how many buildings are in the selected area and how they are divided between the General/Age/Size classifications. In the main curve panel, it is plotted the absolute probability that a specific building of a given selected area (which can be selected by clicking on the map) belong to each of the fragility and cost classes defined for the adopted taxonomy. Note that these classes are also the ones adopted by the fragility and loss models. In order to link the fragility classes to a more standardized structural categories, it is also plotted the probability that a building belonging to one specific fragility category will own to one general structural class: RC, masonry and mixed).

3.3 Fragility

Fragility models are defined only for each hazard intensity and they are not necessarily linked to a specific hazard model. Fragility is expressed as the probability to exceed a defined *limit state* given a certain phenomena intensity (e.g. exceedance probability of roof collapse given a given load of tephra; e.g., NIBS [2004]). Hence the same fragility is valid for all hazard models with the same origin. Here, fragility models are defined per area and general building class, and they include the uncertainty on the inventory. The number of plots in the tab is based on the available building classes for the selected area and the limit states defined for the considered fragility model. In particular, the plots are organized in a table, whose column depends on the building class and the rows contains all the considered limit states. Note that if a building class is not present in the selected area, the corresponding column is omitted. In each plot, the epistemic uncertainty degree is shown using multiple line with different mark styles.

3.4 Loss

Loss models assess the probability of exceeding different levels of (monetary, in this case) losses given the occurrence of a limit state for one random structure within the selected area (e.g. Stergiou and Kiremidjian [2006]). Loss curves depend on the defined cost classes, hence the same loss model can in theory apply to different fragility and/or hazard models.

The number of plots in the tab is based on the limit states defined for the considered fragility model. In particular, the plots are organized with one row for each present limit state. In each plot, one per level, and also in the loss graph, the epistemic uncertainty degree is shown using multiple lines with different mark styles.

Here, loss models are defined per area and limit state, and they include the uncertainty on the inventory. In each plot, the epistemic uncertainty degree is shown using multiple lines with different mark styles.

3.5 Risk

When risk is available for certain hazard model, corresponding textbox and risk tab are enabled. Data plotted in the risk graph are taken from the model previously inserted in the database: no live risk calculation is done (apart for the risk aggregation, as explained later) even if values depend on all the other data and models that have been loaded in order to plot a specific risk.

Here, risk results are calculated per area and are represented (Figure 8) in terms of Risk Curves (reporting the probability of exceedance at different loss levels) and Risk Indexes (in this case, the Annualized Average Loss), along with the epistemic uncertainty on their estimation [see Selva et al., 2013].

3.5.1 Multiple areas aggregated risk

When a user proceeds to multiple area selection (with click-and-drag on hazard map) the aggregated risk is calculated. Given the important spatial correlation of risk results (at least, correlated by the hazard), Risk Curves cannot be aggregated. On the opposite, Risk Indexes can be meaningfully aggregated, since they represent averages.

As it regards the propagation of epistemic uncertainty on the aggregated Risk Index, two opposite assumptions are considered on epistemic uncertainty correlation between areas:

- epistemic uncertainty on risk index values is assumed independent among areas and every curve (epistemic uncertainty distribution) is independently sampled;
- epistemic uncertainty on risk index values is assumed completely correlated among areas and thus all curves (epistemic uncertainty distribution) are sampled jointly.

For multiple areas selection only the risk index plot (Figure 9) is meaningful and the other plots (hazard, inventory, fragility, loss and risk curves) are disabled. On Data panel, values in hazard, probability, area, section ID and centroid coordinates are relative to the selected point and the single area where it is contained, while inventory data (buildings total and classes) are calculated considering the whole selection.

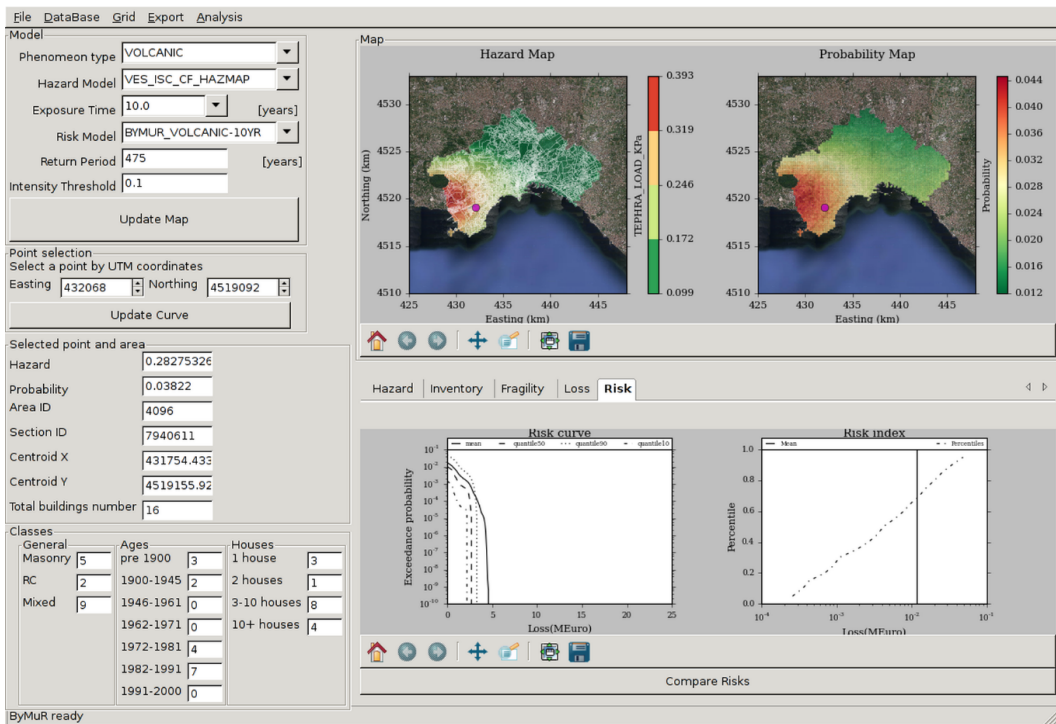


Figure 8. Aggregated risk visualization.

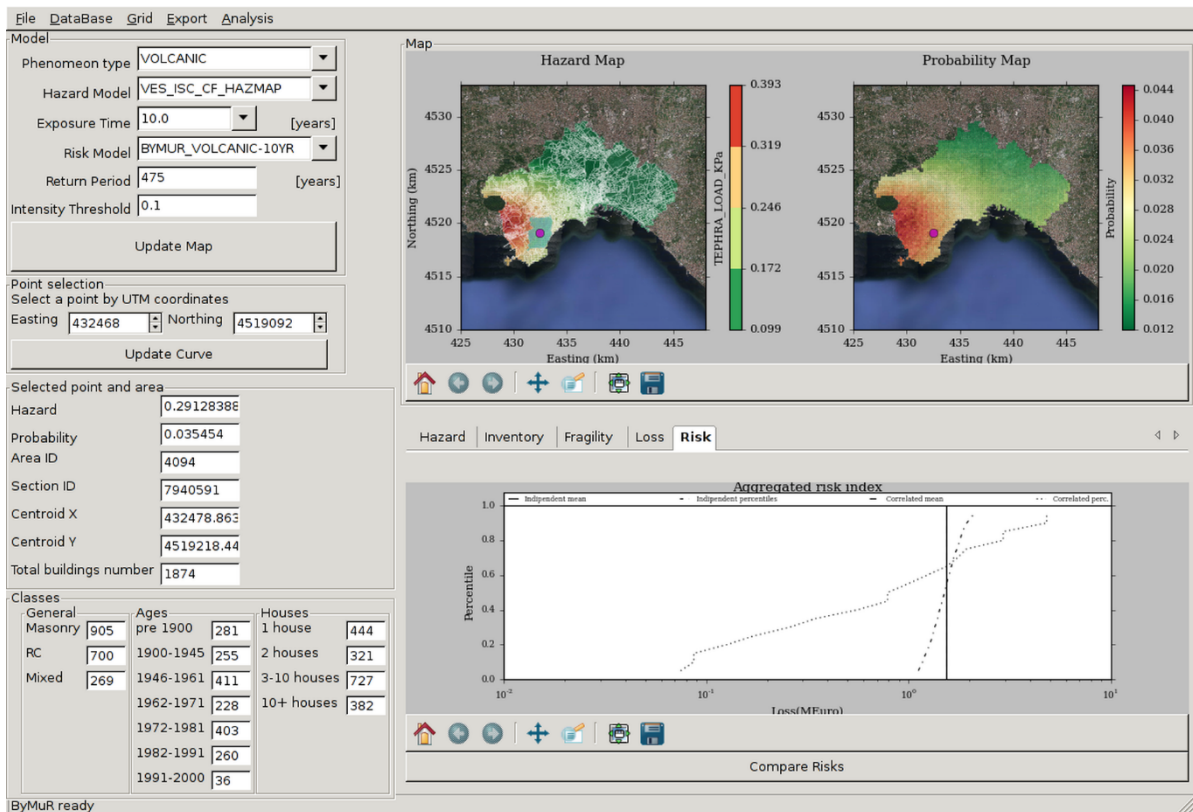


Figure 9. Risk curves.

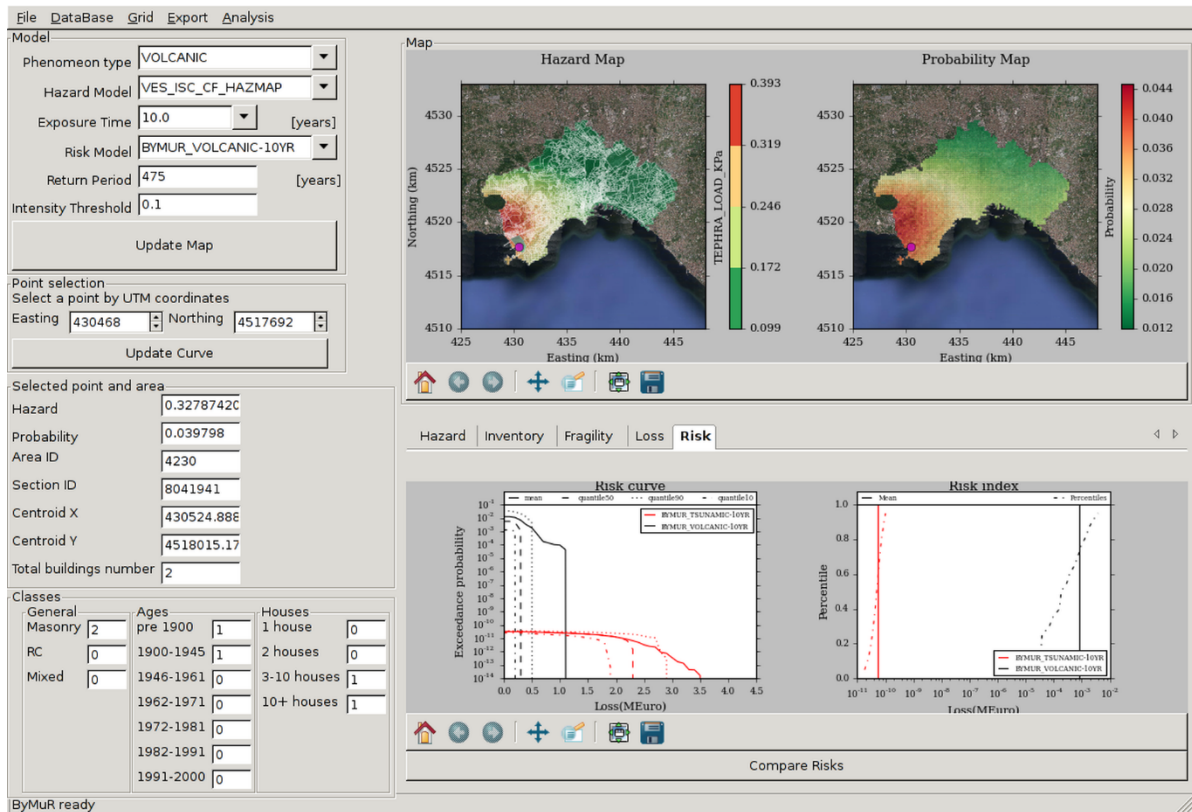


Figure 10. Risks comparison.

3.5.2 Risks comparison

When a risk model is available for the selected hazard model then it is possible to compare it with the risk models defined for other hazardous phenomena. To select which risk model to compare with, users have to open the selection dialog through the *Compare Risks* button located on the bottom of Risk's tab. The resulting risk's curves and indexes will be plotted on the same graph using different colors to distinguish between phenomena. To ease comparison and to avoid overcrowded, and thus unclear graphs, a logarithmic x-axis is used and only model's mean and percentiles are plotted.

4. Conclusion

We have presented ByMuR, a Free/Libre/Open Source Software (FLOSS) aimed to visualize and evaluate hazard datasets and relative fragility, loss, exposure data and probabilistic risk assessments. The project consists in two, decoupled parts: ByMuR-DB, the SQL-based database to store and retrieve pre-calculated data and the ByMuR application, a GUI that provides a user-friendly interface to manage and visualize those datasets. The graphical user interface enables users to specify which data source to connect or create, import and export datasets, define new ensemble models, visualize curves for single sampling points or multiple areas, compare hazards impact on assets.

ByMuR-DB database is independent from the visualization interface and thus it is exportable and it can be used as backend for other compliant applications. Input data formats are based on popular XML specifications, and they have been strictly formalized to permit and encourage software reutilization and datasets exchange between researchers and groups.

Both ByMuR and ByMuR-DB are designed to be easily extendable, in order to manage hazards, data types and regions not originally contemplated during the software design and writing. Standard tools, open formats, free licenses and features modularity are the bases to empower researchers (both final users and developers) to take advantage of this tool.

Acknowledgements

The work has been supported by “ByMuR: Bayesian Multi-Risk Assessment: a case study for natural risks in the city of Naples” (<http://bymur.bo.ingv.it>, Ref. RBFR0880SR), a project funded by the Italian Ministry of Education, Universities and Research (MIUR, Ministero Istruzione dell’Università e della Ricerca) and “V1: Probabilistic Volcanic Hazard Assessments”, a project funded by the Italian Civil Protection Department (DPC, Dipartimento della Protezione Civile).

References

- Cornell C., Krawinkle H., (2000). *Progress and challenges in seismic performance assessment*. PEER Center News 3(2), <http://peer.berkeley.edu/news/2000spring/index.htm>.
- Der Kiureghian A., (2005). *Non-ergodicity and PEER’s framework formula*. Earthq. Eng. Struct. Dyn. 34:1643–1652; doi:10.1002/eqe.504.
- Faenza L., Pierdominici S., Hainzl S., Cinti F.R., Sandri L., Selva J., Tonini R. and Perfetti P., (2017). *A Bayesian seismic hazard analysis for the city of Naples*. J. Geophys. Res. Solid Earth, 122, 1990-2012; doi:10.1002/2016JB013507.
- Gamma E., Helm R., Johnson R., and Vlissides J., (1995). *Design Patterns: Elements of Reusable Object-oriented Software*. Addison-Wesley Longman Publishing Co., Inc.
- Grezio A., Tonini R., Sandri L., Pierdominici S., Selva J., (2015). *A Methodology for a Comprehensive Probabilistic Tsunami Hazard Assessment: Multiple Sources and Short-Term Interactions*, Journal of Marine Science and Engineering 3 (1), 23-51.
- INGV-DPC project V1, (2012-2013). *Valutazione della pericolosità vulcanica in termini probabilistici*. funded by Dipartimento Protezione Civile.
- ISTAT (2001). *14° Censimento popolazione e abitazioni 2001*. Tech. rep., <http://www.istat.it/it/censimento-popolazione/popolazione-2001>.
- Liu Z., Nadim F., Garcia-Aristizabal A., Mignan A., Fleming K., Luna B.Q., (2015). *A three-level framework for multi-risk assessment*. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 9:2, 59-74; doi: 10.1080/17499518.2015.1041989.
- Marzocchi W., Sandri L., Gasparini P., Newhall C.G. & Boschi E., (2004). *Quantifying probabilities of volcanic events: The example of volcanic hazard at Mount Vesuvius*. J. Geophys Res 109, 1-18.
- Marzocchi W., Sandri L. & Selva J., (2008). *BET_EF: a probabilistic tool for long- and short-term eruption forecasting*. Bull. Volcanol. 70, 623-632.
- Marzocchi W., Sandri L. & Selva J., (2010). *BET_VH: a probabilistic tool for long-term volcanic hazard assessment*. Bull. Volcanol. 72, 705-716.
- Marzocchi W., Garcia-Aristizabal A., Gasparini P., Mastellone M., Di Ruocco A., (2012). *Basic principles of multi-risk assessment: a case study in Italy*. Natural Hazards, 62, 551-573.
- Marzocchi W., Taroni M., Selva J., (2015). *Accounting for Epistemic Uncertainty in PSHA: Logic Tree and Ensemble Modeling*. Bulletin of the Seismological Society of America, 105 (4); doi: 10.1785/0120140131.
- Mignan A., Wiemer S., Giardini D., (2014). *The quantification of low-probability–high-consequences events: part I. A generic multi-risk approach* Natural Hazards. Springer Netherlands, 73, 1999-2022.
- National Institute of Building Sciences (NIBS), (2004). *HAZUS-MH: user’s manual and technical manuals*. Report prepared for the Federal Emergency Management Agency, Washington, DC.
- Perfetti P., Tonini R., (2018). *perfettp/ByMuR: Technical report release - v0.5.0*. Zenodo; doi:10.5281/zenodo.1188816.
- Ray S., Lindsay B.G., (2005). *The topography of multivariate normal mixtures*. Ann. Stat., 33(5), 2042-2065.
- Ratdomopurbo A., Widiwijayanti C., Newhall C., Win N.T.Z., (2013). *WOVOdat Database: Progress Report 2013*. IAVCEI 2013 Scientific Assembly - July 20- 24, Kagoshima, Japan.
- Sandri L., Costa A., Selva J., Tonini R., Macedonio G., Folch A., Sulpizio R., (2016). *Beyond eruptive scenarios: assessing tephra fallout hazard from Neapolitan volcanoes*. Sci. Rep., 6, 24271; doi: 10.1038/srep24271.
- Selva J., (2013). *Long-term multi-risk assessment: statistical treatment of interaction among risks*. Natural Hazards, 67, 701-722.

- Selva J., Doumaz F., Reitano D., Troiano A., Vinci S., and the DIVO Consortium, (2009). *DIVO Database for Italian Volcanoes*. AGU Fall Meeting, 14-18/12/2009 San Francisco (USA) WOVodat database, www.wovodat.org.
- Selva J. and Sandri L., (2013). *Probabilistic Seismic Hazard Assessment: Combining Cornell-like approaches and data at sites through Bayesian inference*. Bull. Seism. Soc. Am. 103(3):1709-1722; doi:10.1785/0120120091.
- Selva J., Argyroudou S., Pitalakis K., (2013). *Impact on loss/risk assessments of inter-model variability in vulnerability analysis*, Natural Hazards, 67 (2), 723-746.
- Selva J., Costa A., Sandri L., Macedonio G., and Marzocchi W., (2014). *Probabilistic short-term volcanic hazard in phases of unrest: A case study for tephra fallout*. J. Geophys. Res. Solid Earth, 119; doi:10.1002/2014JB011252.
- Selva J., Tonini R., Molinari I., Tiberti M.M., Romano F., Grezio A., Melini D., Piatanesi A., Basili R., Lorito S., (2016). *Quantification of source uncertainties in Seismic Probabilistic Tsunami Hazard Analysis (SPTHA)*. Geophys. J. Int., 205: 1780-1803; doi: 10.1093/gji/ggw107.
- SSHAC (Senior Seismic Hazard Analysis Committee), (1997). *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*. U.S. Nuclear Regulatory Commission, U.S. Dept. of Energy, Electric Power Research Institute, NUREG/CR-6372, UCRL-ID-122160, Vols. 1/2.
- Stergiou E., Kiremidjian A.S., (2006). *Treatment of uncertainties in seismic risk analysis of transportation systems*. Technical report no. 154, John A. Blume Earthquake Engineering Center, Civil Engineering Department, Stanford University, Stanford, CA.
- Tonini R., Sandri L., Costa A., Selva J., (2015). *The effect of submerged vents on probabilistic hazard assessment for tephra fallout*. Nat. Hazards Earth Syst. Sci. Discuss., 2, 7181-7196; doi:10.5194/nhessd-2-7181-2014.
- Tonini R., Sandri L., and Thompson M.A., (2015). *PyBetVH: a Python tool for probabilistic volcanic hazard assessment and for generation of Bayesian hazard curves and maps*. Comput. Geosci., 79:38-46; doi: 10.1016/j.cageo.2015.02.017.
- UNDRO United Nations Disaster Relief Organization, (1982). *Natural disasters and vulnerability analysis*. Geneva: Office of the United Nations Disaster Relief Coordinator.

Appendices

A. Code and structure

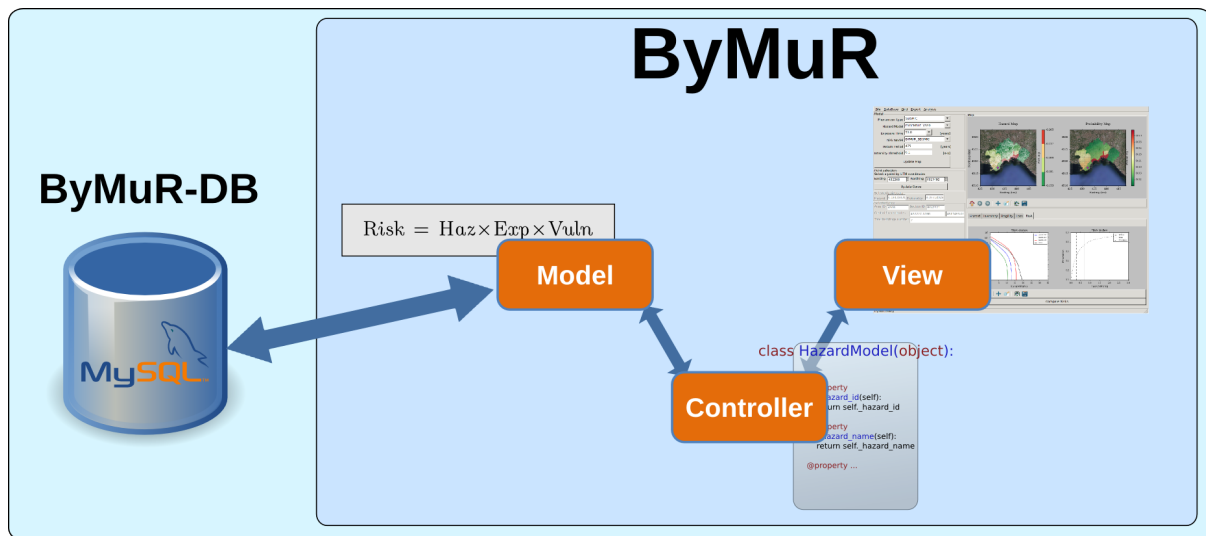


Figure 11. ByMuR software architecture.

In order to facilitate the development of a complete graphical tool, without sacrificing any requested feature and maintaining a reasonable complexity, Python has been chosen as main programming language. Python (<http://www.python.org>) is a general-purpose, high-level, dynamically-typed programming language, very popular in the scientific community mainly because of the high level of integration, extensibility and scalability (in terms of “line of code”).

Python’s facilities for C/Fortran along with the availability of its interpreter (CPython) as free and open source software on all most common operating systems (GNU/Linux, Windows, Mac-OSX) have determined the spread of Python’s usage within research community. One of the main advantages in the utilization of Python consists in the large amount of third party modules available.

The modules included in ByMuR are:

- wxPython (<http://www.wxpython.org>) a cross-platform GUI toolkit that allows to create applications that can run unmodified on different platform like MS Widows, OS X, GNU/Linux and most Unix-like systems;
- MySQLdb (<http://mysql-python.sourceforge.net>) a Python DB API-2.0-compliant interface to MySQL databases;
- Numpy (<http://www.numpy.org>) a scientific library for high level mathematical operations with large arrays, linear algebra and random numbers;
- Matplotlib (<http://www.matplotlib.org>) a Python 2D plotting library which can be integrated in wxPython toolkit to build accurate and interactive plots;
- lxml (<http://lxml.de>) a XML toolkit providing a Pythonic binding for the C libraries libxml2 and libxslt. Its interface is compatible with the ElementTree API but it is faster, especially in serialization.

ByMuR software has been developed by following the MVC (model, view, controller) architecture and, even if not rigorously, the application code has been split in three main files implementing the three different roles:

- `bymur.py` contains the main loop needed to provide a GUI using wxPython library and all GUI widget declarations. Here it is defined the look and feel of the application but the application logic is taken from controller module;
- `bymur_core.py` contains the implementation of the model and it is the only module that read/write data from/to the database. “HazardModel” is the main Python Class which implements a user defined hazard model. Area and point selections, threshold computation and application status reside here;

- `bymur_controller.py` realizes the coordination between model and controller. It responds to the user input and invokes corresponding methods to update model status. This module also keeps the GUI synchronized with the model, propagating events and data.
- The following additional files collect classes and functions for some specific functionalities.
- `bymur_db.py` keeps the definition of ByMuR-DB class and all its methods. All ByMuR-DB related functions are implemented here to simplify maintenance and storage update/modification.
- `bymur_plot.py` encapsulates matplotlib graph and functions in specific classes. No data computation logic is implemented here, just queries and visualization logic. This module plots the datasets while all callback functions and data are taken as input parameters.
- `bymur_functions.py`: contains all models and wx events classes, ancillary data, functions, and threads management procedures which are included in all other files.

A multi-threading approach is necessary in order to keep the application responsive even during CPU-intensive computations or long data transfers from database. Taking advantage of the event-loop construct implemented by the wxPython library, in ByMuR several custom events are declared to manage user actions and data updates. When a long process is about to be executed, a new thread is spawned to run the computationally expensive routine while the main thread creates a waiting dialogue and returns the execution control to the wx event loop, adding a listener to wait for the routine end event.

This solution permits to keep ByMuR responsive under every circumstance and to refresh the GUI even if a heavy computation is running. Since the number of records managed in a single model visualization is very large, the data is read from ByMuR-DB on demand, avoiding, when possible, to keep big structures in the program memory. For instance, examining a hazard model, only the mean statistic of hazard curve is loaded for all points to draw hazard and probability maps. When the remaining statistics of a specific coordinates are necessary to plot hazard curve, they are fetched from ByMuR-DB just in the case the user selects that point as the one to examine the curve for.

B. Installation

ByMuR has been written, tested on and targeted for GNU/Linux operating system and Debian/Ubuntu distributions have been used as main development platform. Python libraries are available in almost all Linux packaging systems, but not always their versions are updated nor the user has administration right to install system-wide dependencies. The tool *virtualenv*¹, often used through his wrapper *virtualenvwrapper*², provide a viable solution to the previous problems and they both has been used during development.

1. Install requirements

```
sudo apt-get install python-pip python-virtualenv virtualenvwrapper \  
python-wxgtk2.8 build-essentials libpng-dev \  
libmysqlclient-dev python-dev libfreetype6-dev libxml2-dev \  
libxslt-dev libatlas-base-dev gfortran
```

Tested on Debian 6.0.9, Ubuntu 14.04., packages name could change slightly using other GNU/Linux distributions.

2. Configure virtualenvwrapper

If you want to use a non-standard directory to store your virtualenvs you should comment out (or delete) `/etc/bash_completion.d/virtualenvwrapper` and then add these instructions in `~/.bashrc` or `~/.profile` :

```
export WORKON_HOME=$HOME/dev/virtualenvs  
export VIRTUALENVWRAPPER_LOG_DIR="$WORKON_HOME"  
export VIRTUALENVWRAPPER_HOOK_DIR="$WORKON_HOME"  
source /etc/bash_completion.d/virtualenvwrapper
```

Check on *virtualenvwrapper* documentation for more informations.

3. Create the new virtualenv

```
$ mkvirtualenv bymur-dev
```

4. Workaround to make wxWidget works

While the new virtualenv is activated link the external wxPython module inside it

```
(bymur_dev)$ ln -s /usr/lib/python2.7/dist-packages/wx*  
$VIRTUAL_ENV/lib/python2.7/site-packages/
```

5. Check out the git repository

```
(bymur_dev)$ git clone git@gitlab.bo.ingv.it:perfetti/bymur.git
```

6. Install requested python modules

```
(bymur_dev)$ pip install --upgrade distribute  
(bymur_dev)$ pip install -r requirements.txt
```

Updated and detailed install procedure can be found in project's *Readme* file on Gitlab³

1 <https://virtualenv.pypa.io/en/latest/>

2 <http://virtualenvwrapper.readthedocs.io/en/latest/>

3 <https://gitlab.bo.ingv.it/perfetti/bymur/blob/master/README.md>

C. Database internals

ByMuR-DB is a database model specifically designed to store and properly organize all the data required to perform a multi-risk analysis in tables which are connected each other with well defined relations (Figure 12). ByMuR-DB has been realized using MySQL⁴, one of the most used and open source relational databases management systems (RDBMS), and compatible with existing databases developed to store volcanological data as DIVO [Selva et al., 2009] and WOVOdat [Ratdomopurbo et al., 2013].

All the data relating to inventory assets, areas definition, reference grids and point's coordinates are stored in ByMuR-DB and referenced by the models data entries to avoid unfavorable replications.

Four kinds of data models are stored in the database: hazard, fragility, loss and risk. All the data and the corresponding uncertainties are also stored in dedicated tables and their referential integrity is ensured by the use of primary and foreign keys constraints provided by MySQL. The file format specifications used to exchange data are thoroughly explained in the following sections. This XML-based format is used to feed models and data to the ByMuR procedures without the need of any SQL syntax knowledge to import data in ByMuR-DB.

The utilization of an external data repository permits the separation of the two different functionality of the ByMuR program. Indeed, the data storage is completely independent from the visualization framework, enabling the possibility to modify (or even completely redesign and rewrite) the GUI without any concern about the data. Vice versa, the decoupling between GUI and storage is obtained by the use of a Model-View-Controller software architecture to unbind visualization functionalities from physical data memorization and application logic (GoF, 1995), as described in Appendix A.

4 <http://www.mysql.com>

D. XML file formats

The XML schema has been defined to ease the exchange of datasets originated from different groups, sources or algorithms. The inspiration has been taken from the formats defined by GEM⁵ (Global Earthquake Model) and the XML schema defined by the sub-project Natural hazards' Risk Markup Language (NRML) for hazard, fragility and loss models has been modified to meet the ByMuR needs.

An initial simplification of the XML Schema Definition (XSD) was done to avoid useless complexity in parsers (e.g. no continuous functions are used in ByMuR) while later, to reach the expressive power required, some more tag and attribute have been added. Indeed GEM formats are designed to manage seismic hazard and risk only, while ByMuR project is primarily focused on risks comparison between different origin phenomena. Because of this, the first necessary modification was the introduction of specific XML tags and attributes along all file formats, allowing users to specify everywhere which kind of hazard they are referring to. In all exchange files, epistemic uncertainty of results is reflected in the presence of multiple files, one for each percentile taken in exam plus one for the average hazard (or fragility, loss, risk).

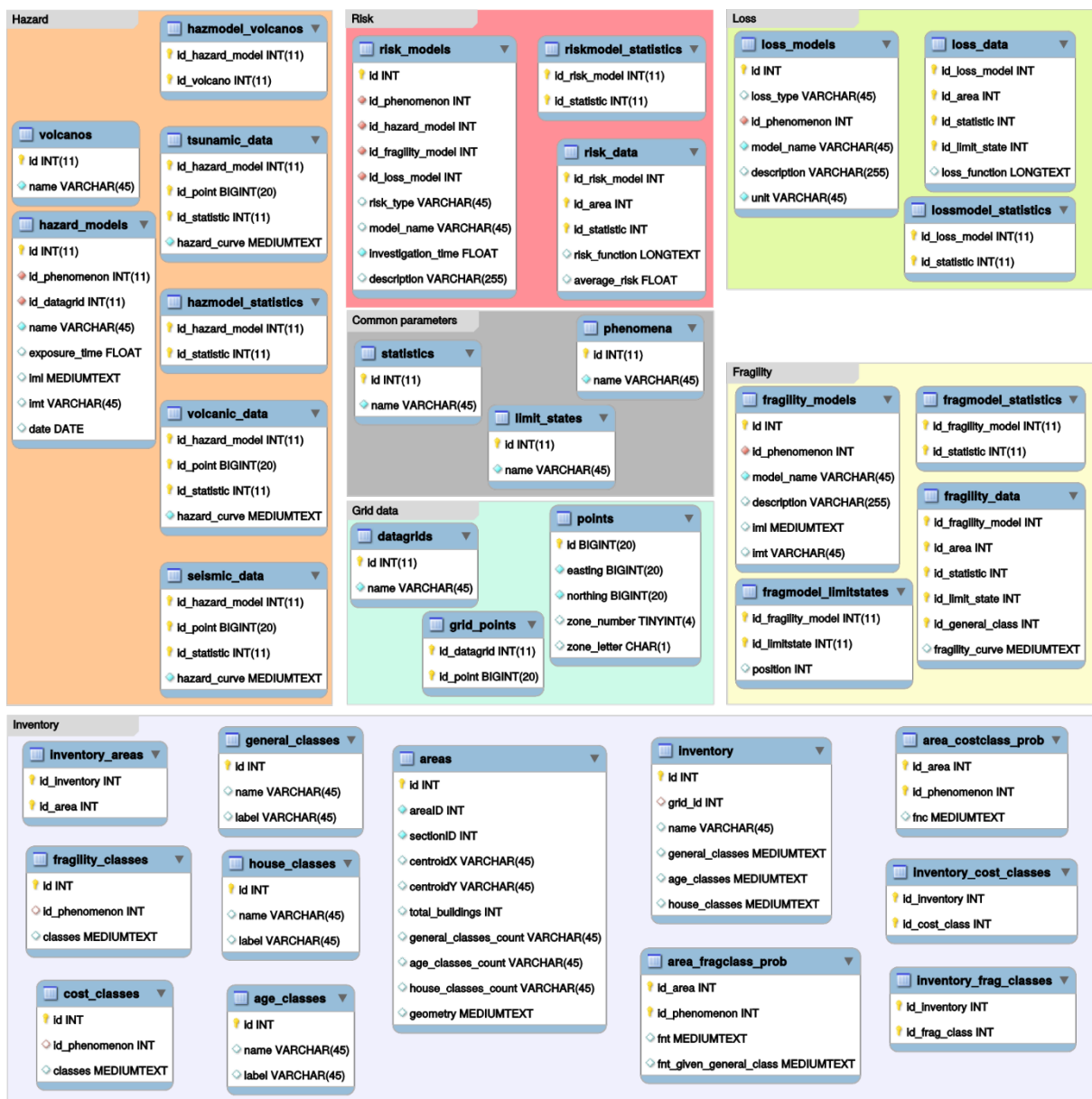


Figure 12. ByMuR-DB SQL schema.

5 <http://www.globalquakemodel.org>

The effort to keep a rough compatibility with original file specifications has determined some ugliness and semantic inconsistency in ByMuR XML schema. Because of this and because of possible changes of the XML schema port by GEM⁶, the ByMuR schema could be refined and slightly modified in the future – hopefully guaranteeing retro-compatibilities.

D.1 Hazard XML format

As previously stated, the first need was to distinguish hazardous phenomenon of different origin. Three mutually exclusive tags were introduced: *volcano*, *tsunami* and *completeness* to determine which kind of phenomena the model is referred to (Figure 13). Also the *hazardModel* tag has been introduced with an attribute *Model*: the tag contains the single hazard model name, while the attribute contains the name of theoretical model used to calculate hazard values.

```

<xs:element name="hazardResult">
  <xs:complexType>
    <xs:sequence>
      <xs:choice minOccurs="0" maxOccurs="1">
        <xs:element name="volcano">
          <xs:complexType> <xs:attribute name="volcanoName" type="xs:string"/> </xs:complexType>
        </xs:element>
        <xs:element name="tsunami">
          <xs:complexType> <xs:attribute name="tsunamiName" type="xs:string"/> </xs:complexType>
        </xs:element>
      </xs:choice>
      <xs:element name="completeness" minOccurs="0" maxOccurs="1">
        <xs:complexType>
          <xs:attribute name="TypeComple" type="xs:string"/>
        </xs:complexType>
      </xs:element>
      <xs:element name="hazardModel" minOccurs="1" maxOccurs="1">
        <xs:complexType>
          <xs:simpleContent>
            <xs:extension base="xs:string" <xs:attribute name="Model" type="xs:string"/> </xs:extension>
          </xs:simpleContent>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
</xs:element>

```

Figure 13. Hazard XML schema.

D.2 Inventory XML format

Inventory file is expected to contain:

- generic buildings class definitions;
- phenomenon specific buildings fragility and cost classes definitions;
- areas definitions with geometry, centroid, buildings number and probability distribution between general, fragility and cost classes.

In ByMuR, the inventory file is originated from the most updated ISTAT shape files, available for the municipality of Naples. In Figure 14, an example is shown.

D.3 Fragility XML format

Fragility models (as for loss and risk) are calculated on the basis of the areas defined by the user and the root tag is called *arealFragility*. A curve is defined for each area, each general building class (specified by the *taxonomy* tag attribute), and for each limit state. As in the hazard case, the most noticeable modification from GEM’s original specification consists in the addition of the attribute *hazardType* in the root tag (Figure 15).

D.4 Loss XML format

For each area in a loss model there are multiple curves to plot, one for each limit state defined in the corresponding fragility model. This correspondence is not enforced in *areasLossModel* tag, but in the resulting risk model XML file, where all models that contribute to risk computation are listed in main tag attributes (Figure 16).

6 <http://micheles.github.io/2014/09/14/dropping-XMLSchema-support/>

D.5 Risk XML format

Risk models are the final results of computation and the attributes *hazardModelName*, *fragilityModelName* and *lossModelName* are specified in the root tag *arealRiskModel* (Figure 17). As in hazard models, ByMuR uses model names and forecast time windows to compose the key to identify a specific risk model and, thus, it is important to keep this duple unique across considered datasets.

```
<arealRiskModel modelName="ByMuR_SEISMIC" hazardType="SEISMIC" format="discrete"
  fragilityModelName="SEISMIC-BCM" hazardModelName="ENSEMBLE_VS30"
  investigationTime="50" lossModelName="SEISMIC-truncGauss" riskType="structural" statistics="mean">
  <riskCurve arealID="213">
    <poEs>0.24417 0.24417 0.24417 0.24414 0.24358 0.242 0.23945 0.23628 0.23221 ... 0 0 0 0 0 0</poEs>
    <losses>0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5 1.6 ... 27.4 27.5 27.6 27.7 27.8</losses>
    <averageRisk>0.57246</averageRisk>
  </riskCurve>
```

Figure 17. Risk XML example.

D.6 Dataset files validation

During the development of ByMuR, an helper program called *bymur_validate.py* has been created to check the syntax of dataset files and to validate them against properly defined XML schemas. The script interprets the first argument as selector of which XML schema verify and remaining arguments on command line as file names to scan (see Figure 18). The goal of this script is to give to contributors an easy and comfortable tool to check their own files, reducing the necessary interaction to successfully load the datasets.

```
$ python bymur_validator.py -h
usage: bymur_validator.py [-h] [-q] (-H | -F | -R | -L) [files [files ...]]
```

Tool to validate ByMuR XML files

positional arguments:

files file list

optional arguments:

-h, --help show this help message and exit
-q, --quiet Suppress normal output, print malformed files only
-H, --hazard Input files are hazard model XML files
-F, --fragility Input files are areal fragility XML files
-R, --risk Input files are areal risk XML files
-L, --loss Input files are areal loss XML files

Figure 18. bymur_validator.py help text.

Glossary

- **Aleatory uncertainty:** the intrinsic randomness of a process or phenomenon.
- **Database:** organized collection of data, made of schema, tables, queries and other objects.
- **Dataset:** the collection of data and files needed to model a function, in this context a hazard/loss/fragility/risk model. A dataset represents the smallest unit of information with an independent meaning.
- **Decision support system:** computer-based information system that supports management and operations decision-making activities.
- **Ensemble:** a statistical mixing among two or more alternative status of system.
- **Epistemic uncertainty:** the lack of knowledge of a processes or phenomenon.
- **Exposure:** all the elements (people, structures, etc.) present in hazard zones that are thereby subject to potential losses.
- **Exposure time:** the time window considered for which the element at risk are exposed to the hazard.
- **GUI:** Graphical User Interface, to allow user to interact with program using graphical icons and widgets.
- **Hazard:** any source of potential damage or harm on human beings, their settlements or the environment
- **Hazard curve:** it shows the exceedance probability (y axis) as function of different intensity thresholds (x axis) in a specific target point.
- **Hazard map:** it shows the spatial intensity for a selected exceedance probability. These maps are obtained from the horizontal intersection between all the hazard curves in the target domain and a selected probability threshold.
- **Intensity:** the unit used to measure the selected hazard (e.g., the thickness of volcanic ash for explosive eruptions, the height of the water column for tsunamis, the peak ground acceleration for earthquakes, etc).
- **Limit state:** the capacity of a structure to satisfy a particular level of performance.
- **Multithreading:** it is a programming and execution model that allows multiple threads (sequences of instructions) to exist within the context of one process. These threads share the process's resources, but are able to execute simultaneously and independently.
- **MVC:** software architectural pattern that divides a given software application into three interconnected parts, so as to separate internal representations of information from the ways that information is presented to or accepted from the user.
- **Probability map:** it shows the spatial exceedance probability for a selected intensity threshold. These maps are obtained by vertical intersecting the intensity threshold with all the hazard curve in each point of the target domain.
- **Return period:** the inverse of the annual probability of exceedance (e.g., the return period of 975 years corresponds to a probability of 1% to exceed a value of intensity).
- **RDBMS:** the Relational DataBase Management System is a computer application that interacts with the user, applications and databases to manage, analyze and interact with data structured in a relational model.
- **Risk:** the potential disaster losses (in terms of lives, health status, livelihoods, assets and services) which could occur to a particular community or a society in a specified time period (see exposure time).
- **XML:** the Extensible Markup Language is a markup language defined by the W3C Specification. As suggested by the name it has been formalized to be extensible to reflect field-specific semantics with the necessary syntax.

Quaderni di Geofisica

ISSN 1590-2595

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/quaderni-di-geofisica.html>

I Quaderni di Geofisica coprono tutti i campi disciplinari sviluppati all'interno dell'INGV, dando particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari, che per tipologia e dettaglio necessitano di una rapida diffusione nella comunità scientifica nazionale ed internazionale. La pubblicazione on-line fornisce accesso immediato a tutti i possibili utenti. L'Editorial Board multidisciplinare garantisce i requisiti di qualità per la pubblicazione dei contributi.

Rapporti tecnici INGV

ISSN 2039-7941

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html>

I Rapporti Tecnici INGV pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico e di rilevante interesse tecnico-scientifico per gli ambiti disciplinari propri dell'INGV. La collana Rapporti Tecnici INGV pubblica esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. L'Editorial Board multidisciplinare garantisce i requisiti di qualità per la pubblicazione dei contributi.

Miscellanea INGV

ISSN 2039-6651

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/miscellanea-ingv.html>

La collana Miscellanea INGV nasce con l'intento di favorire la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV (sismologia, vulcanologia, geologia, geomagnetismo, geochimica, aeronomia e innovazione tecnologica). In particolare, la collana Miscellanea INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc.

Coordinamento editoriale e impaginazione

Centro Editoriale Nazionale | INGV

Progetto grafico e redazionale

Daniela Riposati | Laboratorio Grafica e Immagini | INGV

© 2018 INGV Istituto Nazionale di Geofisica e Vulcanologia

Via di Vigna Murata, 605

00143 Roma

Tel. +39 06518601 Fax +39 065041181

<http://www.ingv.it>



Istituto Nazionale di Geofisica e Vulcanologia