D 3.6

DELIVERABLE

PROJECT INFORMATION

Project Title:	**Harmonized approach to stress tests for critical infrastructures against natural hazards**
Acronym:	**STREST**
Project N°:	603389
Call N°:	FP7-ENV-2013-two-stage
Project start:	01 October 2013
Duration:	36 months

DELIVERABLE INFORMATION

Deliverable Title:	**New Software Package Incorporating Induced Seismicity Hazard in PSHA**
Date of issue:	27 March 2015
Work Package:	WP3 Integrated low probability-high consequence hazard assessment for critical infrastructures
Editor/Author:	Graeme Weatherill (EUCENTRE)
Reviewer:	Fabrice Cotton (UJF)

REVISION: Version 2

SEVENTH FRAMEWORK PROGRAMME

Project Coordinator:	Prof. Domenico Giardini
Institution:	ETH Zürich
e-mail:	giardini@sed.ethz.ch
fax:	+ 41 446331065
telephone:	+ 41 446332610

Abstract

Earthquakes caused by anthropogenic activities can present a risk to both the physical and social infrastructure of those populations affected by them. As a potentially unavoidable consequence of activities that may otherwise benefit a population (e.g., energy generation or revenue from hydrocarbon extraction), it has become increasingly necessary to balance the potential physical and socio-economic consequences of an induced earthquake sequence against the benefit of undertaking such an activity. Probabilistic analysis of induced seismicity hazard and risk is therefore a valuable, even essential, tool for management of man-made processes that may generate earthquakes. To illustrate how induced seismicity models can be integrated with existing state-of-the-art tools for analysis of earthquake hazard and risk, a customised application is developed that incorporates the hazard modelling capabilities of OpenQuake-engine (an open source software for analysis of earthquake hazard and risk), with simulations of induced seismicity from an anthropogenic process. This process highlights the modifications made to the OpenQuake-engine itself to facilitate induced seismicity hazard analysis, such as the incorporation of induced seismicity ground motion prediction equations. The application then incorporates components of the OpenQuake hazard modelling software with the geomechanical seed approach for modelling the seismic response of an enhanced geothermal system.

Keywords: Induced Seismicity, Probabilistic Hazard Assessment, OpenQuake, Geomechanical Seed

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme [FP7/2007-2013] under grant agreement n° 603389

Deliverable Contributors

EUCENTRE Graeme Weatherill

Table of Contents

List of Figures

List of Tables

1 Introduction

The threat that earthquakes pose to the built environment has been well recognised throughout the course of history. As the infrastructure systems within towns and cities become increasingly complex and inter-dependent, whilst the spatial extent of that same environment expands further, the systemic risk from seismicity is continuing to grow. Throughout the 20th century, however, a new type of seismic threat has emerged from anthropogenic causes, which, whilst rarely carrying the massive destructive potential associated with tectonic earthquakes, can be a cause of damage (both in an engineering sense as well as an economic and political sense) within the regions affected. These seismic events that are man-made in origin are known as “induced” events. Induced seismicity has been identified as a potential consequence of several geo-engineering processes. From the early to mid 20th century, increases in seismicity were identified areas of reservoir impoundment (Gupta, 2002), and had already been known as a potential risk associated with mining. More recently, however, technologies in the energy sector have been identified, virtually unambiguously, as the cause of induced seismicity sequences. The most common of these is have been projects for injection or extraction of fluids into the shallow crust; examples being reservoir stimulation for geothermal energy production, hydraulic fracturing associated with shale gas extraction, CO_2 and waste water storage, as well as conventional oil and gas extraction.

The attribution of seismicity to anthropogenic causes opens up a variety of complex challenges, both scientific and political. Damage to property, or loss of economic productivity due to failure of elements within an infrastructure system, cause by man-made seismicity requires careful assessment of the cost-benefit of undertaking activities that may be likely to induce earthquakes (e.g. Bommer et al., 2015). In the case of energy technologies, it is frequently those communities at risk from the seismicity that may stand to benefit the most from the causative activity, be it in terms of energy supply, economic wealth, employment etc. Nevertheless, operators of such technologies may be held financially liable to compensate in the cases that damage is caused.

Assessment of the risk from induced seismicity is a cornerstone of an effective mitigation strategy. For induced seismicity, the risk assessment may operate in real-time, in cases that a triggering process may be modulated or terminated if it is seen that the seismicity is increasing to an unacceptable level (e.g. Bommer et al., 2005). It may also operate as part of a planning strategy, in which the likelihood of damage, and the associated economic cost, are assessed in order to optimise the benefit to cost ratio of production versus risk. It is evident, therefore, that quantitative models of both the physics of the induced seismicity process as well the potential effects on the built environment, are increasingly necessary.

1.1 CHARACTERISTICS OF INDUCED SEISMICTY HAZARD

Induced seismicity earthquake sequences display characteristics of behaviour that may prevent the direct application of models of behaviour that are more commonly defined for tectonic events. With only a handful of very notable exceptions, induced seismicity earthquakes are typically small in magnitude, often within a range $M_W < 4$. This will naturally

limit the likelihood of causing catastrophic damage, but their shallow depths and frequent recurrence resulting from the anthropogenic process can increase the likelihood of causing some damage. Certainly this contributes strongly to the perception of them as a nuisance, causing low-level damage but with sufficient frequency as to affect the social response of the affected population. The persistence of the sequences arises from the rapid reloading of the anthropogenic stresses on the potential weaknesses in the crust, at a rate that is often far greater than the reloading of tectonic stresses for similar natural events. Depending on the mechanism of anthropogenic seismicity it may be the case that this rapid reloading is limiting the ability of ruptures to propagate to greater extents on the fault. This may limit the potential size than an earthquake can reach, and thus the maximum magnitude that needs to be considered in the hazard assessment.

From an engineering perspective such small events are typically below the minimum magnitudes often considered in engineering design. This generally places the resulting strong motion records below the minimum limit of applicability of ground motion prediction equations (GMPEs) that have been fit to strong motion records from larger tectonic earthquakes ($M_W \geq 4.5$). Consequently additional effort is needed to calibrate the ground motion models to the local effects. Bommer et al. (2007) demonstrate the extent to which GMPEs derived for larger ground motions fail to capture the scaling with magnitude for small earthquakes. Furthermore, Chiou et al. (2010) demonstrate that strong motion records from smaller events may demonstrate stronger regionalisation, with local characteristics of the source, path and site playing a greater role in the overall ground motion distribution than is typical for larger tectonic events.

Induced events display temporal recurrence characteristics that are often different from those that can often be assumed of tectonically driven earthquakes. This is, of course, due to the dynamic nature of the anthropogenic process that is the forcing mechanism of seismicity. This presents some additional challenges, as the long-term temporal stationarity often assumed in tectonic seismicity models cannot be applied in this case. Instead the recurrence modelling must take demonstrate the ability to take into account short-term changes in the baseline seismicity in response to the changes in loading from anthropogenic forcing.

1.2 A SOFTWARE FOR MODELLING INDUCED SEISMICITY SEISMIC HAZARD

It must be clarified early within this report that a proposal for a generic “induced seismicity” hazard assessment is not a feasible goal, as the physical mechanisms and the models of earthquake occurrence may vary considerably from one type of induced seismicity to the next (Bommer et al., 2015). A more realisable objective is to identify those parts of the induced seismicity hazard process that may share common elements with seismic hazard from tectonic events. In this context, it becomes possible to separate the modelling problem into those parts that are specific to the nature of the induced seismicity in question, and those parts that may be common to all forms of seismic hazard, and for which well-established modelling techniques can be applied. In the case of seismicity the preferred quantitative modelling approach is probabilistic seismic hazard analysis, or PSHA hereafter (Cornell, 1968; McGuire, 1976).

To demonstrate how to incorporate elements of more conventional PSHA and apply them to an induced seismicity case, the OpenQuake-engine software (Pagani et al., 2014) for seismic hazard and risk assessment is the current tool of choice. As a state-of-the-art, free and open-source software, the mechanics of the OpenQuake calculation engine are favourable for just such an activity.

In this report the aim is to demonstrate how OpenQuake can be linked with models of the spatial and temporal distribution of induced seismicity, for the assessment of the ground-shaking hazard at the surface. In the following chapter an initial appraisal is undertaken of several different approaches for modelling induced seismicity hazard that have been published in the available literature. Chapter 3 will provide an overview of the operation, features and mechanics of the OpenQuake-engine, before a customised application of OpenQuake for an enhanced geothermal system, based on the geomechanical seed model of Goertz-Allmann & Wiemer (2013), is presented in Chapter 4.

2 Induced Seismicity Hazard

To appraise methodologies previously applied to induced seismicity hazard modelling applications, it is important to recall that induced seismicity itself is an umbrella term incorporating many different types of anthropogenic seismic phenomena. Consequently a "generalised" approach that can be applied across the variety of different phenomena (e.g., geothermal injection, mining, shale gas extraction, reservoir induction, etc.) may not be feasible (Bommer et al., 2015). The focus of this appraisal is placed not necessarily on the specific causative mechanism of the induced events but the extent to which the analysis of the ground shaking can be characterised in a probabilistic seismic hazard context. The aim of the appraisal is to identify the common elements within the different process for which a single "induced seismicity hazard module" may provide a natural extension for the purposes of characterising the hazard and risk. Methodologies that depend on a detailed geomechanical characterisation of the fault system will not necessarily be the primary focus.

2.1 TIME-INDEPENDENT INDUCED SEISMICITY

Arguably the simplest approach that can be taken in the characterisation of induced seismicity hazard is to apply a time-independent model in which the long-term rate of earthquakes from the anthropogenic source is considered to be stationary. This approach is effectively a standard application of probabilistic seismic hazard in the manner described by Cornell (1965) and McGuire (1976). A widely cited example of such a model is that of van Eck et al. (2006) for characterisation of seismic hazard due to induced earthquakes in the Netherlands due to natural gas extraction. The assumption of stationarity in the seismic process applied by van Eck et al (2006) is seemingly justified by the observation that the total rate of gas extraction, and similarly the total seismic moment rate, were shown to be relatively stationary. Under these conditions it then becomes possible to specify a long-term rate, assuming the same levels of gas extraction, which can then form the basis of an activity rate for the induced seismogenic source. Lasocki (2005) makes a similar assumption in estimating seismic hazard from mining induced seismicity in Poland.

In both of the above cases the time-independent model may be seen as justifiable if one or more of the following conditions are true:

i) The rate of the forcing mechanism (e.g. gas extraction, mining activity, reservoir load) can be considered approximately stationary within the duration of interest.

ii) The duration of interest may be small enough as to justify a Poissonian model of recurrence in a system in which the dynamic response is slow.

One element to be considered within the process of characterisation of seismic hazard for induced seismicity is that the anthropogenic sources are still embedded within a broader tectonic system, whose seismic hazard may make a small, but not necessarily negligible, contribution to the total seismic hazard at the site. In cases where the tectonic stresses are extremely low, as in very low seismicity regions, it may be acceptable to ignore the tectonic contribution to the overall activity rate. In other places the tectonic rate may form a baseline, below which the induced seismicity rate cannot fall.

2.1.1 Source Geometry Characterisation

The characterisation of the source geometry in is inevitably dependent on the nature of the specific induced seismicity phenomenon in question. In cases where the stress changes on specific faults cannot be identified it is most common to consider the seismogenic source as an area of either uniform seismicity or, in the case of enhanced geothermal systems, an area of seismicity with occurrence probability diminishing further away from the injection point. Constraints upon the depth of the seismogenic source will of course depend on the mechanism by which the seismicity is induced. For enhanced geothermal systems the hypocentral depth distribution probabilities may diminish with increasing distance from the injection source. For natural gas extraction and/or reservoir impoundment the spatial distribution of the seismicity, even if assuming temporal stationarity, may be represented as either an area of uniform seismicity or, if the stress changes on specific fault structures are known, a collection of individual sources

2.1.2 Magnitude Frequency Distribution and Maximum Magnitude

For induced seismicity applications in areas of low to moderate seismicity, the magnitude of the induced earthquakes has, with few exceptions, been small compared to those of more damaging tectonic events. In both the catalogues of induced seismicity that have been observed, and from the physics of the rock fracture process, the size distribution of slip events should generally follow the Gutenberg & Richter (1944) power law:

$$Log_{10}\left(N_C\right) = a - bM \tag{B.1}$$

where N_C is the cumulative number of earthquakes greater than or equal to M, and *a* and *b* are the intercept (for M = 0) and the slope of the distribution. In seismic hazard practice, it is necessary to place upper and lower bounds on the magnitude considered within the magnitude frequency distribution, known as the minimum and maximum magnitude respectively. In the case of time-independent hazard, such as van Eck et al. (2006), the *a*- and *b*-values for the source are fit to the observed seismicity.

Estimation of the maximum magnitude of earthquakes from an induced seismicity may often be inferred from the observed magnitude frequency distribution using statistical techniques (e.g. Kijko, 2004), as opposed to placing constraints on the upper bound from geological information. Van Eck et al. (2006) consider two options, one based on the total cumulative strain energy released over time within the induced system (Makropolous and Burton, 1983), the other based on a Bayesian estimate of the fit of a truncated magnitude frequency distribution. Both approaches produced similar results. Given the potential uncertainties in such an approach, however, it may be more prudent to treat M_{MAX}, or preferably the combination of *a*-, *b*- and M_{MAX}, as an epistemic uncertainty that may be represented by alternative branches on a logic tree (Mignan et al., 2015).

As mentioned previously, for the purpose of analysing seismic hazard within a given time period, it may still be necessary to consider the contribution from tectonic events within the same system. When characterising the magnitude frequency distribution this may present a challenge to the modeller, as it may be determined that the upper bound magnitudes from tectonic events, inferred from, for example, regional proxy distributions (e.g. Johnston et al., 1994). The seismic hazard modeller may choose to address this situation in one of several ways.

The simplest approach may be to consider the anthropogenic source as one that may operate independently of the tectonic source, and as such may simply be superimposed as an additional source in the model. This may be a suitable assumption if it can be demonstrated that distributions of the hypocentral depths of the induced events may be significantly different from those of tectonic events observed in the region. For systems where the source is highly localised, as in the case of enhanced geothermal systems, such a separation can be trivial to make with a low possibility of counting the moment rate from tectonic events within the anthropogenic rate. Furthermore, it is possible to assign a different maximum magnitude for the two sources.

In other systems, especially those where the spatial footprint of the anthropogenic source may be such that the distribution of induced and tectonic events is not well-separated, or those in which the anthropogenic process may be seen as acting to alter the probabilities of occurrence of tectonic events. Hydraulic reservoir impoundment may be a classic example of such a case in which the initial reservoir loading, followed by the subsequent diffusion of water into the surrounding rock, may act to trigger events whose primary stresses remain tectonic. As the anthropogenic and tectonic stress contributions are not inherently separable, the magnitude frequency distribution may reflect the combination of the two elements.

2.2 TIME-VARYING SEISMIC HAZARD

The assumptions presented for considering induced seismicity hazard, as potentially a time-independent model, may frequently be inappropriate to the context of the hazard application. By its very nature, induced seismicity hazard is dependent on activities that will vary with time. The temporal changes in the anthropogenic forcing may be direct and immediate as a result of mitigation decision, such as controlling the rate of fluid injection (for enhanced geothermal systems) or extraction. Furthermore, mitigation of the effects of induced seismicity is dependent upon a short-term assessment of the seismic hazard and risk in order to assess the economic cost-benefit of the mitigation action (e.g. Bommer et al., 2006, 2015; Mignan et al., 2015). In this context an assessment of seismic hazard is needed on a shorter time-period of hours to months, and the frequent re-appraisal of the hazard in response to observations of the seismicity sequence must be possible.

Analysis of time-dependent need not necessarily to be considered only as a means of enacting a short-term response, but also as a long-term assessment of the potential impacts of a particular management strategy (e.g. NAM, 2013; Bommer et al. 2015). The example of the Groningen Natural Gas field will be discussed in further detail in due course; however, it offers a clear example as to how assessment of the potential seismic risk may come to influence the gas extraction strategy to balance the economic productivity against the potential economic consequences of the resulting earthquakes. In this case the expected production strategy may span over several decades, identifying areas of productivity within the gas field and their likely geodetic compaction in response to the volumetric change caused by extraction of gas. Similar issues may be considered in the case of injection in the case of long-term carbon storage. To assess in a quantitative manner the ground-shaking hazard associated with each strategy, geomechanical models of deflation are coupled with seismic models to predict the occurrence of seismicity and its likely magnitude frequency distribution. These can be used in conjunction with a Monte Carlo seismic hazard process (Ebel & Kafka, 1999; Musson, 1999; Weatherill & Burton, 2010; Assatorians & Atkinson, 2013) to generate multiple realisations of seismicity and ground shaking, from which the

probability of exceeding a given hazard or risk level can be determined (Bourne et al., 2014, 2015).

In the following discussion we shall outline several examples of time-dependent seismic hazard assessment for induced seismicity applications. These examples are selected because of their focus specifically on producing estimates of ground shaking exceedance probabilities within a short time period.

2.2.1 Convertito et al. (2012)

The first approach is that of Convertito et al. (2012) who construct a short-term time-dependent model for the Geysers Geothermal field, Northern California, by adopting the conventional PSHA formulation for use with Gutenberg-Richter *a*- and *b*- values that are revised at regular intervals. During the course of an induced seismicity sequence, the probability of exceeding a given level of ground shaking at a particular probability of exceedance are determined by projecting the rate and b-value recorded within the most recent observation period, forward in time to the hazard period of interest (typically on the order of 2 months). Maximum magnitudes are assumed to be held constant and are estimated using the same cumulative strain energy method employed by van Eck et al (2006). The ground motion for each event is determined from the induced seismicity GMPE derived from observed data in the geothermal field.

This approach is a conceptually simple way of trying to adapt the expected seismic hazard within a given time-period in order to reflect the seismicity response of the injection process. In the application to the Geysers Geothermal Field the source model consists of two area sources of uniform seismicity. This relative simplification of the spatial distribution allows for the sources to be adopted directly within a standard PSHA framework, whilst ensuring that each zone is large enough for a sufficient number of events to be recorded in a short time period as to allow for statistically robust estimates of the recurrence parameters.

Whilst the convenience and conceptual simplicity of this approach may be appealing, there are limitations that may make its use unsuitable for forecasting hazard during the induced seismicity sequence. The frequent updating of *a*- and *b*- values, and their subsequent projection forward throughout the forecast period, still requires that the seismicity be assumed to be a stationary Poisson process for the duration of the forecast. This has two consequences; the first is that aftershocks and potential swarms of events cannot be incorporated into the analysis. Within such a short time window, an aftershock sequence may significantly change the overall likelihood of shaking, yet the methodology cannot model aftershock decay within the time frame. Secondly, although observed *a*- and *b*-value may be correlated with properties of the injection process (as shown in Figure 2-1), there is considerable variability in this correlation. Furthermore there is a considerable lag between the occurrence of an intervening action (such as terminating, or substantially increasing or decreasing the rate of injection) and the re-assessment of the b-value. Depending on the circumstance this lag may be prohibitive for operators to be able to utilise such information to mitigate the hazard. An example of the time-varying hazard maps is shown in Figure 2-2.

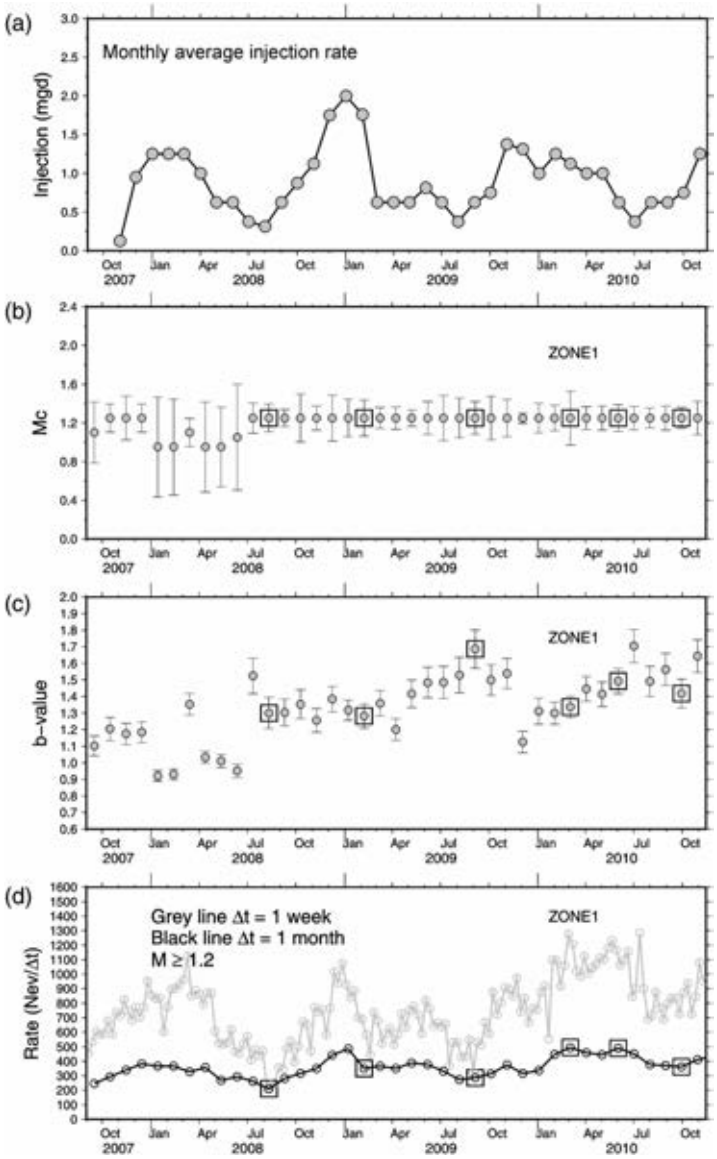

Figure 2-1 Time-varying trends in seismicity parameters with monthly injection rate for the Geysers Geothermal Field, from Convertito et al. (2012). a) Monthly injection rate, b) completeness magnitude, c) b-value and d) Seismicity rate

m

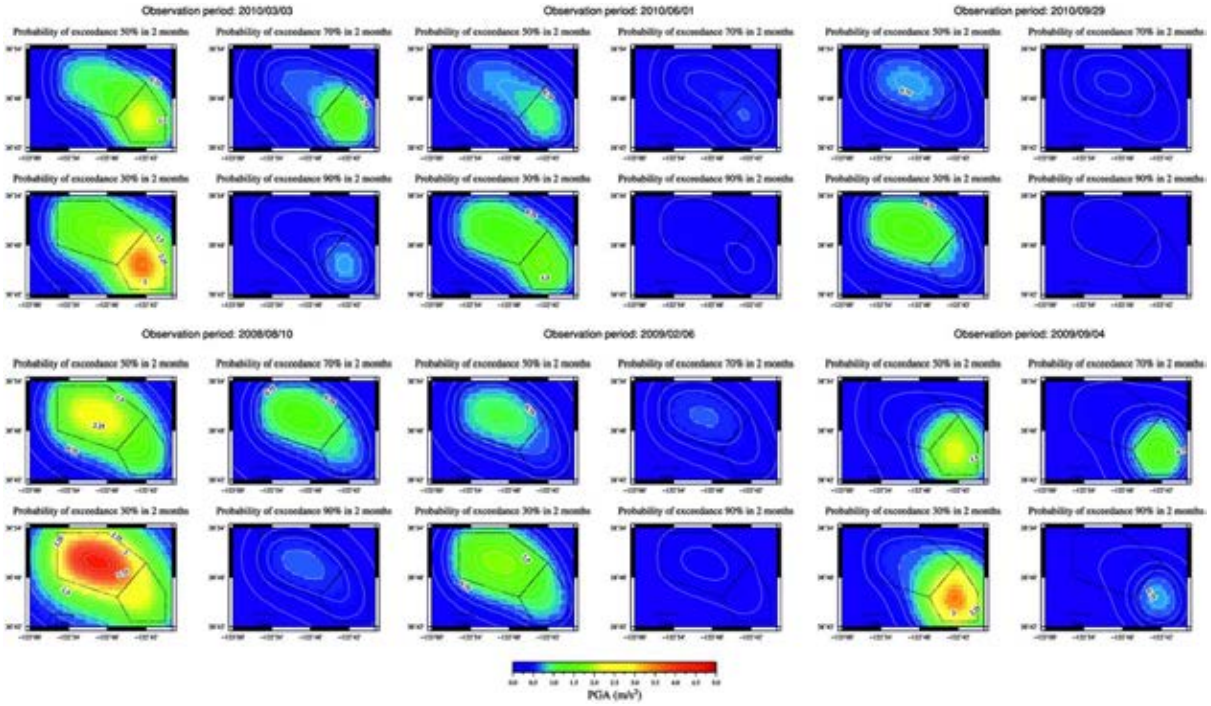

Figure 2-2. Time-varying probabilistic seismic hazard maps for the Geysers Geothermal field based on the approach of Convertito et al. (2012).

2.2.2 Mena et al. (2013) and Mignan et al. (2015)

These two studies present a methodology for the analysis of seismic hazard and risk based on the Basel Enhanced Geothermal System project. The first paper (Mena et al., 2013) outlines different approaches for the characterisation of the activity rate in accordance with the injection process. The second paper outlines an approach for analysis of seismic risk, with a particular focus on epistemic uncertainties, using the seismicity recurrence models presented in the first paper. As a full probabilistic seismic hazard analysis is described within these two papers, further exploration of the seismic hazard model presented therein is warranted. Several elements of the process described resolve, at least partially, some of the shortcomings of the Convertito et al (2012) approach. In particular, Mena et al. (2013) incorporate more commonly used models of time-dependent seismic hazard into the analysis process in order to characterise potential changes in earthquake probability due to triggering of aftershocks.

The main objective of Mena et al. (2013) is the forecasting of seismic activity rates for a 15-day period at the initiation of the injection experiment. Three activity rate models are compared:

1. Reasenberg & Jones (1989)

The Reasenberg & Jones (1989) model is a simple model for estimating the rate of earthquakes greater than or equal to magnitude M, at a time *t* following a mainshock event of a given magnitude M_{MS}:

$$\lambda\left(t,M\right)=\frac{10^{a+b\left(M_{MS}-M\right)}}{\left(t+c\right)^{p}} \tag{B.2}$$

where *a* and *b* are the parameters of the Gutenberg & Richter distribution, and *c* and *p* are the coefficients of the modified Omori law (Utsu, 1961) describing the decay rate of the aftershock activity following the mainshock.

2. Induced-Seismicity Epidemic Type Aftershock Sequence (ETAS)

The ETAS model (Ogata, 1999) has become established as one of the most robust predictors of aftershock activity rate. The rate of aftershocks greater than or equal to a magnitude M_{MIN} at a time *t* days following an event of magnitude M_i is given by:

$$\lambda_i\left(t\right)=\frac{Ke^{\alpha\left(M_i-M_{Min}\right)}}{\left(c+t-t_i\right)^{p}} \tag{B.3}$$

Within an aftershock sequence, however, it can be seen that each aftershock may itself trigger a sequence of aftershocks, and that the total rate of earthquakes greater than or equal to a threshold magnitude is a superposition of the individual aftershock sequences. Furthermore, and more importantly in this application, the sequence itself is superimposed on top of a longer-term background rate of seismicity (λ_0). Therefore the total rate of earthquakes is given by:

$$\lambda\left(t\right)=\lambda_0+\sum_{i:t_i<t}\frac{Ke^{\alpha\left(M_i-M_{Min}\right)}}{\left(c+t-t_i\right)^{p}} \tag{B.4}$$

The standard ETAS model can be adapted for application to induced seismicity by recognising that the background rate is not stationary, but instead will fluctuate in proportion to the anthropogenic forcing (fluid injection in the case of Mena et al., 2013). Bachmann et al. (2011) propose a simple model to correlate the total background rate in a time period *t* to the injection flow rate ($F_r(t)$):

$$\lambda_0\left(t\right)=\mu+c_f\cdot F_r\left(t\right) \tag{B.5}$$

where μ and c_f are fit to observed flow rate. The modified ETAS model has several advantages in that it takes into account both the short-term decay rate of the aftershock sequences, it also permits the activity rate to continue after the end of the injection process.

3. Shapiro et al. (2010)

The third model, presented by Shapiro et al. (2010) is simpler in construction than that of ETAS. The expected number of events in a time period *t* greater than or equal to a given magnitude M is determined from:

$$\log N_C(t) = \log Q_C(t) - bM + \Sigma \tag{B.6}$$

where Q_C is the cumulative fluid injection volume, and Σ the *seismogenic index* that is dependent on the tectonic setting, and is calculated empirically. The above formulation is only valid whilst injection is occurring. To predict the rate of occurrence after the termination of injection a calibrated Omori law (Langenbruch & Shapiro, 2010) is proposed as a means of tapering the rates after the end of injection:

$$R_{0b}\left(\frac{t}{t_0}\right) = \frac{R_{0a}}{\left(\frac{t}{t_0}\right)^p} \tag{B.7}$$

where R_{0a} is the seismicity rate during injection and t_0 is the time at which injection stops. Whilst simpler in functional form than the ETAS model, the Shapiro et al (2010) model has a similar advantage in its ability to forecast activity in the post-injection period.

Each of the short-term time dependent models requires a relatively detailed parameterisation of the activity rate model. Mena et al. (2013) demonstrate how the parameters can be determined and updated after a relatively short time within the induced seismicity sequence based on observed data. The models can also be combined into a single predictive model, or treated on separate branches within an epistemic uncertainty analysis.

The authors illustrate how these models can be extended into a probabilistic seismic hazard and risk analysis. Theoretically this can be done using either a Monte Carlo approach or a more conventional Cornell (1968) and McGuire (1976) approach (see the subsequent chapter for more details), with aleatory uncertainty treated by the use of multiple random simulations of seismicity, or by direct integration of the ground motion or macroseismic intensity uncertainty.

It should be noted though that neither Mena et al. (2013) nor Mignan et al. (2015) discuss the spatial characterisation of the seismicity within the seismogenic source. As the application context is that of the Basel Enhanced Geothermal system, potentially a point source may have been considered. Király et al. (2014) consider the spatial element of these approaches by adopting a three-dimensional Gaussian kernel, whose standard deviations within each dimension are determined from a principal component analysis of the locations of the seismic events after a sufficient time of the sequence. These are shown in Figure 2-3

The standard Gaussian kernels may not be able to capture the physical phenomenon known as the Kaiser effect, in which activity rates in the vicinity of the injection well decrease after injection has stopped. To attempt to capture this effect, Király et al. (2015) propose the use of a temporal weighting scheme when fitting the smoothed seismicity kernel, one in which more recent events are given higher weighting than those earlier in the sequence.

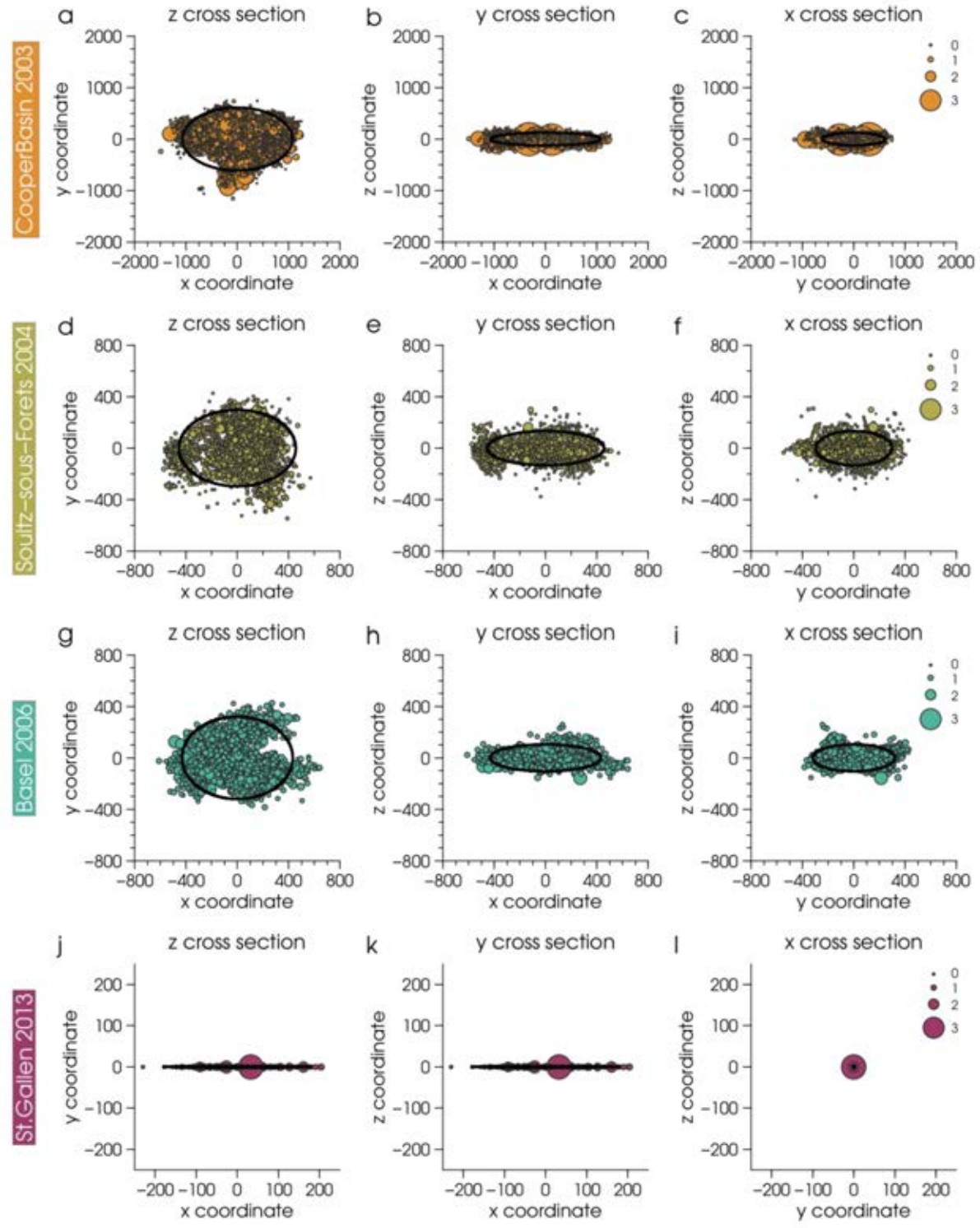

Figure 2-3 Spatial distributions of seismicity in earthquake sequences observed for four different Deep Geothermal Energy projects. Reproduced from Király et al. (2014)

The approaches outlined in the studies of Mena et al. (2013), Mignan et al. (2015) and Király et al (2014; 2015), when taken in conjunction, provide the basis for more widely applicable approach for modelling near real-time seismic hazard and risk from an induced seismicity sequence. The models themselves are largely empirical, requiring the fit of the necessary spatio-temporal activity models to the observed sequence. As Király et al. (2014) highlight, however, when determining the Shapiro et al. (2010) seismogenic index parameter for

different geothermal injection sequences, the potential parameter values may be specific for each injection site, and therefore cannot necessarily be inferred prior to the start of the injection sequence. It should also be noted that assumptions made within both the spatial characterisation of seismicity and the temporal occurrence would not necessarily apply to other forms of induced seismicity.

2.3 GEOMECHANICAL SEED MODELS (GOERTZ-ALLMANN & WIEMER, 2013; GISCHIG & WIEMER, 2013)

The empirical seismicity models shown in section 2.2.2 can be useful in real-time, but are significantly limited for the consideration of seismic hazard and risk outcomes for certain applications. As Gischig & Wiemer (2013) note, by neglecting the physics of the process and relying on observations taken within the seismicity sequence, they may have limited forecasting capabilities for longer time periods (on the order of weeks to months), for exploration of possible injection scenarios and for post shut-in behaviour. The possibility to consider the potential hazard from different injection scenarios may be critical in any cost-benefit analysis of the development of an enhanced geothermal system. Such behaviour may only be described adequately by considering the physical processes that control the system and providing numerical models of the geomechanical evolution of the induced seismicity source. Purely physical models, however, are themselves limited in the case of seismic hazard studies. Arguably the greatest limiting factors of such numerical approaches relate to the poor constraint, and therefore large uncertainties, on many of the parameters controlling the models. The significant computational cost of running the numerical simulations limits the ability of the modellers to integrate the uncertainties into the hazard analysis, a problem that drastically limits their applicability in real-time.

The alternative to a purely empirical or a purely numerical approach is the use of “hybrid” geomechanical seed models proposed for use in enhanced geothermal injection systems by Goertz-Allmann & Wiemer (2013) and Gischig & Wiemer (2013). The geomechanical approach begins by using a numerical model of pore pressure diffusion to determine the spatio-temporal evolution of pore-pressures within a homogeneous and isotropic volume in response to the fluid injection. In the case of Goertz-Allmann & Wiemer (2013) this is done using a simplified analytical model, whilst Gischig & Wiemer (2013) utilise commercial finite element software to generate this information.

Initially a set of seed points is distributed within the isotropic volume surrounding the injection point. The density of seed points may be estimated based on observed data or from considerations of the completeness of the radial crack size corresponding to the minimum magnitude within the volume. For each seed point the maximum and minimum principal stresses (σ_1 and σ_3) are sampled from a Gaussian distribution with the mean values estimated from in-situ values at the site and a variance of 10 %. The samples should be limited within the range of an upper-bound crustal stress limit and the lower bound of the hydrostatic pore pressure. For each of the seeds the normal stress (σ_n) is determined the shear-stress (τ_s) calculated via:

$$\tau_s = \mu\left(\sigma_n - p\right) + c \tag{B.8}$$

where p is the pore pressure, μ is the coefficient of friction and c the cohesion. If the shear stress exceeds the coulomb failure envelope then an induced earthquake is triggered at the location of the seed, otherwise the seed is withdrawn. To sample the magnitude of the event a Gutenberg-Richter model is assumed with b-value calculated from the differential stress ($\sigma_d = \sigma_1 - \sigma_3$) using an empirical fit to the Basel data set:

$$b = -0.022\sigma_d + 4.0 \tag{B.9}$$

This relation is potentially capped such that b-value does not decrease below the tectonic b-value (Gischig & Wiemer, 2013) or above a specified maximum value of approximately $b = 4.0$.

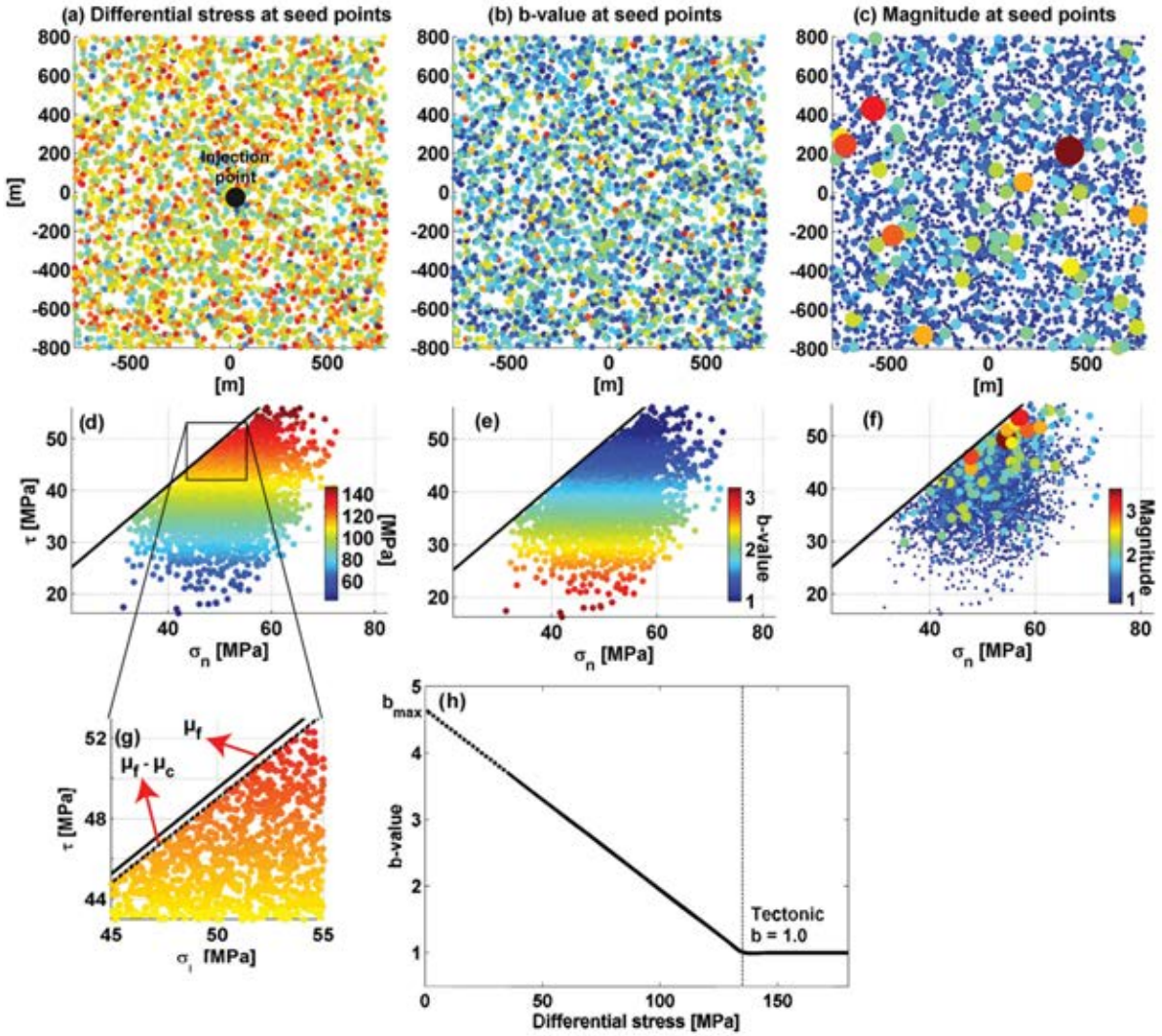

Figure 2-4 Explanation of the geomechanical seed simulation process (reproduced from Gischig & Wiemer, 2013). For a set of randomly located seeds, a) generation of differential stresses and their representation in Mohr-Coulomb space, b) b-value at the seed points and Mohr-Coulomb representation in (e), c) simulated magnitudes and their Mohr-Coulomb representation in (f), g) limit of criticality below the Coulomb-failure, h) the model of b-value with differential stress.

In the last step of the process when an earthquake is triggered from a seed the maximum principal stress σ_1 is reduced by an amount equivalent to approximately 10 % of the value of the differential stress (σ_d). The corresponding reduction in shear stress is taken as the stress

drop ($\Delta\sigma$) of the earthquake. The reduction in σ_1 effectively stabilises the seed point, but instability can be reached again as the pore-pressure increases. Therefore it is possible for the same seed point to trigger multiple events within a sequence.

For application to seismic hazard analysis, geomechanical seed models are capable of providing multiple realisations of seismicity for a given pressure diffusion model. These realisations can be easily combined with models of ground motion in order to provide random realisations of ground shaking for use in a Monte Carlo-based PSHA, as can be seen in Gischig & Wiemer (2013).

Hybrid physical and statistical models such as these can capture physical elements of the process, such as the potential for large magnitudes and the spatial spreading of increased activity away from the site of injection after shut-in. The process does not explicitly consider background tectonic seismicity, although in the applications for which it has been demonstrated the time-scale for the seismic hazard is generally of sufficiently short duration, or the area under consideration sufficiently small, as to minimise the likelihood of observing background events within the time-frame. Also absent from such models are the means of capturing interaction in the stress changes between the seed points once an event has been triggered. Catalli et al. (2013) demonstrate the role of Coulomb stress transfer in explaining both the spatial patterns and the rates of induced seismicity within the Basel Geothermal sequence, particularly after the injection shut-in, or at greater distances from the injection point, where the influence of the pore-pressure changes diminish with respect to the Coulomb stress changes.

As an approach for incorporating elements of the direct physical process into the seismic hazard analysis, the geomechanical seed models offer several significant advantages. Whilst background seismicity may not be incorporated directly, the use of synthetic catalogues for the PSHA does not exclude the possibility that tectonic events (both stationary and non-stationary) can be incorporated into the analysis simply by merging the processes together. Whilst the applications of this process illustrated in Goertz-Allman and Wiemer (2013) and Gischig & Wiemer (2013) are limited to the case of an enhanced geothermal system, it is not unreasonable to expect that elements of the process could be applied in other induced seismicity contexts such as hydrocarbon extraction or storage.

2.4 COMPREHENSIVE GEOMECHANICAL SIMULATION MODELS

The last category of induced seismicity seismic hazard models represents a broad spectrum of approaches that can be used to characterise the seismogenic activity from anthropogenic events using detailed models of geomechanical simulation. It should come as no surprise that within this category it is impossible to talk of a uniform or generic approach to the modelling of seismic hazard, as the exact details of each case a specific to the application in question. Nonetheless, to illustrate future potential of such approaches from within a seismic hazard context, and to attempt to identify those elements of the process that may be common with the applications listed previously, a short overview of the seismic hazard model currently in active development for the Groningen gas fields (NAM, 2014; Bourne et al., 2014) is given.

2.4.1 The Groningen Gas Field

The Groningen field is a large natural gas field located in the northern Netherlands (Figure 2-5). Originally discovered in 1959, the area is one of the most extensive, and economic, areas of natural gas extraction in the world. It is the largest in Europe, and the tenth largest globally, contributing to approximately half of the natural gas production in the Netherlands. Located in an area of very low tectonic seismicity, gas extraction in the region has lead to an increase in seismicity in Groningen and the surrounding regions since the early 1990s. A total of 187 events with $M_L \geq 1.5$ have been recorded in the gas field between April 1995 and October 2012. The largest of these events is the M_L 3.6 Huizinge earthquake, in August 2012. This earthquake caused damage to buildings and raised concerns over public safety. Following the Huizinge earthquake, extensive investigation into the seismic hazard associated with the gas extraction has been undertaken, forming one of the most comprehensive induced seismicity PSHA models produced to date.

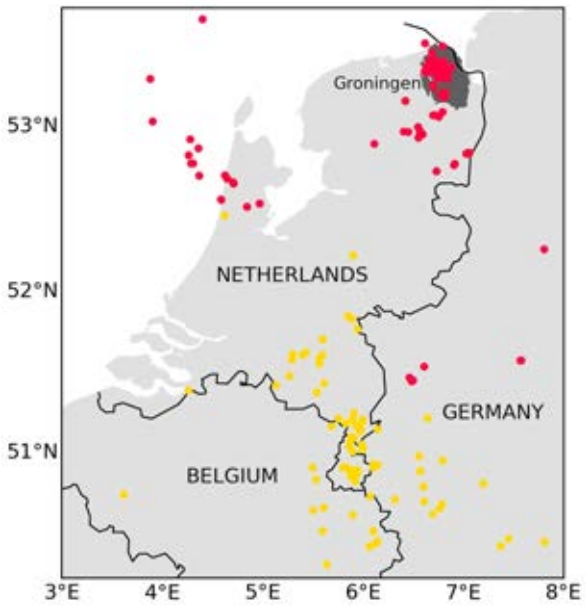

Figure 2-5 Location of the Groningen gas field in the northern Netherlands with natural seismicity shown in yellow, induced seismicity in red (reproduced from Bourne et al. 2014)

Given the occurrence of damaging seismicity from gas extraction in the Netherlands, and the subsequent compensation to residents affected by the earthquake damage, future production strategies require that the potential risk to the population be taken into consideration within the economic cost-benefit. At the time of writing, work is ongoing to continue the assessment of seismic hazard and risk for the Groningen region, potentially forming one of the first implementations of induced seismicity.

The creation of the time-varying seismogenic source model from the predicted geomechanical response from future depletion strategies of the gas field is described in Bourne et al. (2014). From the dense network of GPS sensors deployed across the gas field, the geodetic deformation can be well measured. Starting from the model of Kostrov (1974) the average irrotational seismic strain due to seismicity within in a given volume is proportional to the sum of their moment tensors:

$$\overline{\varepsilon}_{s,ij}\left(t\right)=\frac{1}{2\mu V}\sum_{k=1}^{N\left(t\right)}M_o^k m_{ij}^k \tag{B.10}$$

where $\overline{\varepsilon}_{ij}$ is the *ij*th component of the average seismic strain tensor, N the number of events within the given volume *V* and time interval *t*, μ the shear modulus, and M_o^k and m_{ij}^k the scalar seismic moment and symmetric moment tensor of the kth event. The average incremental strain within a given volume at position ***x*** during the time *t* can is given by:

$$\overline{\varepsilon}_{ij}\left(t\right)=\frac{1}{V}\int_V \varepsilon\left(\mathbf{x},t\right)dV \tag{B.11}$$

For a subsurface reservoir it may be convenient to assume that the lateral extents are significantly larger than the vertical extent, such that the reservoir has the geometry of a thin sheet with lateral incompressibility. Therefore all of the components of the strain tensor can be considered approximately equal to 0, with the exception of the vertical strain ε_{zz}. This simplifies the volume integral in (B.11) to a surface integral over the area of the reservoir, *S*:

$$\overline{\varepsilon}_{zz}\left(t\right)=\frac{1}{V}\int_S \Delta h\left(\mathbf{x},t\right)dS \tag{B.12}$$

where $\Delta h=\int_{z_1}^{z_2}\varepsilon_{zz}\left(\mathbf{x},t\right)dz$ is the change in reservoir thickness. The location of observed seismicity in the field with respect to the volume change of the reservoir for the period 1960 – 2012 is shown in Figure 2-6.

In reality not all of the induced strain will be accommodated by seismogenic slip on faults. A proportion may be considered as elastic train or aseismic creep. The proportion of seismogenic strain with respect to the total strain is given by the partition factor α. The total seismic moment within the deforming volume in the time T, therefore, is given by:

$$M_{o,T}\left(t\right)\simeq\frac{4\mu}{3}\int_S \alpha\left|\Delta h\right|dS \tag{B.13}$$

From this formulation it is evident that the strain-partitioning factor is a crucial parameter to be constrained. Further, α cannot necessarily be assumed constant, but may depend on the cumulative induced deformation, which may itself vary with the proposed depletion scenario. From observations of strain an seismic moment within the reservoir, α an empirical model relating α to thickness change Δh has been proposed:

$$\alpha\left(\Delta h\right)=\frac{e^{f+g\Delta h}}{1-e^{f+g\Delta h}} \tag{B.14}$$

Given that the total seismic moment within the reservoir is uncertain, Bourne et al. (2014) model it with a Pareto distribution, such that the probability density function is given by:

$$\Phi\left(M_o\right)=\frac{\beta}{M_{o,\min}}\left(\frac{M_o}{M_{o,\min}}\right)^{-1-\beta} \tag{B.15}$$

where $M_{o,min}$ is the moment of the minimum bound of the recorded earthquakes, and $\beta = b/1.5$.

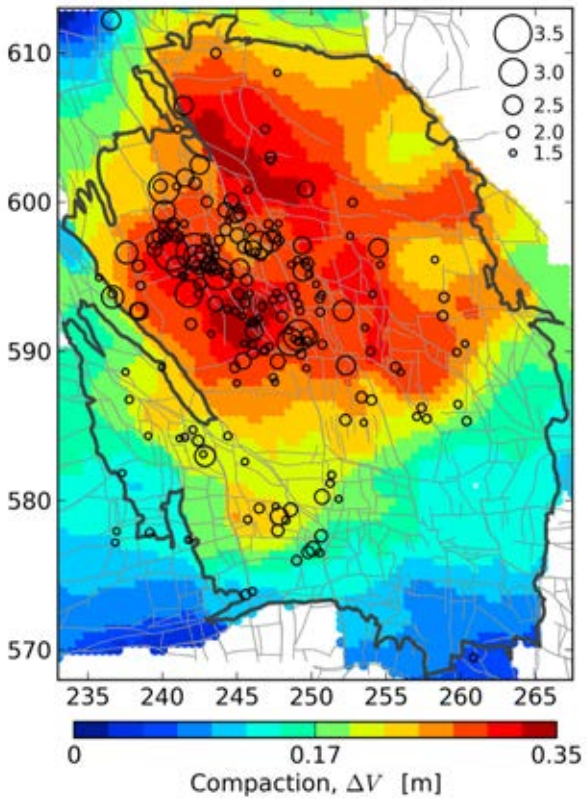

Figure 2-6 Compaction of the Groningen gas reservoir from 1960 – 2012, with induced earthquake events (black circles) recorded during the period 1995 – 2012 (Reproduced from Bourne et al., 2014)

Finally, the last element of the model to be constrained is the spatial density of the events with respect to the spatial distribution of the volume change within the reservoir. For this, Bourne et al. (2014) propose that the median number of events with $M_o \geq M_{o,min}$ per unit area is determined via:

$$N_d(\mathbf{x},t) \simeq \frac{1}{2}\left(\frac{4\mu\alpha(\Delta h)|\Delta h|}{3C_\beta M_{o,\min}}\right)^\beta \qquad \text{(B.16)}$$

where $C_\beta = \left[\Gamma(1-\beta)\cos(\pi\beta/2)\right]^{\frac{1}{\beta}}$ and $\Gamma(\cdot)$ is the Gamma function.

This theoretical formulation allows for the creation of a process by which the volume change in response to different depletion scenarios, as inferred from geomechanical and geodetic models, can be used to generate simulations of seismicity. An example of the evolution of the density of events over a time for a theoretical gas production scenario is shown in Figure 2-7.

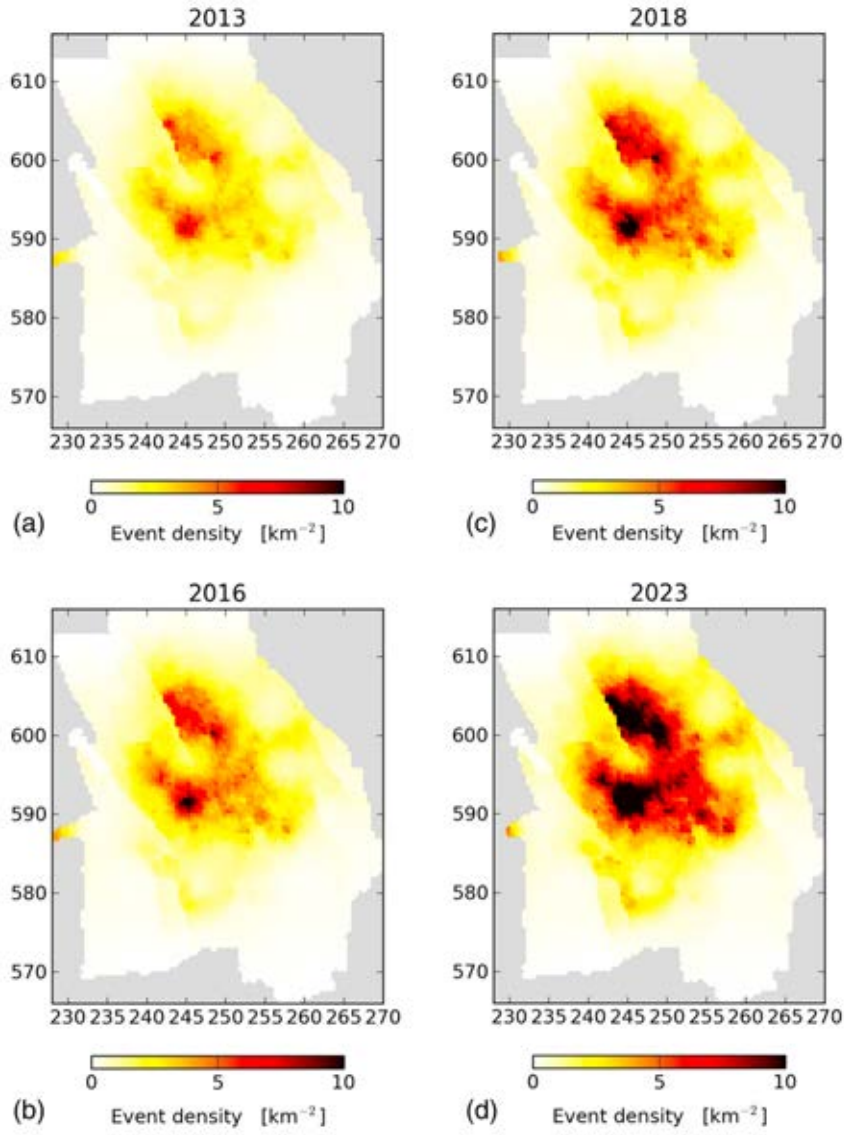

Figure 2-7 Predicted event density from the start of gas production up to 2013, 2018, 2016 and 2023 based on one of the proposed seismological models of the Groningen field and the 2013 gas production plan (Reproduced from Bourne et al., 2014)

Bourne et al., (2014) outline the process for generating an earthquake catalogue from the model of volumetric change in the following steps:

1. Sample the total seismic moment distribution from the probability density function given in (B.15)
2. Sample an event location within the region with weights according to the normalised density distribution in (B.16), i.e. $\Delta N_d(\mathbf{x}) \Big/ \int_S \Delta N_d(\mathbf{x}) dS$
3. Sample the magnitude from the expected magnitude-frequency distribution, which in many cases can be the standard Gutenberg & Richter model, truncated by an upper bound M_{Max}. Initially M_{Max} may be taken as the magnitude equivalent to the total moment distribution, which must be subsequently reduced by the moment of each event after each event it sampled.

4. Repeat steps 2 and 3 until the total seismic moment of the catalogue is approximately equal to that of the total seismic moment sampled in step 1.

As in the case of the geomechanical seed models, this process is deriving stochastic event sets whose properties are consistent with the underlying geomechanical model. Multiple realisations of seismicity can be determined for any particular model of pressure depletion, incorporating the time-dependent properties of the process. In both the geomechanical models and the geomechanical seed models the tectonic background processes are neglected. In case of enhanced geothermal systems this may be because the dynamic response of seismicity occurs over a significantly shorter time-scale than that of the tectonic processed. In the Groningen case it is simply because the background tectonic seismicity is sufficiently small that the total contribution can be considered negligible. In both cases it is possible to put tectonic seismicity back into the model, permitting that in a small number of realisations a tectonic event could be inserted into the synthetic catalogues, although the influence of doing to may be minimal.

To complete the PSHA process in the case of Groningen, each earthquake within the synthetic catalogue is then used as the point source for the generation of strong ground motion using a locally calibrated GMPE. Using the Monte Carlo PSHA process the number of exceedances of a given hazard level from multiple realisations of synthetic seismicity are then used to generate the hazard curves directly. The resulting hazard model is described in more detail in Bourne et al (2015) and NAM (2013). Of particular relevance in this example, however, is the use of a logic tree to explore the epistemic uncertainties on the key model and model parameter that control the hazard process. These were identified as being: i) the compaction model (i.e. the model to predict deformation given a pressure depletion within the gas field), ii) the strain partitioning factor α, iii) the b-value of the region, and iv) the GMPE. In the final logic tree the b-value model is believed to be dependent on the compaction within the reservoir, and hence three alternative models are suggested, one (the central branch) in which b-value is held fixed. The resulting epistemic uncertainty analysis shows that the parameter with the greatest contribution to the overall epistemic uncertainty is the strain partitioning factor α, indicating that a significant effort may be needed to better constrain this part of the model.

2.5 SUMMARY OF SEISMIC SOURCE CHARACTERISATION AND HAZARD CALCULATION PROCESSES FOR INDUCED SEISMICITY

The overview of the four types of PSHA process outlined in sections 2.1 to 2.4 highlight some of the different approaches taken. It is interesting to note, that in the examples shown we see two contrasting approaches for each of the two induced seismicity cases: time-independent PSHA versus full geomechanical modelling in the case of the natural gas extraction example, and time-dependent empirical modelling verses the geomechanical seed approach in the case of the Basel enhanced geothermal system. This may highlight the potential advantages and disadvantages of each approach when considering their use in other applications. Both the Groningen gas field and Basel EGS are areas that are exceptionally well monitored, which considerable amounts of seismological and geophysical data available. In the case of Basel, the models are being applied *a posteriori*, which means that irrespective of the strength of their physical basis, these approaches have not

necessarily been tested in real-time then the constrain on some of the critical parameters is not so great.

With the exception of the time-independent PSHA model, which operates in a manner that is common to any such PSHA application, one of the common factors is the use of synthetic catalogues as the common means of characterising the seismogenic source. In this regard the probability definitions are not always directly comparable with those of conventional time-independent or time-dependent PSHA, as we are more often considering the absolute probability of exceeding a given level of ground motion within the time period. Instead, for the case of the geomechanical-based models, we are only considering the probability of exceeding a given level of ground motion from exclusively the events triggered by the anthropogenic process. In these cases the dynamic nature of the induced seismicity forcing can be taken into account and instead of producing a probability of exceedance in T years, the result is a probability of exceedance with respect to the forcing sequence. Neither approach, however, excludes the possibility that the induced seismicity catalogue could be integrated with a similar realisation of tectonic seismicity in areas, or under conditions, where the latter could be considered to give a non-trivial contribution to the overall hazard.

2.6 GROUND MOTION PREDICTION IN INDUCED SEISMICITY APPLICATIONS

Until recently the study of the seismic ground motions arising specifically from induced seismicity events, and their variability in comparison to tectonic events of a similar size, has not been well understood. This has, in part, been driven by a relative paucity of strong motion data being made available publically from various induced seismicity projects. Application of existing GMPEs derived from strong motion records of tectonic events can be problematic in these environments for several reasons. With only a small number of exceptions, induced seismicity events generally fall below the range of magnitudes considered in the existing databases used for developing GMPEs. The typical magnitude range required for induced seismicity studies may span M_W 1.0 to M_W 5.5 – 6.5, whilst few GMPEs are well constrained for use with earthquakes fall below M_W 4.5 – 5.0. Studies such as Bommer et al., (2007) and Chiou et al., (2010) demonstrate that extrapolation of conventional tectonic GMPEs to such small magnitudes tends to overestimate the expected ground motions (see also Mignan et al., 2015). They may also fail to adequately capture local attenuation affects that may be more apparent for small magnitudes than for the range typically considered for GMPEs. Other forms of potential bias in induced seismicity ground motions with respect to tectonic GMPEs may be due to shallow hypocentral depths, high heat flow and potential variations in stress drop.

Given the limitations of using many existing GMPEs as published, what options are available in to the hazard modeller, and what should be the criteria for selection? The latter question is addressed by NAM (2013) who, for application to the Groningen gas field, outline the main requires, which we summarise here:

1. *Compatible equations to predict both PGA and PGV*. Given the small magnitude of the induced events and their relatively shallow depth, the most adverse types of ground motions may be those with relatively high accelerations for high frequencies, usually at short durations. Such motions may be damaging to acceleration sensitive system elements (such as electrical transformers), non-structural building elements

(e.g., chimneys, balconies, balustrades etc.), or in the worst case to low-rise masonry structures themselves. It is therefore expected that the preferred GMPE should be able to constrain PGA, PGV and spectral accelerations in the high frequency range.

2. *Applicability of the magnitude range under consideration (M_W 1.5 to M_W 6.5)*.
3. *Individual earthquakes in the seismicity model are represented as points (hypocentres), therefore it is important to use a GMPE based on either hypocentral distance or epicentral distance, as opposed to those based on finite fault rupture*. It is shown by Bommer & Akkar (2012) that employing GMPEs with finite fault requirements to point sources can underestimate the hazard contribution from large earthquakes.
4. *The GMPE should be calibrated to the appropriate soil sites, or else include a site term such as V_{S30} as an explicit predictor variable.*
5. *The GMPE must report the standard deviation to characterise the logarithmic distribution of ground motion residuals, as well as the decomposition of the standard deviation into its inter-event and intra-event components.*

Given these requirements, and perhaps not neglecting the GMPE selection criteria proposed by Cotton et al. (2006), the modeller has several options. The first is to adopt an existing GMPE, either one derived from induced seismicity recordings or one for small magnitude tectonic events. If strong motion records are already available these can be used to compare existing GMPEs, or if necessary to re-calibrate them to take into account the local conditions (e.g. magnitude, hypocentral depth, site amplification etc.). Of these the hypocentral depth distribution may require some additional care, as GMPEs derived from small tectonic events may display a different hypocentral depth distribution than that of induced seismicity. Alternatively the modeller may wish to make use of stochastic ground motion simulation methods to provide simulated records with properties calibrated to the local conditions of the region (e.g. stress drop, kappa). These too could be supplemented with local records if available.

The following are potential GMPEs for use in induced seismicity applications.

2.6.1 Dost et al. (2004)

One of the first GMPEs derived specifically from induced seismicity recordings was that of Dost et al. (2004). This uses a set of records measure from induced seismicity events near the Roswinkel gas field in the Netherlands

$$\log_{10}\left(PGV\right) = -1.53 + 0.74M_L - 0.00139R_{HYP} - 1.33\log_{10}\left(R_{HYP}\right) + \varepsilon\sigma \quad \text{(B.17)}$$

$$\log_{10}\left(PGA\right) = -1.41 + 0.57M_L - 0.00139R_{HYP} - 1.33\log_{10}\left(R_{HYP}\right) + \varepsilon\sigma \quad \text{(B.18)}$$

where PGV and PGA are the geometric mean horizontal velocity and acceleration, given in units of cm/s and m/s^2 respectively. σ is fixed at 0.33 in base 10 units, 0.76 in natural log units, for both models. The strong motion records span the magnitude range 2.3 ≤ M_L ≤ 3.9, whilst hypocentral distance ranges from approximately 2.0 km to 25 km. The hypocentral depth range of the local events is also limited to a very narrow range of values. From the induced events the depths were mostly located at 2 km, with an error of ± 0.2 km.

The whilst the model of Dost et al. (2004) may have been derived for small magnitude events, due to the linear scaling of the ground motion with magnitude it does not capture the effect of saturation of the ground motion if extrapolated to larger magnitudes. A modification of this model is proposed by Bommer (2013) to attempt to address this shortcoming [reported in Dost et al., (2013)]:

$$\log_{10}\left(PGV\right)=-1.3972+0.7105M_W-0.0829\left(M_W-4.5\right)^2 \\ -0.00139R_{HYP}-1.33\log_{10}\left(R_{HYP}\right)+\varepsilon\sigma \quad \text{(B.19)}$$

$$\log_{10}\left(PGA\right)=-1.609+0.614M_W-0.1116\left(M_W-4.5\right)^2 \\ -0.00139R_{HYP}-1.33\log_{10}\left(R_{HYP}\right)+\varepsilon\sigma \quad \text{(B.20)}$$

Bommer (2013) also proposes separating the total variability into an inter-event (τ) and intra-event (ϕ) component using an approximate ratio $\phi/\tau\approx 2$. The resulting intra-event and inter-event standard deviations are 0.2952 and 0.1476 respectively.

2.6.2 Convertito et al. (2012)

The second GMPE is developed by Convertito et al. (2012) from strong motion records deployed around the Geysers Geothermal field in California. The model is fit to a database of strong motion records from 220 earthquakes, recorded at 29 stations within the field. The magnitude range is 1.0 ≤ M_W < 3.5, and the hypocentral distance range approximately 0.5 km to 20 km.

$$\log_{10}\left(PGA\right)=-2.268+1.276M_W-3.528\log_{10}\sqrt{R_{HYP}^2+3.5^2}+0.053R_{HYP}+0.324S+\varepsilon\sigma \quad \text{(B.21)}$$

where PGA is the larger of the two horizontal components (in m/s^2), a S is a site term taking the value of 1 for soil, 0 otherwise. σ is found to equal 0.324. The hypocentral depth range of the events is considered by Convertito et al. (2012) is not stated explicitly, although the maximum depth value is 6 km. As shown in the model, the effective depth (h_{eff}) is fixed at 3.5 km, indicating persistently shallow earthquakes.

2.6.3 Douglas et al. (2013)

The set of GMPEs presented by Douglas et al. (2013) represent the most extensive suite of induced-seismicity specific ground motion models. The authors of the study present two types of GMPE. The first one based on a collection of induced seismicity records taken from small magnitude shallow earthquakes including several sets due to induced seismicity sequences (e.g. Basel, Geysers, Roswinkel, Soultz-sous-Forêts). The second is based on strong motion records generated stochastically, using the SMSIM software (Boore, 2003), and based on an appropriate seismological model.

For application, the empirical GMPE is strongly limited by the paucity of records with M_W ≥ 3.0. Instead, it is the GMPE derived from stochastic records (hereafter referred to as the "stochastic" GMPE) that is preferred. From the limits of the strong motion simulations, the stochastic GMPE is considered applicable to magnitudes in the range 1.0 ≤ M_W ≤ 5.0, and for hypocentral distances up to 40 km. However, as the strong motion records are based on

a seismological model, Douglas et al. (2013) consider the epistemic uncertainties on three critical parameters: source stress drop $\Delta\sigma$, attenuation quality factor *Q* and the site/station response parameter κ. Therefore, multiple set of coefficients (36 in total) are derived, exhausting the possible combinations of values for $\Delta\sigma$ (1, 10 and 100 bar), *Q* (200, 600, 1800) and κ (0.005, 0.02, 0.04, 0.06 s). The each coefficient set defines the coefficients for PGA, PGV and Sa(T) in the range 0.01 to 10.0 s. The Douglas et al. (2013) stochastic GMPE assumed the following form:

$$\begin{aligned}\ln\left(Y\right) = b_1 + b_2\left(M_W - 3\right) + b_3\left(M_W - 3\right)^2 + b_4\left(M_w - 3\right)^3 \\ + b_5 \ln\left(R_{HYP} + b_h\right) + b_6\left(R_{HYP} + b_h\right)\end{aligned} \tag{B.22}$$

where Y is the ground motion in m/s^2 (for PGA and Sa) or m/s (for PGV). The hypocentral depth is not considered explicitly within the stochastic model, as simulations are based on R_{HYP} exlusively. Comparisons with the recorded ground motions from geothermal earthquake sequences, whose hypocentral depths are typically limited to 1 – 4 km, would not suggest any clear bias with depth. Thus suggesting that the stochastic model would be suitable for earthquakes within the shallowest few kilometres of the crust. One caveat, however, may be that some observations of shallow earthquakes indicate a significant change in stress-drop with hypocentral depth in the very shallow crust (e.g. Goertz-Allmann & Wiemer, 2013; Atkinson, 2015). Applications of the Douglas et al. (2013) stochastic GMPE in practice may need to take this trend into consideration. A sensitivity analysis of the Douglas et al. (2013) GMPE models is presented by Mignan for the Groningen field in STREST report 3.1.

The aleatory uncertainty model for the stochastic GMPE is intended to represent as close as possible a site-specific and region-specific standard deviation. To accomplish this they divide the total standard deviation into a single-station intra-event term (ϕ_{SS}) and a region-specific inter-event term. The region specific inter-event term (τ) is taken as the mean value from the observed inter-event variability from the Basel geothermal sequence (τ_{BASEL}) and the Soultz-sous-Forêts sequence (τ_{SOULTZ}). Therefore the total aleatory variability term of this GMPE is determined via:

$$\sigma_{TOTAL} = \sqrt{\phi_{SS}^2 + \left(\frac{\left(\tau_{SOULTZ} + \tau_{BASEL}\right)}{2}\right)^2} \tag{B.23}$$

2.6.4 Atkinson (2015)

The GMPE of Atkinson (2015) differs from the other GMPEs presented thus far as it is not fit specifically to observed (or even simulated) induced seismicity recordings. Instead it is fit to a subset of the NGA-West 2 database (Ancheta et al. 2014) that contains records most representative of the magnitude and distance range commonly considered for induced seismicity applications. The applicable magnitude range is considered to be approximately $3.0 \leq M_W < 6.0$, and the hypocentral distance range < 40 km. The GMPE provides

coefficients for orientation-independent component PGA, PGV and pseudo-spectral acceleration at periods ranging from 0.03 s to 5.0 s. The general form of the model:

$$\ln(Y) = c_0 + c_1 M_W + c_2 M_W^2 + c_3 \log\left(\sqrt{R_{HYP}^2 + h_{eff}^2}\right) + c_4 \sqrt{R_{HYP}^2 + h_{eff}^2} \quad \text{(B.24)}$$

where $h_{eff} = \max\left(1, 10^{(-1.72+0.43M_W)}\right)$ is the effective depth parameter to constrain near-source behaviour. Little information is given regarding the hypocentral depth range of the records. However, as the records are largely from tectonic sources their distribution of hypocentral depths may cover a greater depth range than is usually considered by induced seismicity sources.

Atkinson (2015) notes that the anelastic attenuation term should have an insignificant effect of amplitudes within the 40 km distance range considered for the GMPE, but is included to ensure an appropriate curvature should the GMPE be extrapolated to longer distances. For periods greater than 1.0, c_4 is taken to be equal to 0.0 and that the parameter should decrease linearly with log frequency to reach the value of -0.002 for periods less than or equal to 0.1 s (and PGA). It is assumed to be -0.0006 for PGV.

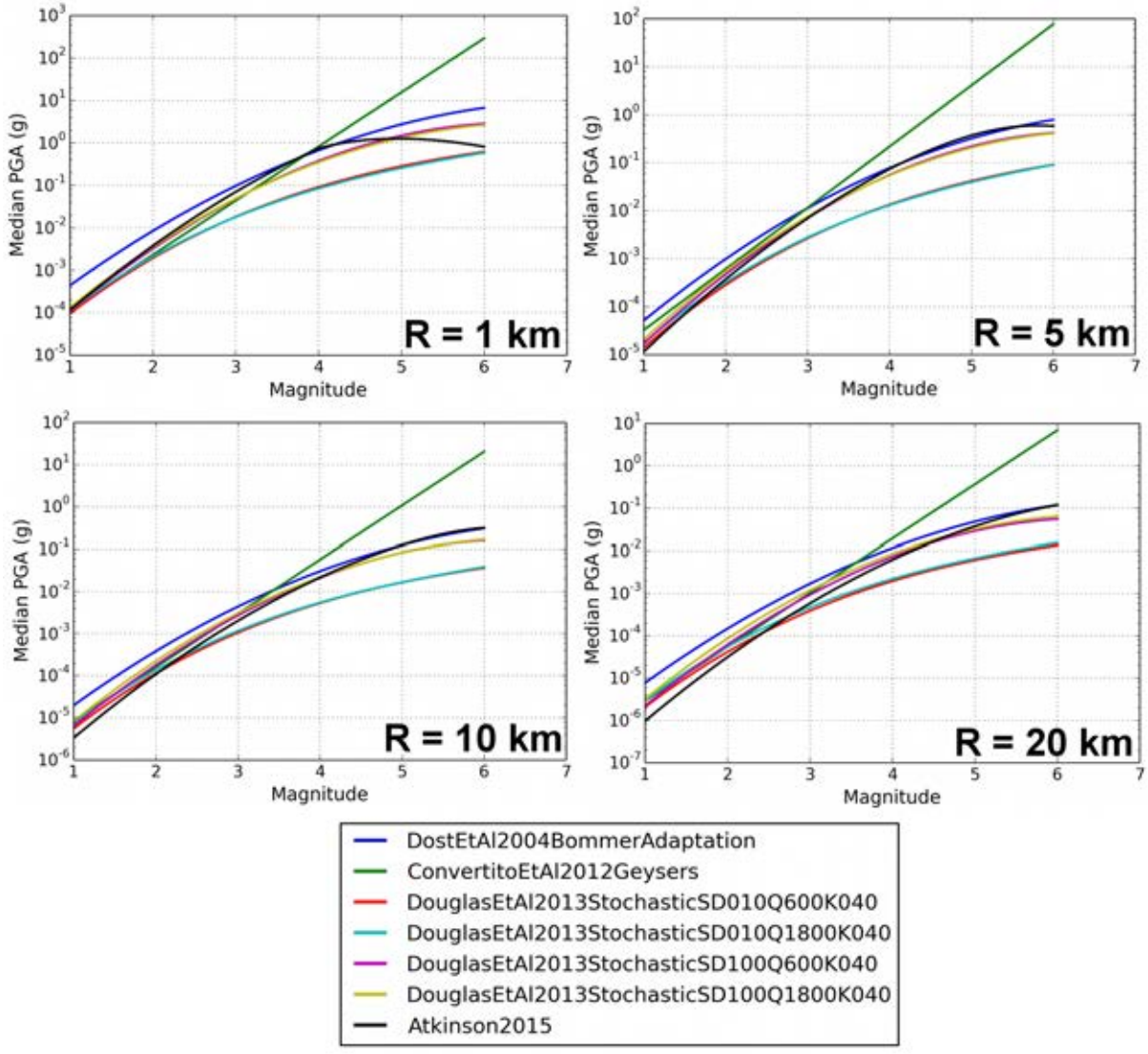

Figure 2-8 Scaling of the Peak Ground Acceleration (PGA, g) with magnitude for seven induced seismicity GMPEs, at four hypocentral distances (1.0 km, 5.0 km, 10.0 km and 20.0 km).

2.6.5 Induced Seismicity GMPE Comparisons

To understand the potential limitations of the GMPEs we perform a comparison of the four models. The GMPE's themselves have been implemented in the OpenQuake software (see the next chapter for the overview, and section 4.1 for implementation details), which provides the possibility for direct comparison of the models. To undertake the comparisons we use the OpenQuake Ground Motion Toolkit (Weatherill et al., 2014), a suite of tools for visualising GMPEs using the OpenQuake engine and, if records are available, comparing the GMPEs directly with the strong motion records. The trellis plots in Figure 2-8 to Figure 2-11 compare the magnitude and the distance scaling of PGA and PGV for the following models: Dost et al., (2004) (with the magnitude correction), Convertito et al., (2012), Douglas et al., (2013) and Atkinson (2015). From the Douglas et al. (2013) GMPE it is too cumbersome to compare all of the 36 combinations; therefore the analysis is limited to just four of the models based on the analysis by Edwards & Douglas (2013): i) $\Delta\sigma$ = 10 bar, Q = 600 and κ = 0.04 s, ii) $\Delta\sigma$ = 10 bar, Q = 1800 and κ = 0.04 s, iii) $\Delta\sigma$ = 100 bar, Q = 600 and κ = 0.04 s and iv) $\Delta\sigma$ = 100 bar, Q = 600 and κ = 0.04 s. For the sake of comparison M_L is assumed equivalent to M_W for the Dost et al. (2004) GMPE.

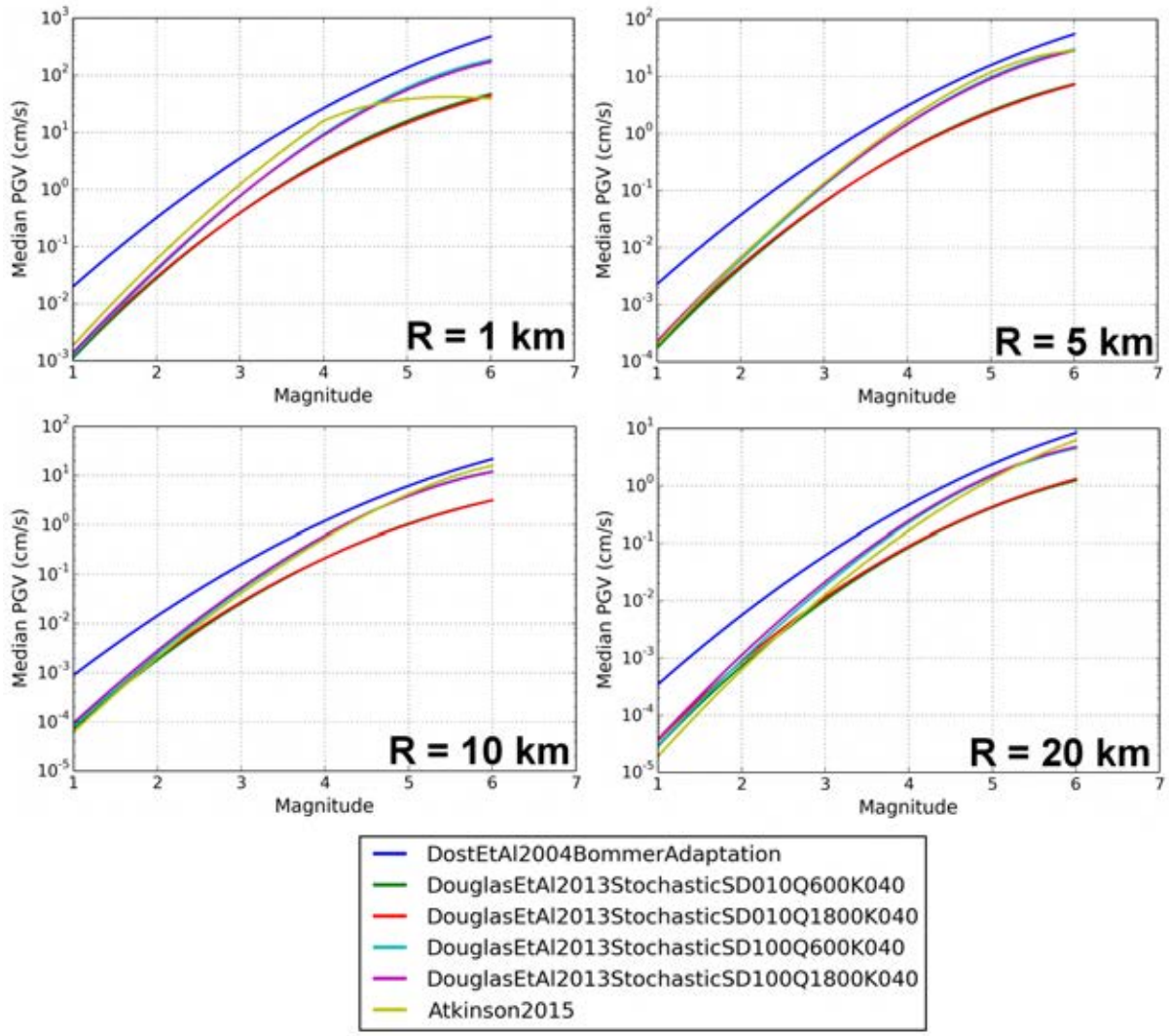

Figure 2-9. As Figure 2-8 for Peak Ground Velocity (PGV, cm/s)

From the comparison of the magnitude scaling of the GMPEs for PGA (Figure 2-8) and PGV (Figure 2-9) there are some interesting trends that may require careful interpretation. For the lower-magnitude range $1.0 \leq M_W \leq 3.0$ there is reasonable convergence between the GMPEs. For larger magnitudes, however, there is greater divergence. Naturally due to absence of a magnitude saturation term for Convertito et al. (2012), that particular GMPE produces ground motion values well in excess of those of the remainder of the set for magnitudes larger than M_W 4.0. For hypocentral distances greater than approximately 5 km, Atkinson (2015) is in very close agreement with Douglas et al., (2013) models for which $\Delta\sigma$ = 100 bar. As Atkinson (2015) is derived from the NGA-West 2 database, which contains a large number of records from California, this is consistent with an $\Delta\sigma$ estimate of 100 bars for the Western United States (Campbell, 2003).

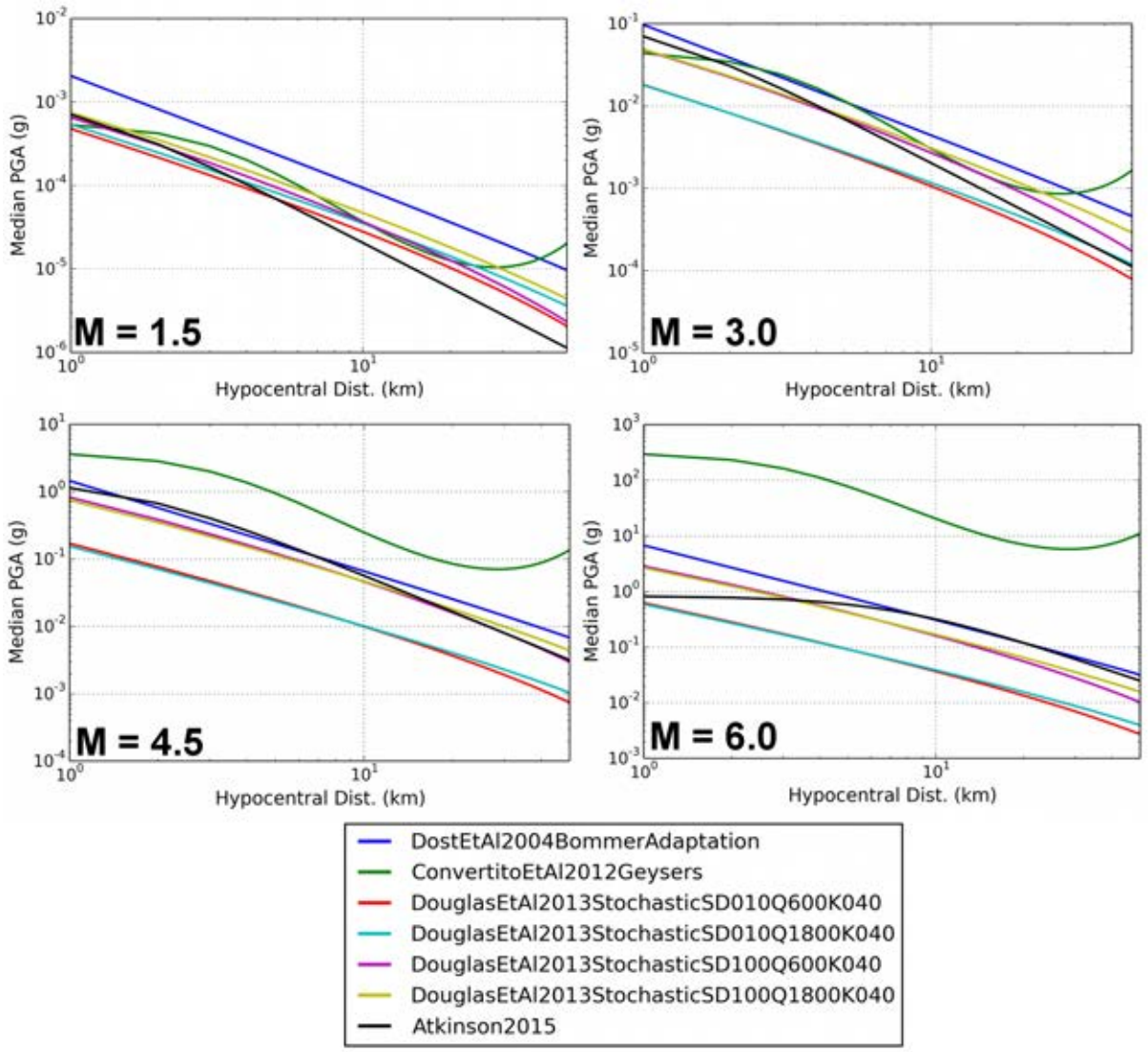

Figure 2-10 Scaling of the Peak Ground Acceleration (PGA, g) with distance for seven induced seismicity GMPEs, at four hypocentral magnitudes (M = 1.5, 3.0, 4.5 and 6.0

Figure 2-10 and Figure 2-11 compare the attenuation of the GMPEs with distance, for PGA and PGV respectively. For low magnitudes we see better agreement between all of the GMPEs, within a hypocentral distance of approximately 20 km. For larger magnitudes there is considerably greater divergence. For Convertito et al. (2012) the high accelerations are cause by the magnitude scaling term. However, for this specific GMPE a new problem emerges at longer hypocentral distances. Here we see a reversal in the attenuation trend with higher accelerations predicted at greater distances. The reason for this can be seen in the positive coefficient for the anelastic attenuation term. Such behaviour is clearly unphysical, and this should be taken into consideration in application of the GMPE to PSHA, as this would prevent extrapolation of the GMPE beyond the specified distance range.

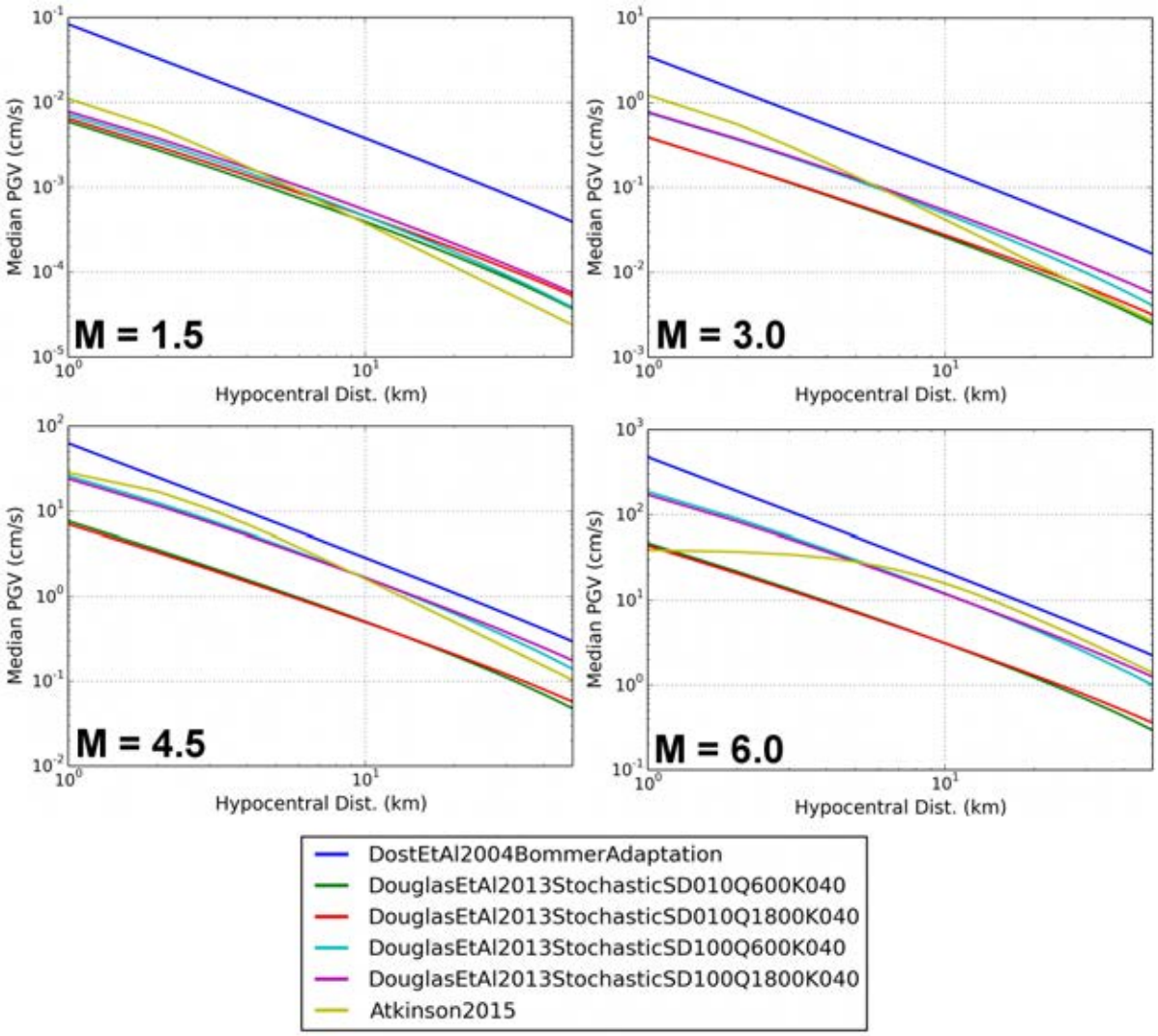

Figure 2-11 As Figure 2-10 for PGV (cm/s)

Also of note in the comparisons is the fact that the Dost et al. (2004) GMPE is persistently higher than the rest, particularly for small magnitude earthquakes. There are several factors that might explain this trend. The first is due to the comparatively shallow hypocentral depth distribution of the events considered, which is close to 2 km with very little variability. For other GMPEs the depth distribution of the events may be closer to 3 – 4 km or deeper. Another factor may be higher regional stress drop, which would typify an extremely stable seismic region such as the northern Netherlands.

The comparison of the different source modelling approaches and induced seismicity GMPEs provides some context of for defining the requirements of an induced seismicity PSHA. The next step, therefore, is to understand how the more conventional PSHA process can be adapted for such applications. From the studies considered here it seems that there are some fundamental requirements of probabilistic seismic hazard software that would facilitate its use in an induced seismicity application. The first is the capability to estimate seismic hazard using simulated earthquake catalogues, either in place of or in addition to the classical PSHA approach. Further requirements are the capability to model the source in a manner that is consistent with the definitions encountered here. This includes small

magnitude, at close distances, with temporal recurrence properties that can be readily adapted for a given context. Finally the need to consider epistemic uncertainty in both the source model and ground motion model is also a necessity. As will be seen in the next chapter, the OpenQuake-engine is one such software that can accommodate these requirements, and can do so in a way that if sufficiently flexible for applications such as these.

3 OpenQuake – Outline and Operation

The OpenQuake-engine is an open-source Python-based software for the calculation of probabilistic seismic hazard and risk, created by the Global Earthquake Model (Pagani et al., 2014). It is designed to undertake state-of-the-art calculations of seismic hazard and risk in a manner that is entirely flexible for application across the globe. The software has been designed to support the characterisation of seismogenic sources for a wide variety of source typologies, and contains an extensive and growing library of ground motion prediction equations for a wide variety of tectonic environments. In addition to its range of functionality, which shall be expanded upon in due course, the manner in which the software is developed makes it a suitable candidate for customisation in a context such as induced seismicity. OpenQuake is constructed using a test-driven development approach, which ensures that a transparent and documented quality assurance process is built into the development process. For each function within this software an accompanying test suite is developed. This additional code tests the function under a set of, ideally, exhaustive conditions to ensure that each unit of the code operates in the expected manner. Both the source code and its corresponding test suite can be explored in its public open repository (http://github.com.gem). OpenQuake is distributed under a GNU Affero General Public License 3.0.

The OpenQuake-engine is divided into three parts:

i) the OpenQuake hazard library (oq-hazardlib), which contains the core functionality for calculation of the seismic hazard

ii) the OpenQuake risk library (oq-risklib), which contains the code for the calculation of seismic risk as well as the “OpenQuake common library” (oq-commonlib), which contains the code used to glue the operations of the hazard and risk calculations together (including input and output).

iii) The OpenQuake engine (oq-engine) is the main OpenQuake program that is responsible for the management of calculation tasks in a distributed computing environment.

Of the three modules, both the *oq-hazardlib* and *oq-risklib* can operate as standalone Python libraries. This permits the possibility that the core scientific functionalities of OpenQuake can be extracted from the OpenQuake-engine and adapted into other codes. Examples of this can be seen in GEM’s Hazard Modeller’s Toolkit (Weatherill et al., 2014a) and Ground Motion Toolkit (Weatherill et al. 2014b). It is this usage of OpenQuake’s hazard and risk libraries that will form the basis for the application to induced seismicity herein.

3.1 THE CORE CALCULATION KERNELS

OpenQuake contains four primary kernels for performing different types of seismic hazard calculation:

i) “Classical PSHA” – the primary kernel for the calculation of probabilistic seismic hazard following the approach of Cornell (1968) and McGuire (1976)

ii) "Event-Based PSHA" – an adaptation of the PSHA approach using Monte Carlo sampling of the source and ground motion models to determine the probability of exceeding a given level of ground motion at a site.

iii) "Disaggregation" – the calculator to perform the disaggregation of the seismic hazard model for a given ground motion intensity measure at a specific probability of being exceeded, adopting the formulation proposed by Bazzurro and Cornell (1999)

iv) "Scenario" – the calculator can produce multiple realisations of ground motion fields arising from a given earthquake scenario (source and magnitude).

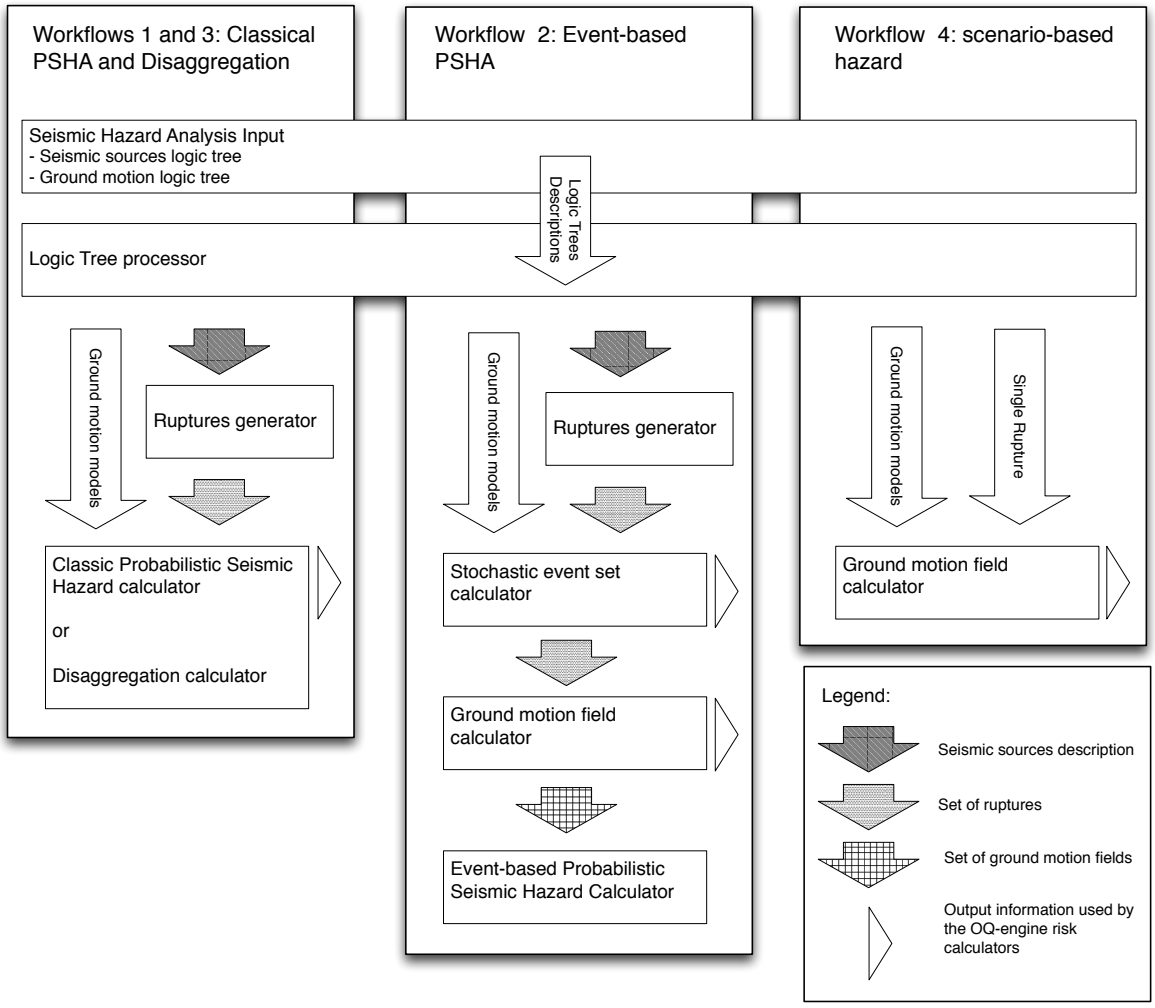

Figure 3-1 Workflow of the main OpenQuake calculation modules (reproduced from Pagani et al., 2014a)

The mathematical formulation of the hazard calculation engine is described in detail in Pagani et al. (2014a,b), but for completeness we shall summarise it here. Adopting the approach first presented by Field et al. (2003), OpenQuake makes two core assumptions that form the basis of the hazard calculations:

i) The seismicity in a region is described by a collection of *independent seismic sources* such that the occurrence of an earthquake within a specific source does not alter the probability of occurrence of earthquake ruptures in other sources

ii) Each rupture generated within the source is independent of other ruptures generated by the same source.

It will be seen in due course that it can remain possible to proceed with a PSHA calculation using OpenQuake if one or both of the assumptions i) and ii) are not true, albeit this requires customisation of the OpenQuake-hazardlib beyond its conventional approach. From the above assumption, each of the four main calculation workflows can be broken down into two components: the seismic source model (SSM) and the ground motion model (GMM). The seismic source model is defined as a set of ***I*** seismogenic sources:

$$SSM = \left\{ SRC_1, SRC_{2,} SRC_3, \ldots, SRC_I \right\} \tag{B.25}$$

Each seismogenic source (SRC_I) then can then generate a set of individual ruptures:

$$SRC_i = \left\{ Rup_{i1}, Rup_{i2,} \ldots, Rup_{ij} \right\} \tag{B.26}$$

3.1.1 Classical Probabilistic Seismic Hazard

The main output of a probabilistic seismic hazard analysis is the hazard curve. For a set of ground motion intensity levels the hazard curve indicates for each level the probability that the given level of ground motion will be exceeded at least once in the specified timespan ***T*** years $P\left(X \geq x \,|\, T\right)$, and consequently the probability that the ground motion level will not be exceeded in ***T*** years is taken as $P\left(X < x \,|\, T\right) = 1 - P\left(X \geq x \,|\, T\right)$

From the assumption of independent sources, the probability that none of the sources results in a ground motion value exceeded a given level is taken as the product of the probabilities that each source does not cause an exceedance:

$$P\left(X \geq x \,|\, T\right) = 1 - \prod_{i=1}^{I} P_{SRC_i}\left(X < x \,|\, T\right) \tag{B.27}$$

If it is assumed that each source generates independent earthquake ruptures then the probability that each source does not cause an exceedance of the ground motion level, $P_{SRC_i}\left(X < x \,|\, T\right)$, is simply the product of the probabilities that each individual rupture generated by the source does not cause an exceedance:

$$P_{SRC_i}\left(X < x \,|\, T\right) = \prod_{j=1}^{J_i} P_{rup_{ij}}\left(X < x \,|\, T\right) \tag{B.28}$$

If a rupture does not cause any exceedance of the specified ground motion level in the time span T, this may be due to the fact that the rupture has not occurred at all in the time span,

or that it has occurred one or more times without causing an exceedance of the ground motion level. As these conditions are mutually exclusive, from the total probability theorem the hazard level can be written as:

$$\begin{aligned} P_{rup_{ij}}\left(X<x\,|\,T\right) &= P_{rup_{ij}}\left(n=0\,|\,T\right)+P_{rup_{ij}}\left(n=1\,|\,T\right)\cdot P\left(X<x\,|\,rup_{ij}\right)+ \\ &\quad P_{rup_{ij}}\left(n=2\,|\,T\right)\cdot P\left(X<x\,|\,rup_{ij}\right)^2+\ldots \\ &= \sum_{k=0}^{\infty} P_{rup_{ij}}\left(n=k\,|\,T\right)\cdot P\left(X<x\,|\,rup_{ij}\right)^2 \end{aligned} \tag{B.29}$$

where $P_{rup_{ij}}\left(k\,|\,T\right)$ is the probability of the j[th] rupture in the i[th] source occurring *k* times in the time span *T*, and $P\left(X<x\,|\,rup_{ij}\right)$ is the conditional probability that parameter *X* does not exceed the given level *x* given the occurrence of rup_{ij}. Returning to the initial assumptions of independence between sources and independence of ruptures within each source, the probability that ground motion level *X* will exceed the threshold value *x* in the time span T is given as:

$$P\left(X\geq x\,|\,T\right)=1-\prod_{i=1}^{I}\prod_{j=1}^{J_i}\sum_{k=0}^{\infty}\mathrm{P}_{rup_{ij}}\left(k\,|\,T\right)\cdot P\left(X<x\,|\,rup_{ij}\right)^k \tag{B.30}$$

At this point in the formulation the assumption of earthquake rupture as a Poisson process does not necessarily have to be made. From equation (B.29) we see that any type of probability model can be supported provided that $P_{rup_{ij}}\left(k\,|\,T\right)$ can be defined explicitly. This permits the use of time-dependent temporal occurrence models such as the commonly used Brownian Passage Time (BPT) model (Matthews et al., 2002; Petersen et al., 2007).

In the case of time-independent seismic hazard it is normally assumed that the source model may consist of sources whose temporal occurrence is Poissonian, thus each rupture may have an average occurrence rate in the time interval *T* defined as λ_{ij} :

$$P_{rup_{ij}}\left(k\,|\,T\right)=e^{-\lambda_{ij}T}\frac{\left(\lambda_{ij}T\right)^k}{k!} \tag{B.31}$$

Putting equation (B.31) into (B.29) and making use of the Taylor series approximation:

$$e^x=\sum_{k=0}^{\infty}\frac{x^k}{k!} \tag{B.32}$$

for the Poisson case

$$\begin{aligned} P_{rup_{ij}}\left(X<x\,|\,T\right) &= e^{-\lambda_{ij}T}e^{\lambda_{ij}T\cdot P\left(X<x|rup_{ij}\right)} \\ &= e^{-\lambda_{ij}T\cdot\left(1-P\left(X<x|rup_{ij}\right)\right)} \\ &= e^{-\lambda_{ij}T\cdot P\left(X\geq x|rup_{ij}\right)} \end{aligned} \tag{B.33}$$

From this distribution the probability of at least one occurrence of rup_{ij} in the time span T is:

$$P_{rup_{ij}}\left(n\geq 1\,|\,T\right)=1-e^{-\lambda_{ij}T} \tag{B.34}$$

so equation (B.33) becomes:

$$P_{rupij}\left(X<x\,|\,T\right)=\left(1-P_{rup_{ij}}\left(n\geq 1\,|\,T\right)\right)^{P\left(X\geq x|rup_{ij}\right)} \tag{B.35}$$

which, then placed into (B.30) produces:

$$P\left(X\geq x\,|\,T\right)=1-\prod_{i=1}^{I}\prod_{j=1}^{J_i}\left(1-P_{rup_{ij}}\left(n\geq 1\,|\,T\right)\right)^{P\left(X\geq x|rup_{ij}\right)} \tag{B.36}$$

It is this probability calculation that forms the basis of the Classical PSHA kernel, and its equivalence with the more traditional rate-based formulation of PSHA (McGuire, 1995) is demonstrated by Field et al. (2003).

The probability definition defined in (B.36) allows for the structure of the calculator to be divided into two separate sections: the definition of the exhaustive set of ruptures possible within the source, and their corresponding probabilities $P_{rup_{ij}}\left(n\geq 1\,|\,T\right)$ and the means of calculating the calculation of the distribution of ground shaking given the occurrence of the rupture $P\left(X\geq x\,|\,rup_{ij}\right)$. As OpenQuake originally started life as a fork of the OpenSHA project (Field et al., 2003), see Pagani et al (2014a) for more detail, it retains the terminology adopted therein. Therefore for the remainder of this document we shall describe the exhaustive set of ruptures from all possible sources, and their corresponding probabilities of occurrence within a given time span *T*, as the ***EARTHQUAKE RUPTURE FORECAST***. The calculation of the probability distribution of the ground motion at a given site is determined from the GMPE.

3.1.2 "Event-Based" PSHA

The "Classical" PSHA calculator defines, for a given seismogenic source model, the exhaustive set of ruptures, and their corresponding probabilities, that can be generated given the configuration of the source. From the PSHA formulation described in (B.36), it is evident that the hazard curve is constructed by considering the full earthquake rupture forecast and, for each rupture, the probability of exceeding a given level of ground shaking from that rupture. In some applications, particularly those pertinent to seismic hazard inputs for critical infrastructure, a Monte Carlo application may be preferable (e.g. Musson, 1999; Weatherill & Burton, 2010; Assatorians & Atkinson, 2013). In such an application the seismic source model is sampled stochastically in order to generate a synthetic catalogue of events,

which can be considered as a possible realization of seismicity from the model within a given time span *T*.

OpenQuake-engine has the capability to generate synthetic catalogues (otherwise known as *stochastic event sets*) by sampling from the earthquake rupture forecast. As each rupture in the earthquake rupture forecast has an associated probability of occurrence one or more times within the time span (and assuming independence both within the rupture set for the source and between the sources), an occurrence of each ruptures within the time span can be determined by sampling the expected number of occurrences in the time span (for a Poissonian rupture) or applying inverse transform sampling (for a non-Poissonian source model).

For each event the stochastic event set OpenQuake can generate a field of ground motion, taking into account the uncertainty of the GMPE. From the rupture (described in a stochastic event set in exactly the same manner as it is in the classical PSHA) the logarithm of the expected ground motion and its corresponding variability are calculated from the GMPE. OpenQuake will then generate a random field of ground motion residual values, which are then multiplied by the standard deviation and added to the median value. If the user wishes to take into account spatial correlation in the ground motion residual terms in the manner described by Park et al. (2007), this option is available using the empirical spatial correlation model of Jayaram & Baker (2009). Each sample ground motion field is a random realisation of the ground motion at one or more sites, given the rupture occurrence. For an event set of total duration T_0 (the product of the sampling time span and the number of realisations), containing *K* ruptures, the rate of exceedance of the given level of ground motion is then determined directly from the set of Monte Carlo generated ground motions at the site:

$$\upsilon\left(X \geq x \,|\, T\right) = \frac{\sum_{k=1}^{K} H\left(X_k - x\right)}{T_0} \tag{B.37}$$

where X_k is the ground motion value *X* produced at the site from the kth rupture.

It has been shown by Musson (1999; 2012) that for a given site the hazard curves produced by the event-based approach and the classical approach converge for a large total duration. The use of the ERF for generating the ruptures in the stochastic event set further ensures that the two approaches can be considered equal for a sufficiently large number of event sets, albeit there is an initial computational overhead associated with generating the ERF from the source model.

3.1.3 Disaggregation

As is now common practice in PSHA, OpenQuake-engine can produce a disaggregation of ground motion hazard at a specific probability of exceedance. The OpenQuake-engine disaggregation calculator considers a model space of six dimensions: magnitude (M), source to site distance (R), longitude and latitude of the surface projection of the closest point of the rupture (λ,ϕ), the tectonic region type and the ground motion residual value, ε , defined as:

$$\varepsilon = \frac{x_y - \ln\left(Y\right)}{\sigma^{Y}_{TOTAL}} \tag{B.38}$$

where x_y is the ground motion value, and $\ln\left(Y\right)$ and σ^{Y}_{TOTAL} are the logarithmic mean and standard deviation of the expected ground motion respectively.

For a given model space bin, represented by $\mathbf{m} = \left(M, R, \lambda, \phi, TRT, \varepsilon\right)$, the probability of exceeding ground motion level *x* in the time-span is given by:

$$P\left(X > x \mid T, \mathbf{m}\right) = 1 - \prod_{i=1}^{I}\prod_{j=1}^{J_i}\begin{cases} P_{rup_{ij}}\left(X < x \mid T\right) & \text{if } rup_{ij} \in \mathbf{m} \\ 1 & \text{if } rup_{ij} \notin \mathbf{m} \end{cases} \tag{B.39}$$

By default, OpenQuake will produce the six-dimensional matric output with the contribution from each of the parameters bins **m**, across the whole dimensional space. It can also produce outputs in the reduced model space, such as for i) magnitude disaggregation $P\left(X > x \mid M\right)$, ii) distance disaggregation $P\left(X > x \mid R\right)$, iii) tectonic region type disaggregation $P\left(X > x \mid TRT\right)$, iv) magnitude-distance disaggregation $P\left(X > x \mid M, R\right)$, v) magnitude-distance-epsilon disaggregation $P\left(X > x \mid M, R, \varepsilon\right)$, vi) longitude-latitude disaggregation $P\left(X > x \mid \lambda, \phi\right)$, vii) longitude-latitude-magnitude disaggregation $P\left(X > x \mid \lambda, \phi, M\right)$ and viii) longitude-latitude-tectonic region type disaggregation $P\left(X > x \mid \lambda, \phi, TRT\right)$.

3.1.4 "Scenario" Hazard

As a final calculator, OpenQuake-engine can produce multiple fields of ground motion from a given model of rupture, sampling the uncertainty in the ground motion residuals in the same manner that is described for the event-based calculator. The inputs, in this case, are simply a rupture scenario model (a combination of a rupture geometry and a given magnitude) and the choice of GMPE. The users are then able to select the number of fields that they may wish for the sample.

3.2 SEISMIC SOURCE CHARACTERISATION

For the widest possible application, OpenQuake can consider many different source typologies that may often be used in many different PSHA models. These include sources that directly characterise the fault geometry, with varying degrees of complexity, as well as those that are based in distributed seismicity (point or uniform area sources). Each source requires that enough information be supplied to define the rupture geometry for each magnitude within the magnitude frequency distribution, in addition to the probability of occurrence of the specific magnitude.

Each source, regardless of type, requires three common attributes: a unique identifier (typically a string), a name and a tectonic region type. The latter is used within the context of the PSHA calculation to identify the choice of GMPEs that may be associated with the specific source. The tectonic region type provides a mapping, such that the GMPE logic tree (discussed in more detail in 3.4) can apply the specific set of GMPEs for the given region type to the source in question. No "general" regionalisation is assumed in OpenQuake, and thus the modeller has the capability to assign whatever region type they wish to the source, and they can also associate any GMPE logic tree to that region type, regardless of whether the GMPE was derived for that source. So, for example, it is possible to apply GMPEs derived from active shallow crustal regions to sources in stable regions, and vice-versa.

3.2.1 Magnitude Frequency Distributions

For all but the non-Parametric source typology, each source must be associated with a specific magnitude frequency distribution. At the time of writing, OpenQuake supports the definition of three such distributions:

- Truncated Gutenberg & Richter

The standard exponential model parameterised by an a-value, a b-value and a minimum and maximum magnitude

- Youngs & Coppersmith (1985)

This is the hybrid exponential model proposed by Youngs and Coppersmith (1985), which contains an exponential model with a "boxcar" function at the high magnitude. The boxcar function is constrained between $M_{ch} - 0.25 \leq M_{ch} \leq M_{ch} + 0.25$, where M_{ch} is the characteristic magnitude. The function can be created in two different ways. The first is from the rate of the characteristic magnitude, requiring a minimum magnitude, a b-value, a characteristic magnitude and its corresponding rate. The alternative is to define the magnitude frequency distribution (MFD) from the total moment rate, which may be more appropriate in cases where the recurrence is modelled directly from the slip on a specific fault, or from the moment rate within a deforming region. In that case the total moment rate is required as an input, rather than the occurrence rate.

- Incremental Magnitude Frequency Distribution

The final approach requires no specific model parameterisation, but instead the occurrence rate for each magnitude bin is specified directly. In this case the modeller must provide a vector of occurrence rates for each magnitude bin (i.e. incremental rates rather than cumulative), in addition to the minimum magnitude of the bins and the corresponding bin width. This approach therefore permits magnitude frequency distributions of any form to be supported by OpenQuake.

3.2.2 Temporal Occurrence Model (TOM)

In the formulation of the seismic hazard integral in (B.36) it is clear that for each rupture it must be possible to define the probability of one or more occurrence of the rupture within the time-span *T*. The TOM is the choice of model that can be used to determine this information from the recurrence model for the rupture. For time-independent applications it is the homogeneous Poisson model that is used (referred to as PoissonTOM), which determines the probability of one or more occurrences of the rupture from the occurrence rate.

3.2.3 Point Source

The simplest form of seismogenic source is one in which the activity rate is characterised at a single point. The location of the point on the Earth's surface is given by a longitude and latitude value. This typology can be useful is the activity rate is defined for a regular grid of cells that may be represented by a seismogenic source in the centre of the cell.

Whilst the use of gridded seismicity may be common in PSHA modelling, the greatest challenge in adopting such an approach lies not necessarily in characterising the activity rate, but in characterising the finiteness of the rupture within the source. The vast majority of modern GMPEs require that the seismogenic source be characterised in terms of a finite rupture plane rather than a point. This presents a considerable number of challenges to the modeller in order to constrain the probability distribution of finite ruptures that can be anchored to the point but that are also consistent with the broader tectonics of the region.

For a point source with a known magnitude frequency distribution, OpenQuake generates finite planes whose respective areas are related to the magnitude via the magnitude scaling relation (e.g. Wells & Coppersmith, 1994; Hanks & Bakun, 2008), and whose aspect ratio is specified as a property of the source. For each magnitude in the MFD of the point source, the area is calculated from the magnitude scaling relation, and the corresponding length and width of the finite rupture determined from the aspect ratio. The centroid of this finite rupture will correspond to the hypocentral location of the point; therefore the modeller must define the possible hypocentre depths using a probability mass function. The probability mass function is input simple as a list of values and their corresponding probabilities. For example, a modeller may wish to consider three different hypocentral depths: 5.0 km, 10.0 km and 15.0 km, so each must be specified with a corresponding probability (e.g. 0.2, 0.6, 0.2), the sum of which must equal to 1.0. Probability mass functions are used in many different parts of OpenQuake and this terminology will be referred to throughout.

Given that a finite rupture plane is then generated with a given length and width, and that the centre of this plane is anchored to the three dimensional location of the point, the modeller can constrain the upper and lower depths to which the rupture can propagate. This is done so that the modeller can limit the rupture thickness to the seismogenic crust in the case of large magnitude earthquakes. If the length, width and location are such that the rupture cannot be contained within the specified seismogenic thickness, the aspect ratio is then no-longer preserved and instead the length will increase with and the rupture width with be constrained to the seismogenic thickness.

Finally, the orientation of the rupture within the half space needs to be specified. To do this the modeller must specified another probability mass function, this time providing one or more "nodal planes" (literally a rupture plane defined by a strike, dip and rake), each with their corresponding probability.

An illustration of the way in which finite ruptures are generated for point sources is given in Figure 3-2.

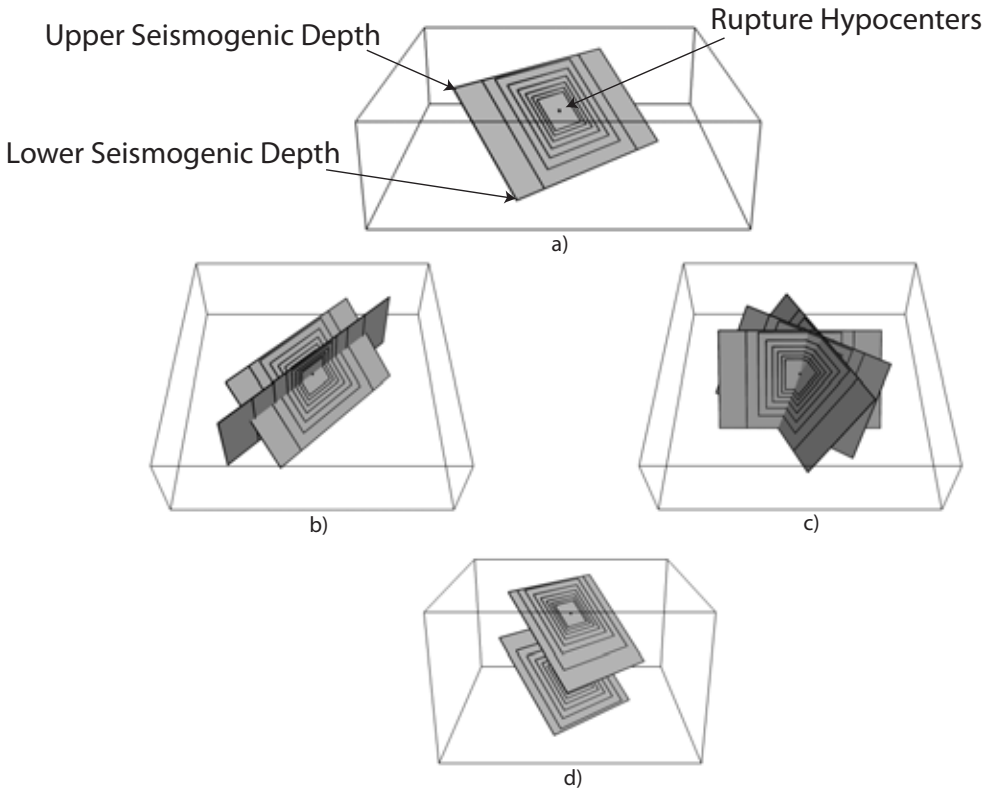

Figure 3-2: Generation of the finite rupture from the point source typology: a) spatial configuration, b) example of a case with two dip angles, c) case for multiple strike angles, and d) case for multiple hypocentral depths. (Figure from Pagani et al., 2014b)

3.2.4 Area Sources

The representation of the seismogenic source in terms of an area of uniform activity rates is a long-standing, and still commonly used, means of characterising the seismogenic source. Within most PSHA software it is common practice to discretise the uniform polygon into a discrete mesh of points, with the activity rate divided evenly between the points (possibly compensated for latitudinal change in spacing if the points are spaced evenly in a geodetic coordinate system). As the area source can be considered a collection of point sources, whose attributes are shared for each point, then the list of parameters is the same. The only exception is that the source geometry is specified as a polygon, with the user inputting the longitude and latitude coordinates. The effective manner in which OpenQuake generates the rupture set within an area source is illustrated in Figure 3-3.

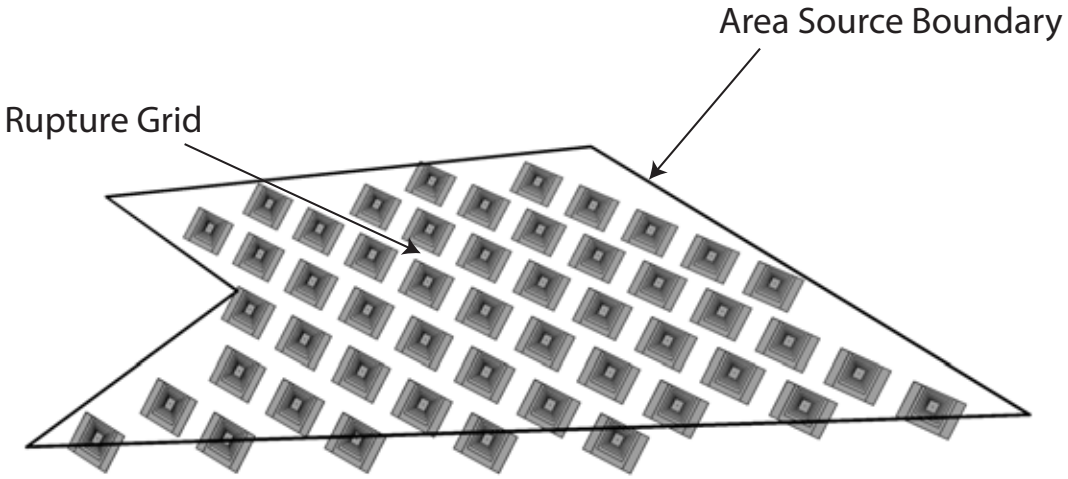

Figure 3-3: Distribution of finite ruptures within an area source

3.2.5 Simple Fault Sources

The simple fault source is the first of several fault typologies that enables the generation of potential ruptures across a fault surface. The geometry of the fault is defined such that the upper and lower edges are parallel. A mesh of evenly spaced points is then rendered across the fault, with a spacing that is configured by the user. To generate the earthquake rupture forecast from this source, given a defined magnitude frequency distribution, for each magnitude the corresponding rupture area is determined from the magnitude scaling relation. The user must specify a rupture aspect ratio, which is then used to determine the length and width of the rupture for the calculated area. Once the rupture geometry is determined, OpenQuake will then place that rupture on the fault mesh, moving the rupture incrementally across the mesh until the full set of possible rupture locations for that given magnitude is gathered. Each rupture for the given magnitude is assigned an equal probability, which is taken from the corresponding bin of magnitude frequency distribution and divided evenly between the numbers of ruptures for the magnitude. The process will be repeated for the following magnitude bin until all of the possible ruptures for the whole MFD are collected. If the largest magnitude for the fault source gives and area than is greater than or equal to that of the whole fault surface then the whole surface is used as the rupture (thus implying that larger magnitudes are due to greater slip on the fault surface). The rupture floating process is illustrated in Figure 3-4.

Where a hypocentral location is needed (as in the case of GMPEs relying on hypocentral distance or hypocentral depth) this is currently taken as the centroid of the rupture surface. However, at the time of writing, new functions are currently under development to allow for the possibility to float the hypocentre location within the rupture plane for the purposes of modelling directivity in the ground motion.

To generate the fault geometry four pieces of information are needed: i) the trace of the fault as a line in terms of longitude and latitude, which represents the projection ***up-dip*** from the top of the fault edge to the Earth surface (or the top fault edge if the fault reaches the surface), ii) the upper seismogenic depth of the fault, iii) the lower seismogenic depth of the fault and iv) the dip of the fault. Each of the upper and lower edges has a constant depth, thus making them parallel to each other, and to the Earth surface. In addition to the fault

geometry and the other critical information (i.e. identifier, name, tectonic region type and magnitude frequency distribution). The modeller must also assign a magnitude scaling relation, a rupture aspect ratio and a rake value to each fault source.

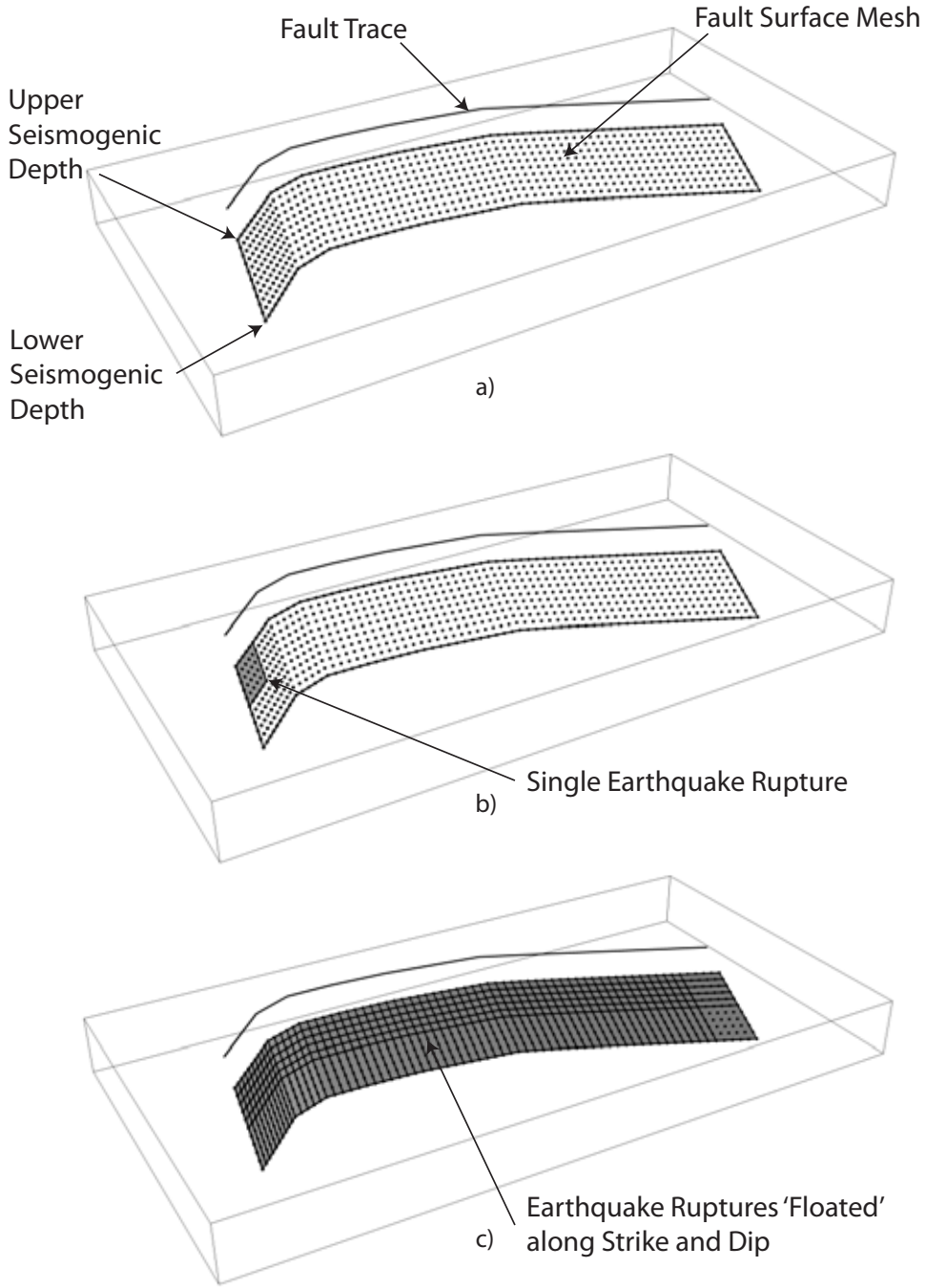

Figure 3-4: Simple fault source and "rupture floating" within OpenQuake, a) generation of the mesh, b) creation of the earthquake rupture, c) enumerated set of possible rupture locations for the rupture set in (b). Figure from Pagani et al. (2014b)

3.2.6 Complex Fault Sources

The simple fault source typology is appropriate for many applications in which the faults are shallow, or may, within the understanding of the tectonic environment, be considered geometrically regular. There are cases, however, in which we may expect the fault dip to change with depth (such as in the case of a listric fault), or where we may expect the lower edge of the seismogenic zone to no longer be parallel with the upper edge (as is common for subduction zones). To support these cases OpenQuake defines the complex fault typology. Here the geometry of the fault surface is specified by a series two or more fault edges, each of which is a line of longitudes, latitudes and depths. The edges no longer have to be parallel and do not require that the depth must be constant within each edge. From this irregular surface a mesh of points is then generated (Figure 3-5); however, unlike the case for the simple fault, the mesh is not evenly spaced (only the spacing of the upper edge is controlled by the user). The rupture floating procedure works in a similar fashion as it does for the simple fault source typology. However, because of the irregularity in mesh spacing, for each rupture the algorithm will search for the best combination of mesh points that satisfies the predefined area and aspect ratio. Whilst the optimised area of the rupture may not exactly correspond to the area that is expected for the magnitude from the given scaling relation, the overall disagreement is relatively small.

The remainder of the attributes of the complex source are the same as for the simple fault source typology.

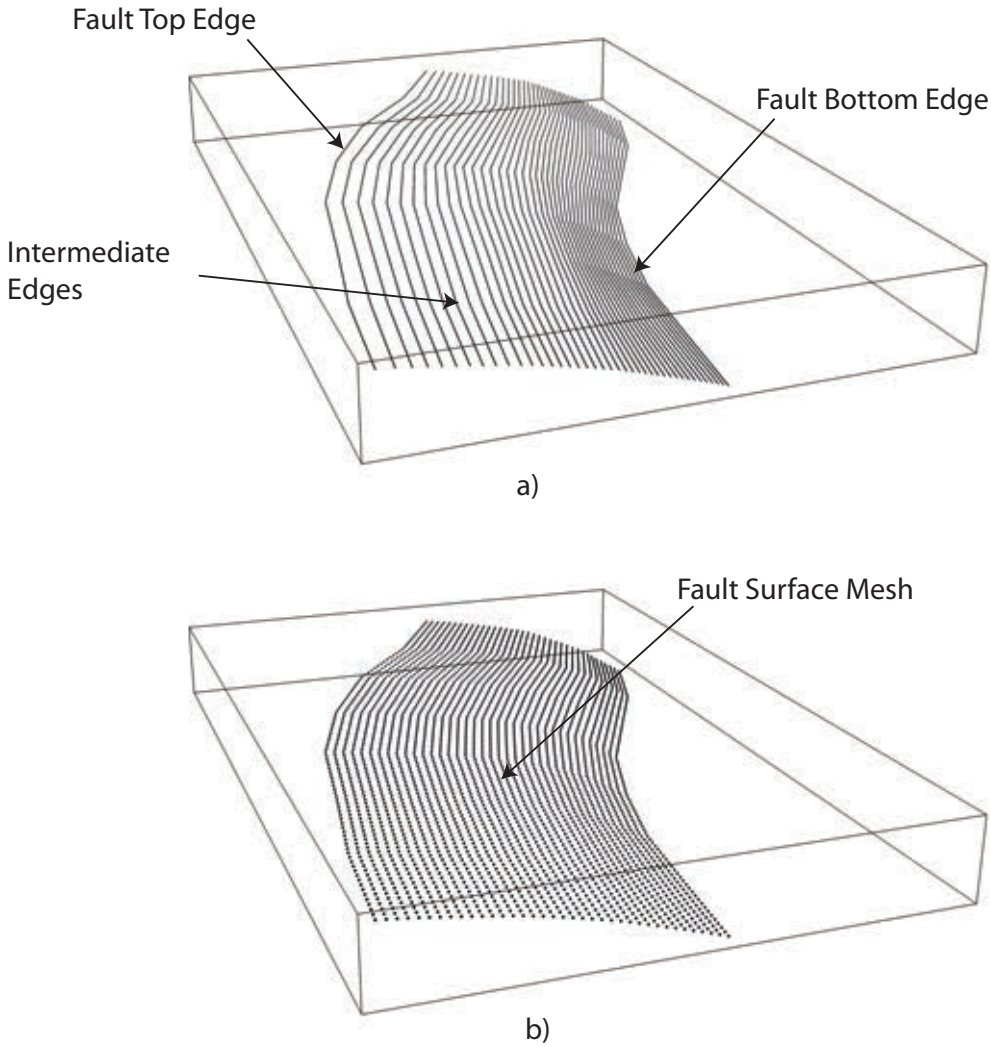

Figure 3-5: Generation of the fault mesh for the complex typology, a) definition of the fault edges, and b) generation of the fault surface mesh.

3.2.7 Characteristic Fault Source

The characteristic fault source is one in which the geometry of the rupture is fixed, regardless of the magnitude frequency distribution on the fault. It is assumed that the rupture will extend across the entire geometry for each magnitude. Such a typology can be useful for purely characteristic models of fault behaviour in which future ruptures are believed to occur on the same extent fault, and that if there is variability in magnitude then this may be due to variability in slip on the surface rather than rupture area.

As the rupture extent is assumed fixed to the specified rupture surface (i.e. no floating of the rupture across the surface) then the modeller has more freedom to specify the geometry of the surface. OpenQuake event supports the possibility of discontinuous surfaces, such as multi-segment ruptures with possible offsets and/or step-overs, and even different typologies from segment to segment. The modeller may choose to define one or more segments, each of which may have the following geometry typologies:

i) Planar Surface – a regular rectangle described by the longitude, latitude and depth of the four corner points (see Figure 3-6)

ii) Simple fault surface – as for the simple fault, a regular surface defined by a surface trace (up-dip projection of the fault to the surface), an upper seismogenic depth, a lower seismogenic depth and a dip.

iii) Complex fault surface – as for the complex fault, an irregular surface is defined by a series of edges

As no floating of the ruptures is undertaken, the user does not need to specify a magnitude scaling relation or rupture aspect ratio. A rake value is still required to characterise the style of faulting.

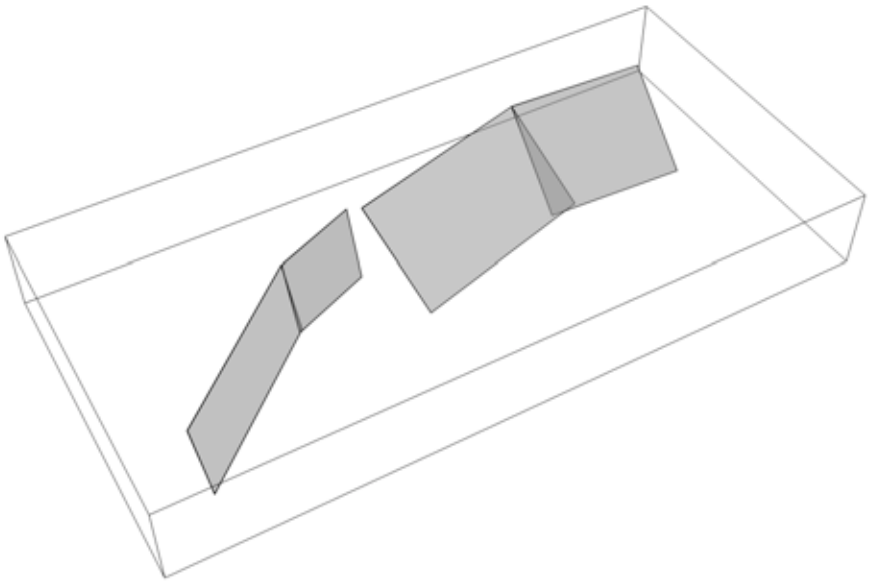

Figure 3-6 Multi-segment planar ruptures for a "characteristic" source. Note that in this case the rupture surface is considered to always correspond to the whole multi-segment rupture

3.2.8 Non-Parametric Seismic Source

The Non-Parametric seismic source differs from the other OpenQuake source typology in that they do not represent a single source with a defined magnitude frequency distribution and TOM, from which the earthquake rupture forecast is generated. Instead the Non-parametric seismic source is itself an earthquake rupture forecast. That is to say that it represents a collection of ruptures for which the probability of N occurrences is specified explicitly for each rupture. When using this typology, OpenQuake no longer needs to assume a Poissonian model, as the probabilities may be defined from a time-dependent process. The probabilities should be defined as a probability mass function in which the numbers of occurrences of the ruptures are associated with their corresponding probabilities. So for the case of a rupture in which the probability of zero occurrences in the time span, *T*, is 0.9, the probability of one occurrence is 0.09 and the probability of two occurrences is 0.01, the input would be given as a list of paired values: *[(0, 0.9), (1, 0.09), (2, 1.0 – (0.9 + 0.09)].*

Table 1 Required source model attributes by source typology

Source Typology	Geometry Attributes
Point Source	• Identifier • Name • Tectonic Region Type • Point Location (Longitude, Latitude) • Upper Seismogenic Depth • Lower Seismogenic Depth • Rupture Aspect Ratio • Magnitude Scaling Relation • Magnitude Frequency Distribution • Nodal Plane Distribution • Hypocentral Depth Distribution
Area Source	• Identifier • Name • Tectonic Region Type • Polygon Coordinates (Longitudes, Latitudes) • Upper Seismogenic Depth • Lower Seismogenic Depth • Rupture Aspect Ratio • Magnitude Scaling Relation • Magnitude Frequency Distribution • Nodal Plane Distribution • Hypocentral Depth Distribution
Simple Fault Source	• Identifier • Name • Tectonic Region Type • Fault Trace* (Longitudes, Latitudes) • Upper Seismogenic Depth • Lower Seismogenic Depth • Dip • Rupture Aspect Ratio • Magnitude Scaling Relation • Magnitude Frequency Distribution

	• Rake
Complex Fault Source	• Identifier • Name • Tectonic Region Type • Edges of the fault plane • Rupture Aspect Ratio • Magnitude Scaling Relation • Magnitude Frequency Distribution • Rake

Characteristic Fault Source	• ID
	• Name
	• Tectonic Region Type
	• Planar Fault Geometry/Geometries:
	  o Corner points (Longitude, Latitude, Depth) of the four nodes of the rectangular plane
	• *OR* Simple Fault Geometry/Geometries:
	  o Fault Trace
	  o Upper Seismogenic Depth
	  o Lower Seismogenic Depth
	  o Dip
	• *OR* Complex Fault Geometry/Geometries:
	  o Fault Edges (Longitudes, Latitudes, Depths)
	• Magnitude Frequency Distribution

Non-parametric Fault Source	• ID • Name • Tectonic Region Type • Planar Fault Geometry/Geometries • OR Simple Fault Geometry/Geometries • OR Complex Fault Geometry/Geometries • Probability mass function as a list of probability of N occurrences (e.g. for a rupture with a 0.01 probability of one occurrence in time T, the list should be: [0.99, 0.01], for n in the range 0 to N)

3.3 GROUND MOTION CHARACTERISATION

OpenQuake boasts an extensive, and ever-growing, library of ground motion prediction equations. For the most up-to-date list of supported GMPEs the reader is encouraged to consult the online documentation for GMPEs implemented in OpenQuake, which is found at the following link (http://docs.openquake.org/oq-hazardlib/master/gsim/index.html).

From a computational perspective, the incorporation of new GMPEs into any PSHA software is expected to be one of the areas of the code where new features will be implemented regularly. Furthermore, for many applications it is increasingly common to take existing GMPEs and to make modifications to calibrate them to local conditions. Examples of this can be seen in the use of single-station sigma or V_S-kappa corrections. Whilst addition of new GMPEs, or modification of existing ones, still requires direct access to the source code, the OpenQuake implementation is designed to facilitate this process to make it easier for a user to make such common adjustments. For the main OpenQuake library, however, the GMPEs go through a rigorous test-driven development process. This process requires that the GMPEs be automatically tested against tables of expected values, which should have been generated from independent implementations (i.e. from those of the GMPE authors themselves).

To harmonise the implementation of GMPEs, a "base" class is created, which contains both the template for the GMPE and the methods for calculating the probabilities of exceeding a given ground motion level once the mean and standard deviation of the ground motion are known. Every GMPE implemented in OpenQuake is created, firstly, by inheriting this abstract base class. The actual function of the GMPE is implemented inside the method within the class called `.get_mean_and_stddevs`, which executes the main GMPE code. The modeller must, however, specify a series of attributes that contain additional information about the GMPE. These attributes include the tectonic region type for which the GMPE is defined, the intensity measure types supported, the definition of the horizontal component, the standard deviation types (total, inter-event, intra-event) provided by the GMPE, the required site parameters, the required rupture parameters and the GMPE distance metric. Finally, the GMPE coefficients are entered into the code in the form of a standard table, which lists, for each period, the coefficients for each parameter in the GMPE, and when running the code, will automatically generate interpolated coefficients for any spectral periods not directly stated in the coefficients table *within the spectral limits of the GMPE*.

More details specific to a particular GMPE implementation for induced seismicity will be discussed in the next chapter.

3.4 EPISTEMIC UNCERTAINTY

Modern practice of seismic hazard analysis requires the characterisation of both the aleatory, or irreducible, variability of the earthquake process, and the epistemic uncertainty. PSHA incorporates the former directly within the seismic hazard integral, whilst the latter is commonly analysed with the use of logic trees. The logic tree provides a formal framework for combining alternative interpretations, and uncertainties in their respective parameterisations, into a single branching structure from which a spread or distribution of hazard outcomes can be quantified. Alternative models, and/or parameterisations, of the critical elements within the PSHA process are associated with respective weights. From

each of the models, or a sufficient weighted sample, a single hazard curve may be produced. For a given level of ground motion, the distribution of possible exceedance probabilities may be determined from the mean or from fractiles of the weighted output hazard curves.

Given the widespread adoption of logic trees in practice, OpenQuake integrates logic trees into the calculation. In fact, it goes further than this in so much as every calculation is input as a logic tree, in which alternative parameterisations of the seismic source are separated from, and later combined with, the selection of multiple ground motion models. The OpenQuake logic tree definition is broken into three components, illustrated in Figure 3-7. In this example we consider the epistemic uncertainty on two parameters of a source (or source model). Above the red line in Figure 3-7 are the alternative values for the parameters considered, in this case fault dip and upper seismogenic depth. For each of the parameters the alternative values are each represented as a *branch*, requiring a parameter value and a corresponding weight. The total set of alternative values for the parameter should represent the total epistemic uncertainty for the parameter in questions. This is known as the *branch set*, and the sum of the weights of all of the alternative branches within the set must be equal to 1.0. Both the branch set, and each unique branch must be assigned a unique identifier. Below the red line in Figure 3-7 we see how the *branch sets* are placed within the logic tree. The location of the *branch set* within the tree is known as the *branching level.* It can be seen here how multiple *branch sets* are applied such that the possible combinations of parameter uncertainties can be effectively exhausted. Each individual combination of parameters (i.e. each individual pathway through the logic tree from left to right) is an *end branch*, and its individual weight will be the product of the weights of each of the branches traversed to reach it. Provided the weights on all of the individual branches in each *branch set* sum to 1.0, so too does the sum of the weights of all of the possible end branches.

It may not take much imagination to recognise that as more epistemic uncertainties are represented within the logic tree the size of the tree increases exponentially. To assess the final distribution of the epistemic uncertainty of a model, OpenQuake-engine provides two options: i) *full path enumeration*, in which a hazard curve is calculated for every branch of the logic tree, or ii) *weighted random sampling*, in which N_{SAMP} *end branches* are sampled and the final distribution inferred from these. As complex logic trees may result in total numbers of end branches on the orders of hundreds of thousands, or millions, sampling may often be an acceptable approach.

A *branch set* can be applied to all of the sources in the source model(s) or to a subset of sources, or even just to specific *branches* of the precious *branching level*. When defining a *branch set* the modeller has a choice as to which of those options is preferred. Four options are currently supported:

- *applyToBranches*: apply the branch set to one or more branches of the previous branching level, defined by their corresponding unique identifiers.
- *applyToSource*: apply the branch set only to specific sources within a source model
- *applyToSourceType*: applies the epistemic uncertainties only to those sources of a specific type.
- *applyToTectonicRegionType*: the branch set is applied to all sources belonging to a particular tectonic region.

The logic tree structure presented here enables the possibility to support different epistemic uncertainties on different elements within the model. Whilst the logic tree representation is the same for those uncertainties pertaining to the seismic source and those to the ground

motion model, the two systems must be specified in separate files as input to OpenQuake. Various validation checks are performed at the start of the calculation to ensure the logic tree is correct and computationally realisable.

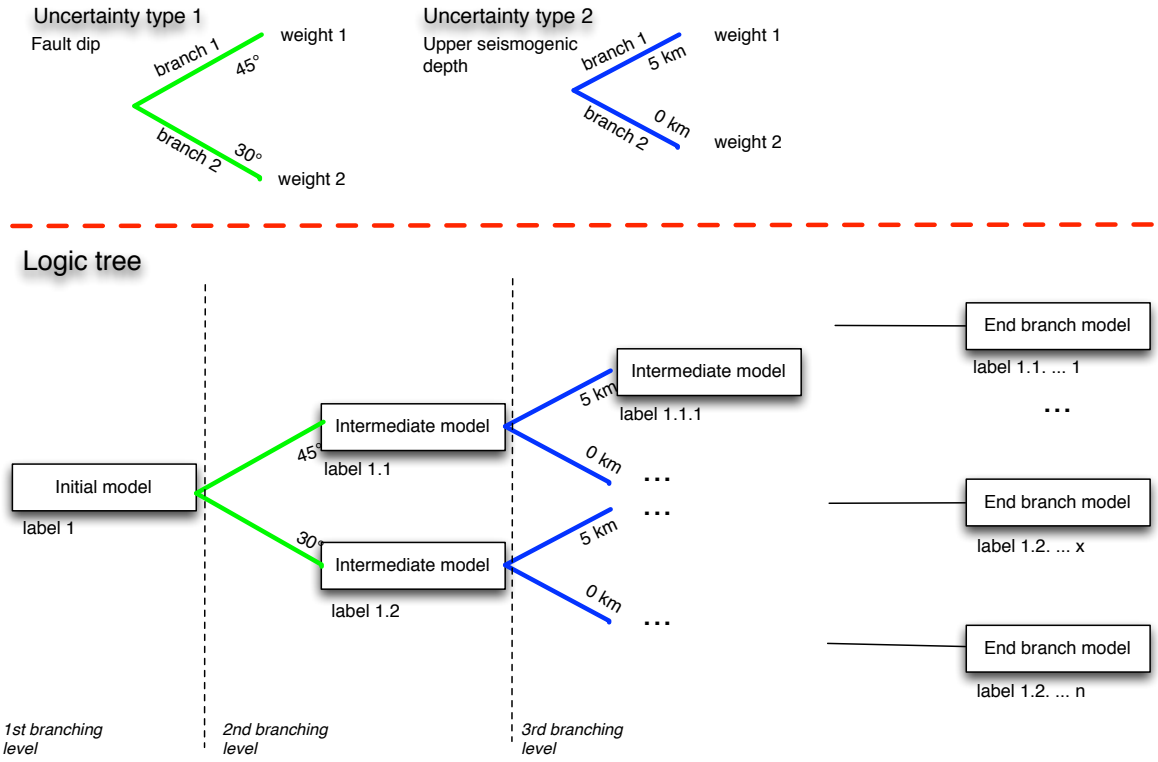

Figure 3-7: Structure of the modular logic tree within the OpenQuake-engine. For a given fault source, two epistemic uncertainties are defined here: fault dip and upper seismogenic depth.

3.5 OPENQUAKE: IN SUMMARY

The OpenQuake-engine was designed for application to seismic hazard and risk assessment in a wide variety of different contexts and environments. As such, it supports multiple methods for assessing earthquake hazard (i.e. classical PSHA, event-based, scenario) and a variety of different seismogenic source representations to capture the range of modelling possibilities. Whilst this flexibility is appealing, so too in its capacity to incorporate GMPEs with relative ease and a high level of quality assurance, it is the general architecture of the software that makes it ideal for customising to new applications. In the next section we shall be making use, not of the whole OpenQuake-engine, but of the *oq-hazardlib*, for induced seismicity hazard assessment. The *oq-hazardlib*, as the name implies, is capable to operate as a stand-alone library of tools for probabilistic seismic hazard assessment. By defining the induced seismicity in a manner that is consistent with the *oq-hazardlib* we can integrate a probabilistic seismic hazard analysis into any custom application.

4 Customising OpenQuake For Induced Seismicity Hazard with Illustrative Applications

In section 2 of this report, the review of existing induced seismicity PSHA applications highlighted the fact that it is not necessarily possibly to consider “induced seismicity hazard” as a single approach or method, but as an adaptation of the more general PSHA process to the behaviour of the phenomenon in question. It should be emphasised therefore, that it is not necessarily possible to develop induced seismicity seismic hazard software, but instead to demonstrate how existing software can be applied to induced seismicity applications.

Largely due to its open and test-driven development process, the OpenQuake engine is an optimum candidate for experimenting with customisations designed for induced seismicity applications similar to those described in section 2. This chapter will outline the additions made to OpenQuake that may facilitate application in induced seismicity contexts, before discussing an illustrative application of OpenQuake for the calculation of seismic hazard from a fluid injection process using the geomechanical seed model described in 2.3.

4.1 INDUCED SEISMICITY GROUND MOTION PREDICTION EQUATIONS IN OPENQUAKE

In the context of the STREST project, four GMPEs for induced seismicity applications have been implemented: Dost et al., (2004), Convertito et al. (2012), Douglas et al. (2013) and Atkinson (2015). The particulars of each GMPE are summarised in Table 2, and described in more detail in section 2.6.

Of the four GMPEs, Dost et al. (2004), Convertito et al. (2012) and Atkinson (2015) are largely trivial implementations. They have a relatively simple functional form and are dependent only upon magnitude and hypocentral distance. For the Convertito et al (2012) GMPE a notional site-scaling term is included; however, the site factor is calibrated only for the Geysers Geothermal field, and may need to be re-adjusted if applied in other regions.

The GMPE of Douglas et al. (2013) presents a greater challenge in the implementation due to the nature of the GMPE. Within its architecture, OpenQuake is designed to be rigid in the manner in which it separates epistemic uncertainties from aleatory variability. In the case of Douglas et al. (2013), the parameters $\Delta\sigma$, Q and κ are not considered as continuous variables. Were they to be so, it would be necessary to include new parameters into the respective source, distance and site classes inside of the code. From this, however, it would need to be assumed that the expected mean and standard deviations of the ground motion intensity at intermediate values of these three parameters could be predicted by interpolation from the GMPE at the values defined. As Douglas et al. (2013) and Edwards and Douglas (2013) illustrate, the 36 coefficient sets represent $\Delta\sigma$, Q and κ as epistemic uncertainties, the selection of the preferable coefficient set (or sets) being dependent on the application in question. OpenQuake reflects this and implements all 36 different options as individual

GMPEs. The nomenclature of the GMPE class names follows the pattern `DouglasEtAl2013StochasticSD###Q###K###`, where SD### takes the value 001, 010, 100 for stress drops of 1, 10 and 100 bars respectively, Q### takes the value of 200, 600 and 1800 for the corresponding Q values, and K### takes the value 005, 020, 040, 060 for κ of 0.005, 0.02, 0.04, 0.06 respectively. To ensure correctness of implementation, we generated verification tables from the Fortran implementation provided by J. Douglas (pers. Comm).

Table 2: Summary of characteristic of induced-seismicity GMPEs implemented within OpenQuake (see section 2.6 for more detail)

GMPE	Intensity Measure Type	Required Parameters	Notes
Convertito et al. (2012)	• PGA • Larger of the two components	• M_W • Hypocentral Distance • Site Term (V_{S30})	Empirical model derived from strong motion records of the Geysers Geothermal System
Douglas et al. (2013)	• PGA • PGV • SA (0.005 s to 0.5s) • Random horizontal	• M_W • Hypocentral Distance • *No site term*	Implements different GMPE coefficient sets for combinations of $\Delta\sigma$ (1, 10, 100 bar), *Q* (200, 600 and 1800) and κ (0.005, 0.02, 0.04, 0.06). In total 36 GMPEs
Dost et al., (2004; updated 2013)	• PGA • PGV • Geometric mean	• M_L • Hypocentral Distance • *No site term*	Two simple GMPEs for induced events in the Groningen field
Atkinson (2015)	• PGA • PGV, • Sa (0.03 s – 5.0 s) • RotD50	• M_W • Hypocentral Distance • *No site term*	GMPE for small magnitude (M_W 3.0 – 6.0) tectonic events at close distances

The implementation of the Douglas et al. (2013) also required some additional details that would not necessarily be apparent from within the paper itself, but were clarified by the author. As discussed previously, the aleatory variability of the GMPE is not derived directly from the simulated records but from empirical recordings of induced seismicity earthquake sequences (see section 2.6.3 for details). For PGA and spectral acceleration at periods less than 0.1 s, the σ_{TOTAL} term for Sa (0.1) is used. PGA is assumed equal to Sa (0.005 s) in this case.

4.2 MODIFICATIONS TO THE SEISMIC SOURCE MODEL

As shall be seen in the following exercise, the flexibility of OpenQuake to support many different source typologies, and the separation of the source geometry from the rupture probabilities, enables it to be customised to different applications without the need for modifications to the code. As such, the source typologies remain unchanged, but we shall see in the following examples how they can be used in a temporally dependent application.

4.3 APPLICATION OF OPENQUAKE FOR AN ENHANCED GEOTHERMAL SYSTEM INDUCED SEISMICITY HAZARD ANALYSIS

To illustrate the ability of OpenQuake to be adapted to an induced seismicity context, a self-contained program has been created to implement the geomechanical seed model approach, as proposed by Goertz-Allmann & Wiemer (2013), and combine it with the OpenQuake hazard library to provide: i) stochastically generated sets of earthquakes based on different injection scenarios, and ii) curves and maps of the probability of exceeding a given level of ground motion using one or more induced seismicity GMPEs. The following outlines the architecture of such a code, and provides an *illustrative* example using data from the Basel injection experiment. It should not, however, be considered a rigorous or robust assessment of the resulting seismic hazard from the Basel injection induced seismicity sequence.

4.3.1 Core-Components and Workflow

The core structure of the software is broken down into two main class sets: i) the generator of the pressure field, and ii) the geomechanical seed model. The first contains the methods by which a pressure field can be generated from a given set of injection parameters, whilst the latter generates the corresponding synthetic event set from the seeds and, when accompanied by a given GMPE, can provide the seismic hazard curves for a set of input sites. The object-oriented nature of the program is designed so that different methods can be incorporated with great ease, provided that the methods follow a process whose inputs and outputs are consistent. This is a process known as *abstraction*, which can be seen throughout the OpenQuake code. It facilitates the addition of new methods into the software, such as GMPEs, magnitude scaling relations or seismic source typologies, whilst minimising, or ideally avoiding, propagation of changes throughout the code architecture.

The precise workflow of the code largely re-states the description of the process given in section 2.3; however, a simplified flowchart is given in Figure 4-1. The inputs are gathered together in two groups: *injection system configuration* and *geomechanical seed configuration*. The former controlling the information regarding the fluid injection process, the latter controlling the geomechanical seed model. The user must specify the start time and the end time for which they wish to generate the synthetic catalogue, whilst the shut-in time of the injection is a parameter within the injection system configuration. This allows the user to define the seismicity following shut-in. In the current formulation background seismicity is not included. However, if an initial background catalogue is needed from a pre-defined source zone accompanying the injection process then this can be generated via OpenQuake standard tools for calculating an event set from a seismic source zone. All dates and times

are represented using Python's "datetime" module to allow date and time operations to be precise to the nearest microsecond (https://docs.python.org/2/library/datetime.html). For the generation of probabilistic seismic hazard curves, the user must specify the choice of GMPE, the intensity measure type (PGA, PGV, Sa etc), and the target intensity measure levels fro calculating the probability of being exceeded.

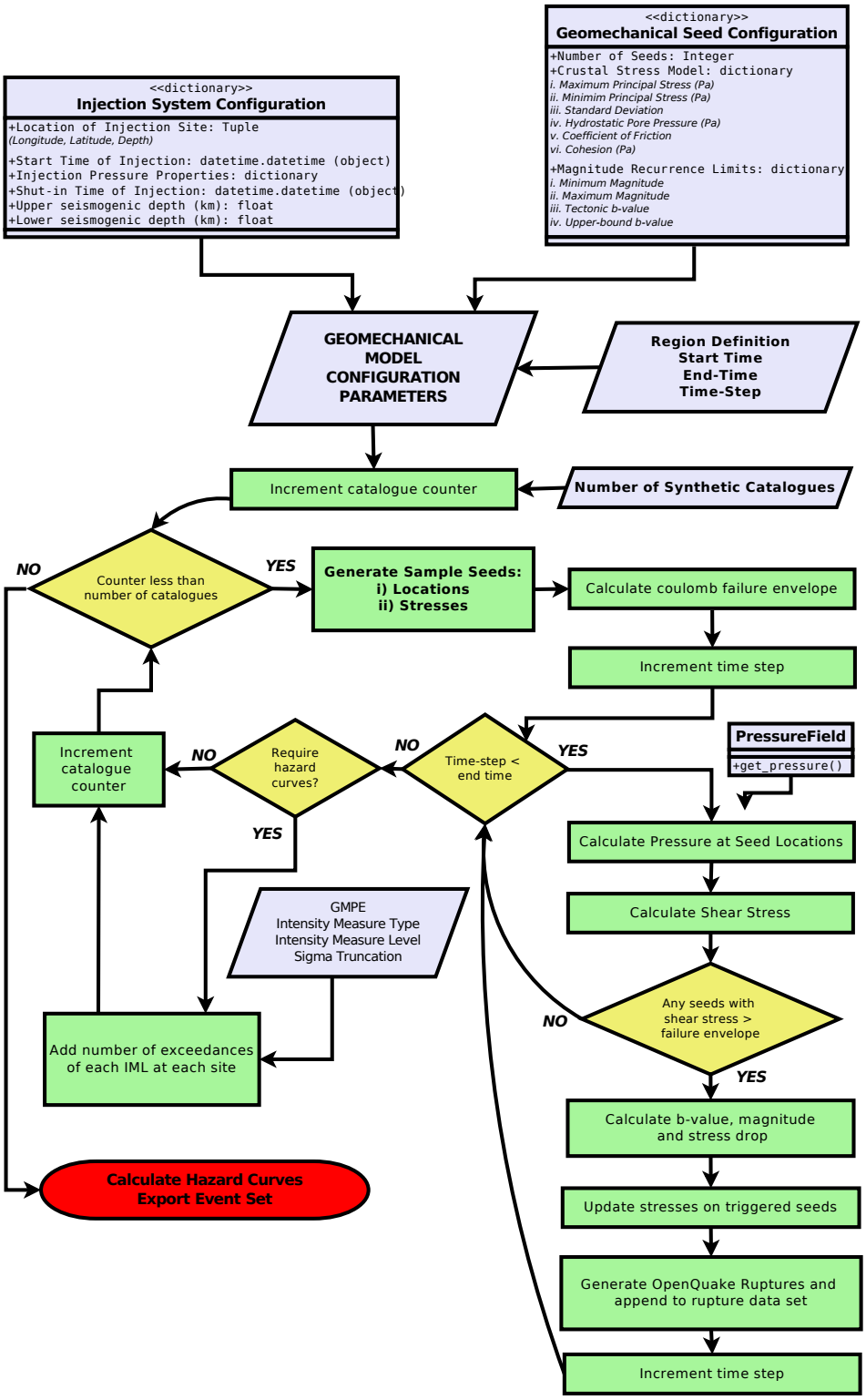

Figure 4-1 Flowchart of Geomechanical Seed Induced Seismicity Hazard Algorithm

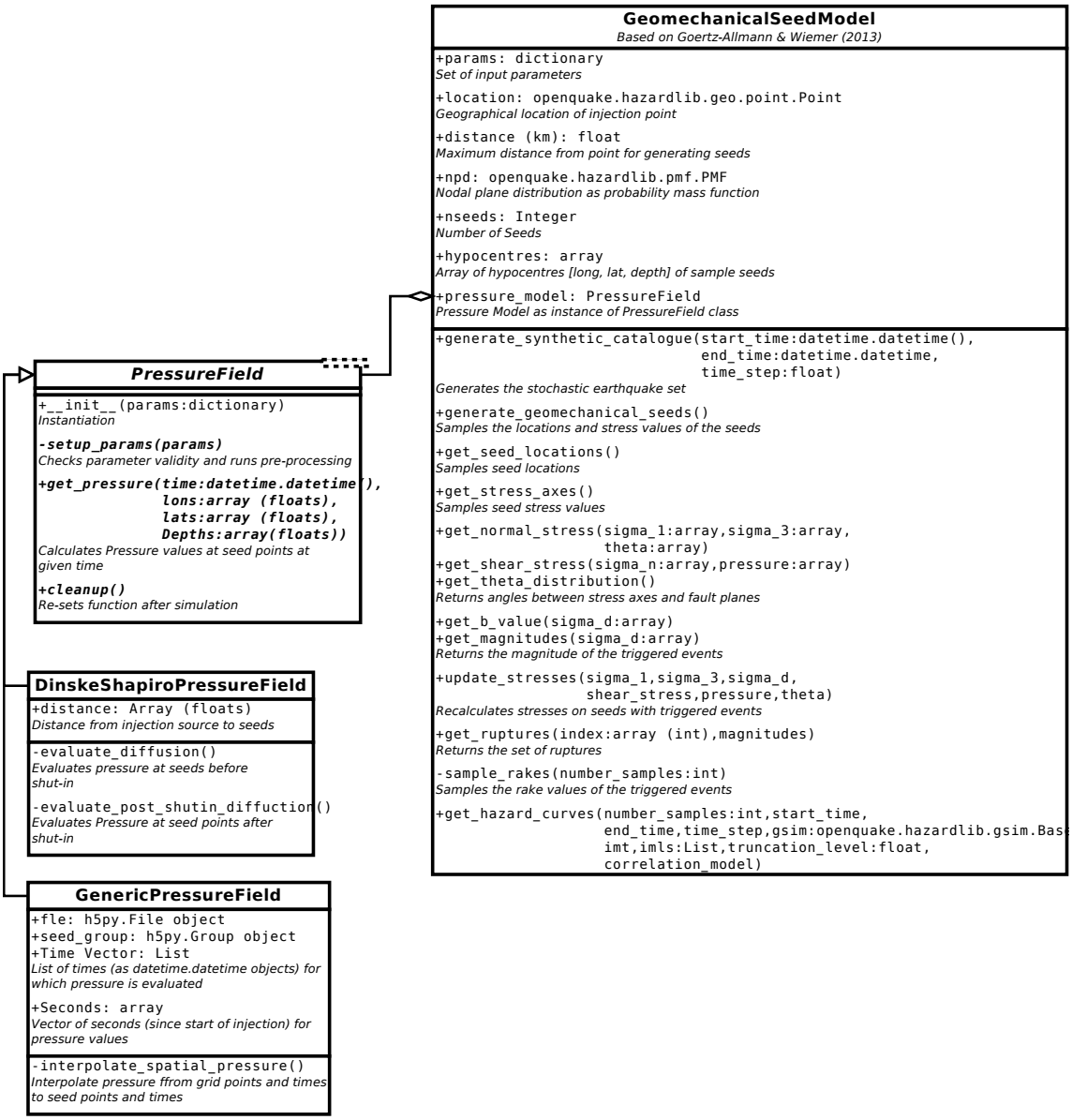

Figure 4-2 UML diagram to indicate the structure and interaction of the two main classes: the pressure field (PressureField) and its sibling classes (DinskeShapiroPressureField and GenericPressureField), and the GeomechanicalSeed class.

4.3.2 The Pressure Field Class

The general operation of any pressure field object is described in an *abstract base class* named `PressureField`. It contains three core methods:

i) *Instantiation (__init__(params))*

Creates an instance of the class and verifies the input parameters conform to the expected (or tolerable) value ranges

ii) *get_pressure(time, lons, lats, depths)*

Evaluates the pore pressure at the given *time*, for a set of geomechanical seed locations defined in terms of longitude, latitude and depth

iii) *cleanup()*

Implements any process to re-set the pressure values once the geomechanical seed simulation is complete. This ensures that when performing multiple simulations the model values of the previous run do not propagate into the new simulation

With the basic definition of the pressure class defined, we use the properties of inheritance to create new objects that can determine the pore-pressure field from a flow model. The first implements the formulation proposed by Dinske & Shapiro (2010). The pore pressure at location distance *r* from the source at time *t*, where *t* is before the shut-in time t_0 is given by:

$$p\left(r,t\right)=\left(\frac{q_0+q_t t}{4\pi Dr}+\frac{q_t r}{8\pi D^2}\right)\cdot\mathrm{erfc}\left(\frac{r}{\sqrt{4Dt}}\right)-\frac{q_t\sqrt{t}}{4\left(\pi D\right)^{3/2}}\cdot\exp\left(\frac{-r^2}{4Dt}\right) \tag{B.40}$$

and for the period after shut-in time (t_0):

$$p\left(r,t\right)=\left(\frac{q_0+q_t t}{4\pi Dr}+\frac{q_t r}{8\pi D^2}\right)\cdot\mathrm{erfc}\left(\frac{r}{\sqrt{4Dt}}\right)-\frac{q_t\sqrt{t}}{4\left(\pi D\right)^{3/2}}\cdot\exp\left(\frac{-r^2}{4Dt}\right)$$
$$-\left[\left(\frac{q_0+q_t\left(t-t_0\right)+q_t t_0}{4\pi Dr}+\frac{q_t r}{8\pi D^2}\right)\cdot\mathrm{erfc}\left(\frac{r}{\sqrt{4D\left(t-t_0\right)}}\right)-\frac{q_t\sqrt{t-t_0}}{4\left(\pi D\right)^{3/2}}\cdot\exp\left(\frac{-r^2}{4D\left(t-t_0\right)}\right)\right] \tag{B.41}$$

where D is the hydraulic diffusivity, $\mathrm{erfc}\left(z\right)$ is the complementary Gaussian error function.

$q_0=4\pi Da_0p_0$ and $q_t=4\pi Da_0p_t$, where p_0 and p_t are the initial pore-pressure and pressure gradient respectively and a_0 the effective injection source radius.

The code subdivides this process into two methods: one for the pre-shut-in period and one for the post-shut-in period. The implementation of the Dinske & Shapiro (2010) model is performed by the class `DinskeShapiroPressureField`, which is a subclass that inherits the properties of `PressureField`. The limitations of using a relatively simplified pressure model for the geomechanical seed model are addressed by Gischig & Wiemer (2013). Instead, they outline the necessity to generate the input pressure field using a nonlinear model of pore-pressure diffusion. To do so, however, they utilise a commercial finite-element software package (COMSOL). It is not currently the intent of this work to implement full finite-element modelling software within the OpenQuake suite, although there may be potential for future integration with open-source finite element packages.

To accommodate the possibility of integrating with more physics-based nonlinear pore-pressure diffusion models a second form of pressure model has been created, and that is the `GenericPressureField` class. To utilise this class it is assumed that the modeller can model the evolution of the pressure field across a regular volumetric grid, at discrete time intervals. This can be undertaken using whichever software, proprietary or open-source, the

modeller chooses. Depending on the density of the volumetric grid, however, the memory size of each field at each time step may become large. To overcome this problem the pressure values must be written to a high-density binary (hdf5) file. The hdf5 format can sustain a considerable degree of nesting in the data structure, whilst being extremely memory-efficient for large arrays of data. The structure of the hdf5 is described as follows:

```
\Nodes
| Locations of Grid Nodes [N, 3] (Longitude, Latitude Depth)
\Pressures
____ YYYY-MM-DD hh:mm:ss.s
  | Pressure Values at Nodes [N, 1]
```

At the top level the locations of the node points of the pressure grid are stored. It is assumed that these remain constant so they do not need to be duplicated on every field. At the same level there is a group named *Pressures*, which contains the data for each time step. The pressure for each time step is stored in a data set, whose name is given by the time-step key. An example of an hdf5 file of data visualised using the HDFView tool (http://www.hdfgroup.org/products/java/hdfview/) can be seen in

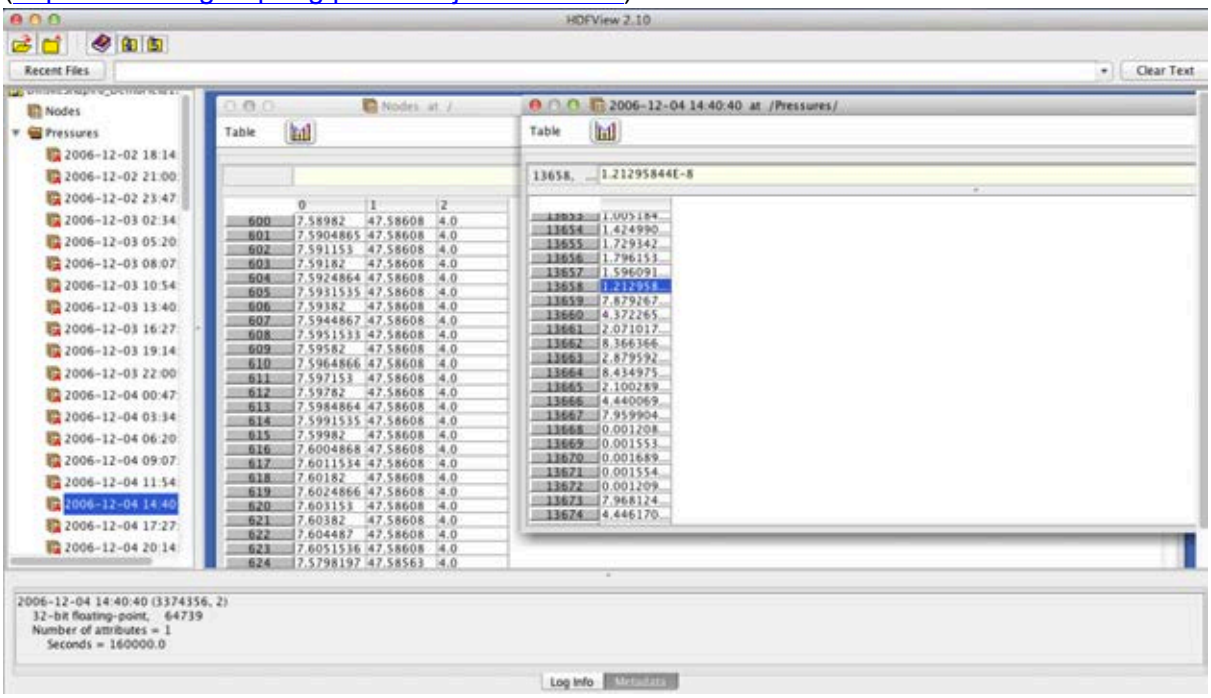

Figure 4-3.

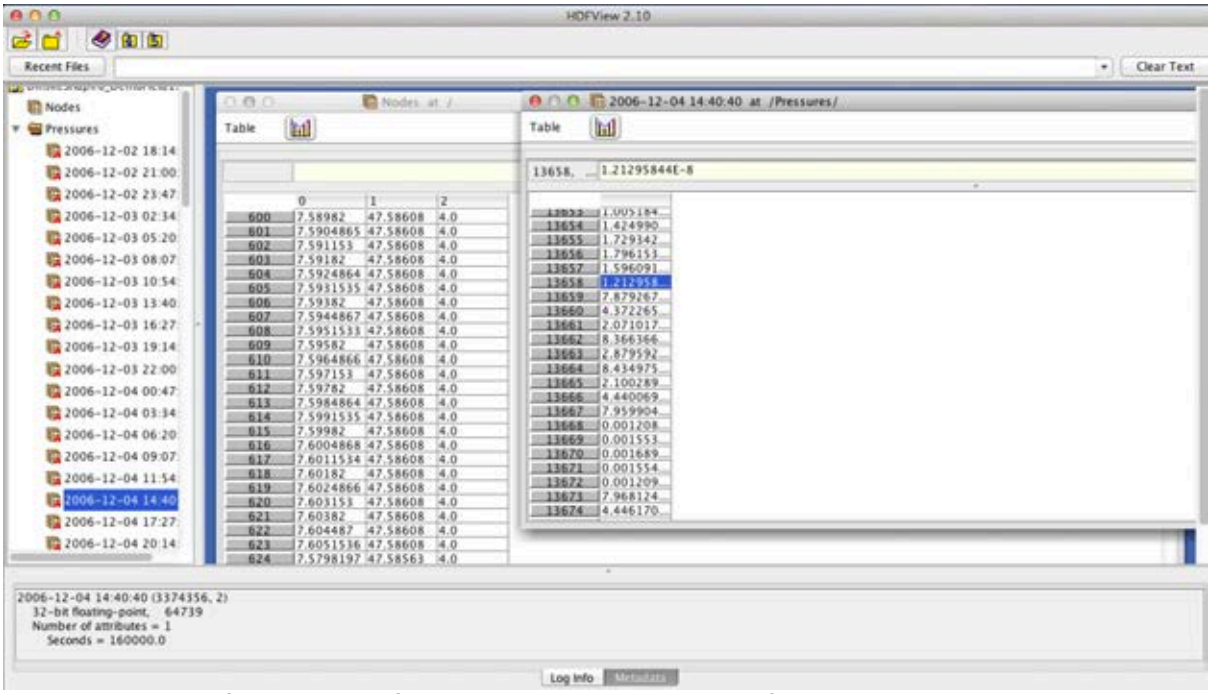

Figure 4-3 HDF5 file structure for a pre-calculated pressure field.

As the pressure values are calculated on a regular grid they will need to be interpolated from the regular grid locations in 3D, to the sampled geomechanical seed location. It is not assumed that the time-step of the pressure files corresponds to that at which the geomechanical seed pressure values are needed. To overcome this a two-step interpolation process is adopted. Once the seeds are initiated the code will apply three-dimensional linear interpolation to determine the pressure values at the seed points for each time-step and store those values in new group inside the hdf5 file. In the second stage, the code will apply simple 1-D linear interpolation to interpolate the pressure values from the initial pressure-field times, to those required for the geomechanical seeds. Once the simulation is completed, the `cleanup()` function will delete the interpolated pressure values at the seed locations from the hdf5 file so that a new simulation process can be started. Naturally the two-step interpolation process incurs an overhead in terms of the computational time, so this process is considerably slower than the use of the Dinske & Shapiro (2010) pressure field.

4.3.3 The Geomechanical Seed Class

The full operation for running the geomechanical seed simulation is managed by the class named `GeomechanicalSeed` (see Figure 4-2). The class is structured to as to break the workflow described in Figure 4-1 into steps such that each step is executed within an individual function. The rationale for doing so is to maximise the flexibility to support modifications to this process where small changes in the modelling assumptions may be desired and thus only minor modifications to individual methods may be needed. Again, this is an example of *abstraction* in the software design.

Depending on the required operation, the main execution of the simulation may be done via one of the following two methods:

```
generate_synthetic_catalogue(start_time, end_time, time_step)
```

This method is the main process by which a geomechanical seed model is generated for a given pressure field evolution and a set of ruptures returned as instances of the OpenQuake Rupture class. This executes the following steps:

1) Generate a set of random seeds by: a) sampling the locations of the seed points from within the three dimensional space centred on the injection point, with distance limits and upper and lower depths specified in the configuration parameters, then b) sampling the maximum principal stress σ_1, minimum principal stress σ_3 and potential fault plane orientation, θ. The mean and standard deviation of the expected principal stress distributions are defined in the parameter inputs, whilst in the current case the angle between fault plane and the principal stress axes is assumed to be random and uniformly distributed. From these values the differential stress is calculated via $\sigma_d = \sigma_1 - \sigma_3$ and the normal stress σ_n via:

$$\sigma_n = \frac{1}{2}\left(\sigma_1 + \sigma_3\right) + \frac{1}{2}\left(\sigma_1 - \sigma_3\right)\cos\left(2\theta\right) \tag{B.42}$$

2) From the stress distributions calculate the corresponding coulomb failure envelope:

$$\tau_s^{CF} = \mu \cdot \sigma_n + c \tag{B.43}$$

where μ is the coefficient of friction and *c* the cohesion (both defined for the specific crustal model.

3) Increment the time counter from the start time by the time step and use the pressure model to calculate the expected pore pressure, *p*, at the geomechanical seed points at the time-step

4) Calculate the shear stress at the seed points from:

$$\tau_s = \mu \cdot \left(\sigma_n - p\right) + c \tag{B.44}$$

5) At the seed points for which $\tau_s \geq \tau_S^{CF}$, events are triggered within this time step. For each of these points calculate the b-value from:

$$b = -0.022\sigma_d + 4.0 \tag{B.45}$$

b-value is constrained within the lower limit of the tectonic *b-value* and an upper-bound b-value (Goertz-Allmann & Wiemer, 2013; Gischig & Wiemer, 2013). Using this given *b-value,* sample the magnitude from the cumulative distribution function of the double-truncated Gutenberg-Richter distribution, with M_{MIN} and M_{MAX} specified in the configuration parameters.

6) For those seeds that have triggered events, reduce the maximum principal stress by an amount equal to 10 % ± 5 % of the differential stress. The resulting difference in τ_S is assumed to be the stress drop of the event (Goertz-Allmann & Wiemer, 2013)

7) Given the seed locations (hypocentres) and magnitudes of the triggered events, create an instance of the OpenQuake “Rupture” class. In this case only the simplest form the of rupture class is required, which takes as input:
    a. Magnitude

b. Rake
c. Tectonic Region Type (defaulted to “Induced” here)
d. Hypocentre
e. Rupture Surface
f. Source Typology

In this case the source is assumed to be a point source; hence this typology is preferred. Likewise the rupture surface is taken as a mesh of one point, from which the distances will be calculated, thus unlike the point source ruptures shown in Figure 3-2, no finite plane is generated and all source to site distances are referred to the epicentre.

One uncertainty remains, however, and that is in the rake value of the source. However, the rake value of a rupture is only used to classify the style-of-faulting. This has an impact only if a finite rupture surface is considered and the magnitude scaling relation has a style-of-faulting requirement, and/or if the GMPE has a style-of-faulting dependency. From the induced seismicity GMPEs listed in section, and recalling from NAM (2013) that one of the key requirements of an induced seismicity GMPE in a context such as this is that coefficients be given for an point source distance metric, it should be evident that rupture finiteness is note considered here, nor is the GMPE dependent on style-of-faulting. In most cases, therefore, the value of rake is irrelevant. Nonetheless the modeller has two options. The rake can be sampled randomly from the uniform distribution $-180 \leq \lambda \leq 180$, or the rake can be sampled from the probability mass function of a nodal plane distribution defined as is typical for a tectonic point or area source.

8) The ruptures of the triggered event are appended to the total event set and the process repeated for the next time step until the end time is reached.

The result of this process is a full stochastic event set for a single realisation of seismicity for the given pressure flow model. This event set is then output to the user, along with a corresponding vector of event times and stress drops.

The second core method of the class is the wrapper method to generate multiple simulations of seismicity via the geomechanical seed model, and for each simulated count the rate of exceedance of a set of ground motion levels in order to retrieve the seismic hazard curve for a set of points.

```
get_hazard_curves(number_samples, start_time, end_time, time_step, gsim, imt, imls, sites, truncation_level, correlation_model)
```

As this method is implementing a full probabilistic seismic hazard calculation more information is needed to describe the ground motion model. Inside of the method for each of the simulations (the number of simulations being set by `number_samples`), the method `generate_synthetic_catalogue()` is called to generate the corresponding rupture set. This requires that the start time, end time and time step of the simulation needs to be specified in exactly the same manner. The remaining attributes describe the ground motion model. `gsim` is the name of the GMPE (from the OpenQuake class list), `imt` the intensity measure type as an instance of either PGA, PGV or SA(T). `imls` is the list of ground motion levels from which the rate (and subsequently the probability) of exceedance is determined. The attribute `"sites"` is an OpenQuake “SiteCollection”, which contains itself a list of OpenQuake “Site” classes. The OpenQuake “Site” class is a holding class that describes the

site with the following parameters: `longitude`, `latitude`, V_{S30}, V_{S30}`_measured` (a Boolean term set to True if the V_{S30} was taken directly from a measurement, or False if it is inferred by a proxy), $Z_{1.0}$ (the depth to the 1.0 km/s shear-wave velocity layer), $Z_{2.5}$ (the depth to the 2.5 km/s shear-wave velocity layer) and a unique site ID.

The remaining two parameters control the `truncation_level`, which determines the truncation of the number of standard deviations of the aleatory variability term of the GMPE, and the `correlation_model`, which chooses the spatial correlation model that can be applied to the ground motion residuals.

4.4 EXAMPLE CALCULATION – BASEL GEOTHERMAL SYSTEM

To illustrate the application of the software to a real case study, we consider the model of the Basel Geothermal System, as provided by Dinske & Shapiro (2010) and Goertz-Allmann & Wiemer (2009). Once again, it is reiterated that the application is purely illustrative, and given the limitations of the model parameters the resulting ground motion exceedance probabilities should not be considered representative of those associated with the actual injection process. The limitations of the linear flow model assume here have already been stated by Gischig & Wiemer (2009), and so require no further discussion here. The parameters used for the simulation are those found in Table 1 of Goertz-Allmann & Wiemer (2013), which for clarity we put into Table 3 here.

The first step is to verify the evolution of the pressure field resulting from the simple fluid injection model. **Error! Reference source not found.** shows the evolution of pressure with distance from the source and time since the start of the injection. As expected, even after shut-in (around 43,200 s) we see the high pressures close to the injection source diminish, whilst still observing a general diffusion to the surrounding medium.

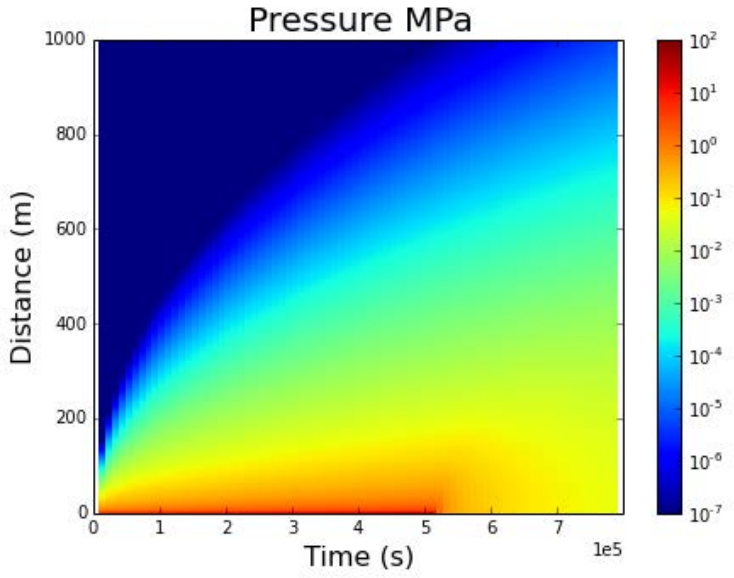

Figure 4-4. Spatio-temporal diffusion of pore-pressure based on the model of Dinske & Shapiro (2010).

Table 3 Model configuration parameters for the Basel example.

Geomechanical Model		Seismological Model	
σ_1	185 MPa	t_{start}	2006/12/2 18:14:0.0
σ_3	75 MPa	t_{end}	2006/12/12 0:27:20.0
$\max(\sigma_1)$	232 MPa	M_{MIN}	0.9
$\mathrm{var}(\sigma_1,\sigma_3)$	0.1	M_{MAX}	6.0
ρ_h	45 MPa	$b(Tectonic)$	1.0
μ	0.85	$b(Upper)$	7.0
N_{SEEDS}	30000	$time-step$	10^4 s
c	185 MPa	Depth_{UPPER}	3 km
a_0	1.0 m	Depth_{LOWER}	7 km
p_0	9.5 MPa		
p_t	48 Pa/s		
t_0	2006/12/8 17:48:0.00		
$D(m^2/s)$	0.05		

For a single simulation of seismicity we can observe the temporal and spatial evolution of the seismicity cloud in Figure 4-5. In this particular simulation the largest event has a magnitude of approximately M_W 3.5, slightly larger than the largest magnitude observed during the actual Basel the sequence.

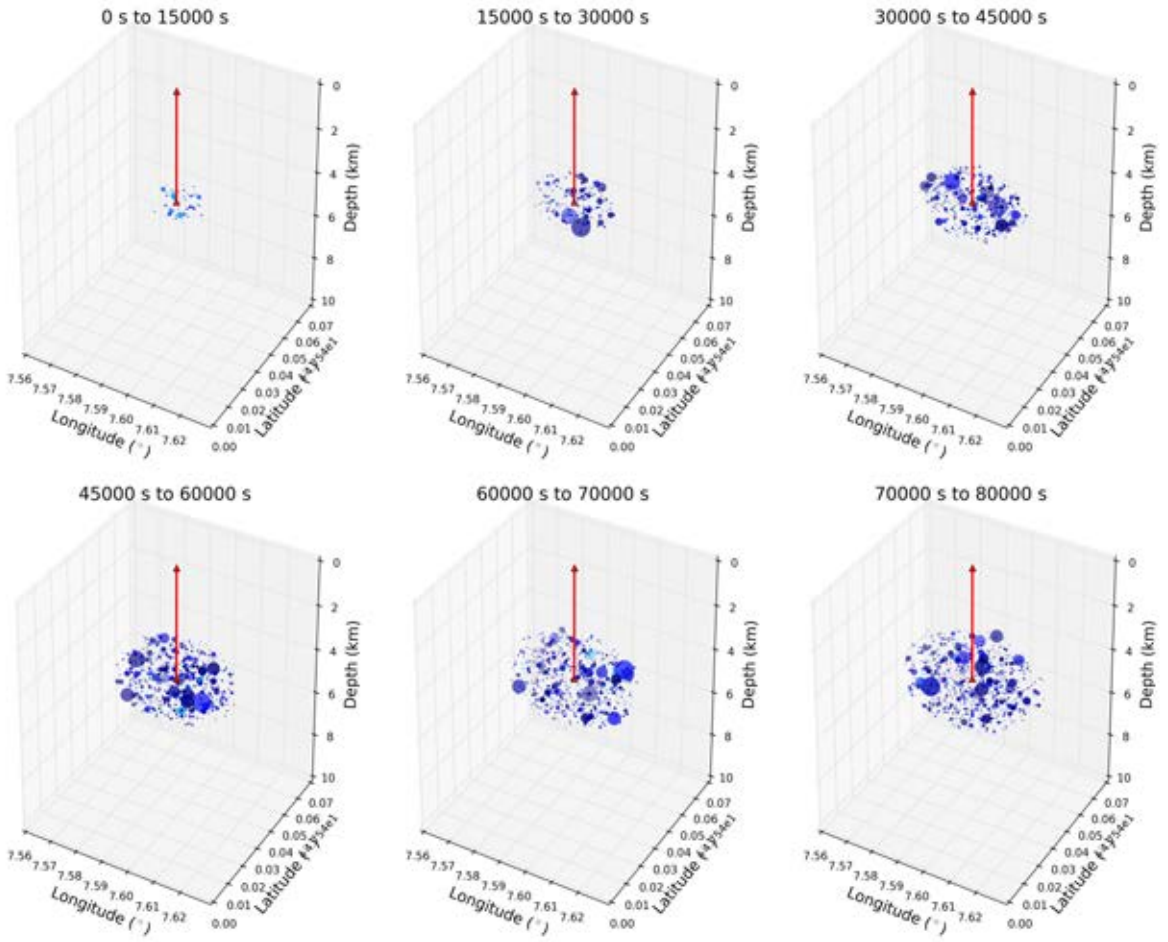

Figure 4-5 Spatio-temporal evolution of hypocentral location during the evolution of the injection sequence. Hypocentres are scaled by magnitude size

To generate a full probabilistic seismic hazard analysis taking into account the aleatory variability, a set of 1000 simulations are run. On each simulation the event set is generated at the expected ground motions calculated at a site (or in this case a grid of sites). The GMPE of Douglas et al. (2013), with $\Delta\sigma$ = 100 bar, Q = 600 and κ =0.04 s is used for calculation of the ground motion. The resulting seismic hazard maps for the 10%, 5% and 2% probabilities of exceedance are shown in Figure 4-6. In the near vicinity of the injection site the peak ground accelerations predicted by the model are surprisingly high; as much as 0.2 g for the 5% probability of being exceeded, and increasing closes to 0.3 g for the 1% probability. This may be largely attributable to the scaling of the GMPE, which predicts median accelerations on the order of approximately 0.1 g at short distances of three to four kilometres.

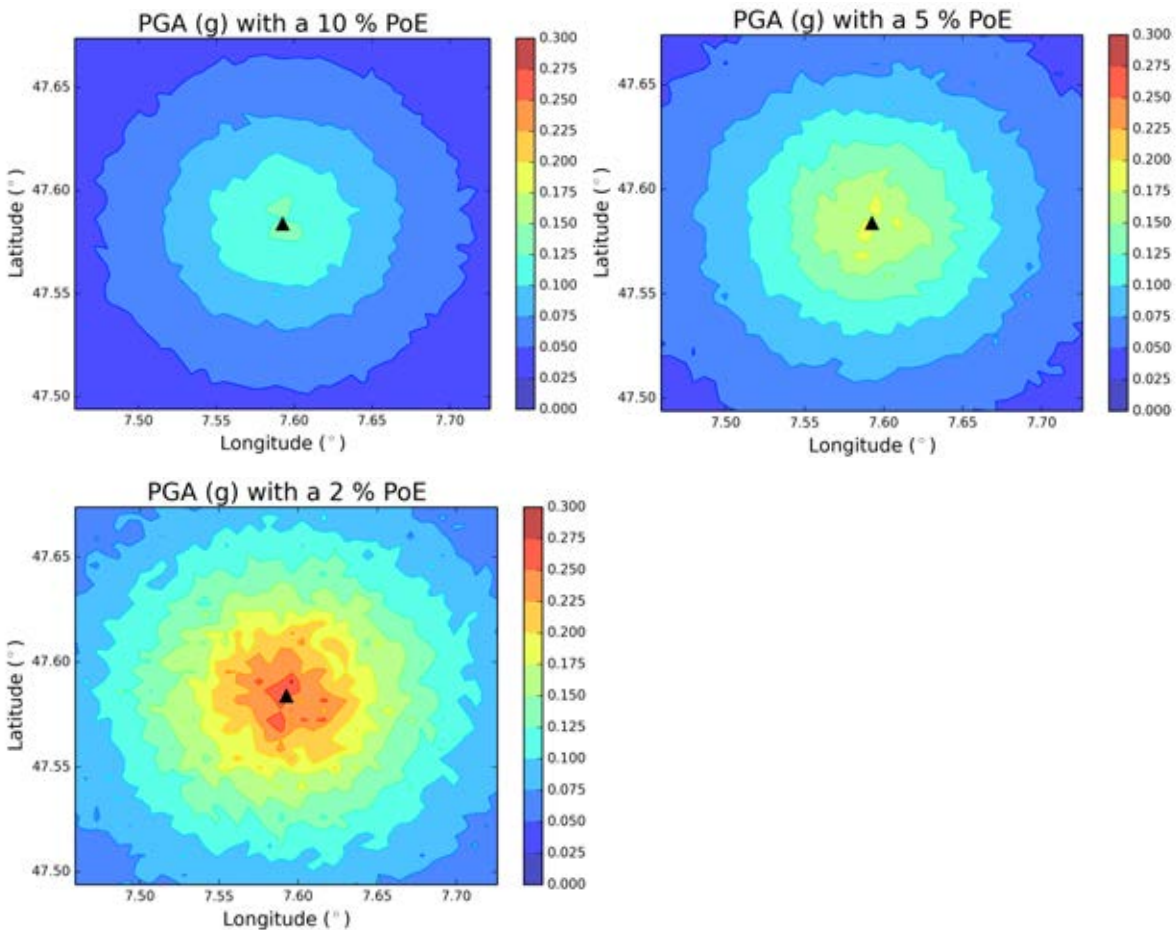

Figure 4-6. PGA with a 10% (*top left*), 5% (*top right*) and 2% (*bottom*) probability of being exceeded during the injection sequence.

Even if the geomechanical model is fixed, the choice of the seismological and hazard configuration may still have a significant impact on the seismic hazard curves. In practice this should be undertaken using a full analysis of epistemic uncertainty. In Figure 4-7, a small sensitivity analysis is performed for a site located 500 m from the top of the injection point. Here, four different modelling assumptions are compared: stress drop (using the Douglas et al. (2013) GMPE with $\Delta\sigma = 10$ bar and $\Delta\sigma = 100$ bar), and maximum magnitude (M_W 5.0 and M_W 6.0). It is remarkable in this case that even close to the site the impact of the maximum magnitude, whilst not trivial, is significantly smaller than that of the stress drop. As both GMPEs have the same aleatory uncertainty term, it is clear this different is attributable to the median ground motion, which as seen in Figure 2-8 produces PGA values almost an order of magnitude higher in the near-field region ($R_{HYP} \leq 5$ km) in the case of $\Delta\sigma = 100$ bar than for $\Delta\sigma = 10$ bar.

One clarification to be made here is that the stress drop defined for the GMPE has a different meaning from that of the calculated stress-drops in the geomechanical seed model. The GMPE stress-drop may be considered more in the context of a regional property, and not necessarily linked to the simulated stress drops of the seed. The former, then, renders stress-drop an epistemic uncertaintly, the latter is aleatory within this process. Seeking consistency between the two definitions is more challenging, and would require further investigation to determine whether the two distinct interpretations can be reconciled.

The sensitivity of the hazard to such parameters makes it clear that in any real application there is a need to incorporate analyses of epistemic uncertainty into the calculation, as highlighted by Mignan et al. (2015). However, initial sensitivity studies such as this also demonstrate the need to establish in which parts of the model it is necessary to focus on uncertainty reduction, and to identify the most realistic uncertainty model for the parameters. In this particular case, it may not be appropriate to exhaust all of the possible combinations of the Douglas et al. (2013) model, as done by Mignan et al. (2015). Many configurations can, and should, be rejected at the outset based on the observed seismological properties of the region. Instead, it may be more desirable to either select a subset of the models based on observed strong or weak motions, such as in the case of Edwards and Douglas (2013). Or, alternatively, existing GMPEs should be re-calibrated so as to capture the centre and range of the likely variation in stress drop; the latter approach corresponding more closely to the proposed for general application in PSHA by Atkinson et al. (2014).

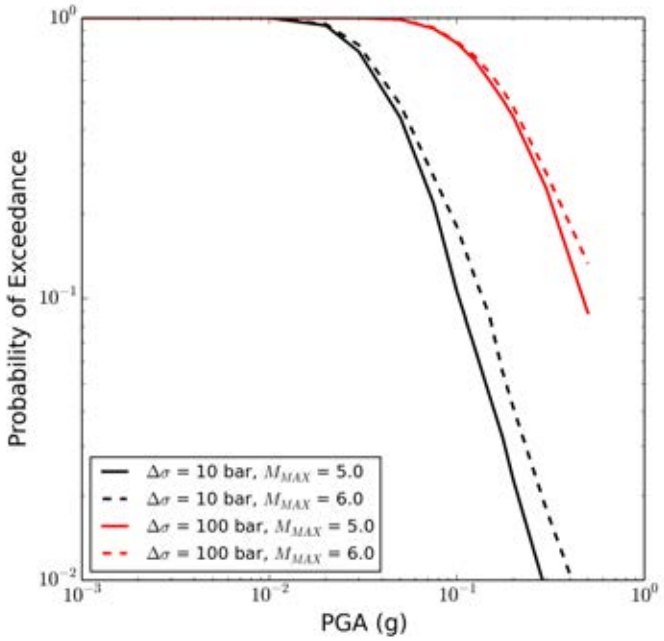

Figure 4-7. Seismic hazard curves for a site located 500 m from the site of the injection well, for different combinations of M_{MAX} and stress drop (using the Douglas et al. (2013) GMPE).

The combination of geomechanical model and PSHA software shown here is intended to be just an illustrative application. Certainly further analysis could be performed to explore the sensitivity of the geomechanical modelling assumptions on the resulting seismic hazard analysis, and indeed this should be an objective of any practical application of the software. More importantly, however, this case demonstrates the relative ease with which the physical modelling can be combined with PSHA in a single framework. Whilst not in itself a general package for induced seismicity, many elements of this process can be adopted in other potential applications, including some of those discussed in section 2 of this report.

5 Conclusions and Future Directions

Within the current report it has been shown how existing software for PSHA can be adapted for an induced seismicity application. In the example shown here, it is the geomechanical seed model approach that is illustrated, albeit other approaches from those listed in section 2 might also have been considered. The geomechanical seed approach highlights one possible well-developed application, but there are elements within this process that may be transferable to other forms of induced seismicity hazard. The most relevant of these elements is the use of Monte Carlo simulations, to generate multiple realisations of seismicity that can incorporate the spatio-temporal evolution of the physical process explicitly into the PSHA. That a similar Monte Carlo based approach is also implemented for the purpose of modelling the induced seismicity hazard in response to conventional gas extraction scenarios in the Groningen gas field (Bourne et al., 2014; 2015) is not at all coincidental.

From a practical perspective, the intent of this study has not necessarily been to incorporate an “induced seismicity package” into OpenQuake, but instead to demonstrate how OpenQuake can be integrated into a more directed application for the problem at hand. Modelling of seismic hazard from induced seismicity cannot be viewed as a single methodology, but one that must be entirely adaptable to the physics of the anthropogenic seismicity. The incorporation of OpenQuake into induced seismicity software has also successfully illustrated many of the reasons why OpenQuake may be a valuable tool for scientific research and industry. Its open-source nature and architecture make it ideal for just such an application, which, with the exception of the addition of new GMPEs, required very few modifications for such a novel applications. The methods related purely to the seismic hazard application can be taken and applied in the form of a library, and can be adapted for the context. This report has demonstrated, albeit in a simplified manner, precisely how this development process can work. Few other open-source codes can boast such flexibility nor prove adaptable to short-term time-dependent models without extensive modification. The addition of induced-seismicity specific GMPEs consolidates the potential for its use in further induced seismicity applications. Anticipated future developments of OpenQuake will likely provide further flexibility in the applications of the GMPEs by allowing direct separation and modification of the mean and aleatory uncertainty elements of the GMPE.

In the absence of application-specific software features requirements, the potential directions of development of such an application are difficult to identify. The simple structure of the geomechanical seed model described in section 4 can be enhanced, particularly with respect to the processing of epistemic uncertainties. This would, of course, require further optimisations to the efficiency of the code, which would naturally be an objective of any software intended for production. Further optimisations, particularly those that can take advantage of OpenQuake’s full parallelisation capabilities, would likely be a high priority

From a scientific perspective, there are several directions that could be taken to maximum the versatility of the approach to consider different types of induced seismicity. Once again, the geomechanical seed approach may have potential for application to many different cases of anthropogenically induced earthquake hazard. The manner in which the

geomechanical evolution drives the triggering of earthquake on stochastically generated seed points may not necessarily be limited in context solely to application for an enhanced geothermal system. Indeed, one of the most significant advantages of this approach is that both the seismicity simulations, and the resulting hazard, do not have a direct link to the stimulation process, but are themselves response to the changes in pressure. It is conceivable that the same hazard modelling process can be applied for other cases such as natural gas and oil production, ground water modulation (extraction or even reservoir impoundment) and potentially even mining. The natural link therefore might be to consider the construction of open-source software that can integrate OpenQuake with other open-source software for modelling thermo-hydromechanical processes via a finite-element approach. Such software can help greatly bridge the gap between the physical modelling of the induced seismicity source itself whilst retaining the capability to incorporate both the aleatory and epistemic uncertainty of the model into the seismic hazard process.

References

Assatourians, K., Atkinson, G.M., 2013. EqHaz: An open-source probabilistic seismic-hazard code based on the Monte Carlo simulation approach. *Seismological Research Letters* 84: 516-524.

Atkinson, G. M., Bommer, J. J., Abrahamson, N. A., 2014. Alternative Approaches to Modeling Epistemic Uncertainty in Ground Motions in Probabilistic Seismic-Hazard Analysis. *Seismological Research Letters* 85(6): 1141-1144.

Atkinson, G.M., 2015. Ground-Motion Prediction Equation for Small-to-Moderate Events at Short Hypocentral Distances, with Application to Induced-Seismicity Hazards. *Bulletin of the Seismological Society of America,* in press. doi:10.1785/0120140142

Bachmann, C.E., Wiemer, S., Woessner, J., Hainzl, S., 2011. Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems: Probability-based monitoring approach for EGS. *Geophysical Journal International* 186: 793–807. doi:10.1111/j.1365-246X.2011.05068.x

Bazzurro, P., Cornell, C.A., 1999. Disaggregation of seismic hazard. *Bulletin of the Seismological Society of America* 89: 501–520.

Bommer, J.J., Akkar, S., 2012. Consistent source-to-site distance metrics in ground-motion prediction equations and seismic source models for PSHA. *Earthquake Spectra* 28: 1–15.

Bommer, J.J., Crowley, H., Pinho, R., 2015. A risk-mitigation approach to the management of induced seismicity. *Journal of Seismology* 19: 623–646. doi:10.1007/s10950-015-9478-z

Bommer, J.J., Oates, S., Cepeda, J.M., Lindholm, C., Bird, J., Torres, R., Marroquín, G., Rivas, J., 2006. Control of hazard due to seismicity induced by a hot fractured rock geothermal project. *Engineering Geology* 83: 287–306. doi:10.1016/j.enggeo.2005.11.002

Bommer, J.J., Stafford, P.J., Alarcon, J.E., Akkar, S., 2007. The Influence of Magnitude Range on Empirical Ground-Motion Prediction. *Bulletin of the Seismological Society of America* 97: 2152–2170. doi:10.1785/0120070081

Bourne, S.J., Oates, S.J., Bommer, J.J., Dost, B., van Elk, J., Doornhof, D., 2015. A Monte Carlo method for probabilistic hazard assessment of induced seismicity due to conventional natural gas production. *Bulletin of the Seismological Society of America,* in press.

Bourne, S.J., Oates, S.J., van Elk, J., Doornhof, D., 2014. A seismological model for earthquakes induced by fluid extraction from a subsurface reservoir. *Journal of Geophysical Research: Solid Earth* 119: 8991–9015

Campbell, K.W., 2003. Prediction of strong ground motion using the hybrid empirical method and its use in the development of ground-motion (attenuation) relations in eastern North America. *Bulletin of the Seismological Society of America* 93: 1012–1033.

Catalli, F., Meier, M.-A., Wiemer, S., 2013. Coulomb stress changes at the Basel geothermal site: can the Coulomb model explain induced seismicity. *Geophysical Research Letters* 40: doi:10.1029/2012GL054147.

Chiou, B., Youngs, R., Abrahamson, N., Addo, K., 2010. Ground-Motion Attenuation Model for Small-To-Moderate Shallow Crustal Earthquakes in California and Its Implications on Regionalization of Ground-Motion Prediction Models. *Earthquake Spectra* 26: 907–926. doi:10.1193/1.3479930

Convertito, V., Maercklin, N., Sharma, N., Zollo, A., 2012. From Induced Seismicity to Direct Time-Dependent Seismic Hazard. *Bulletin of the Seismological Society of America* 102: 2563–2573. doi:10.1785/0120120036

Cornell, C.A., 1968. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America* 58: 1583-1606.

Cotton, F., Scherbaum, F., Bommer, J.J., Bungum, H., 2006. Criteria for Selecting and Adjusting Ground-Motion Models for Specific Target Regions: Application to Central Europe and Rock Sites. *Journal of Seismology* 10: 137–156: doi:10.1007/s10950-005-9006-7

Dinske, C., Shapiro, S.A., 2010. Analysis of Seismicity Resulting from Time-dependent Fluid Injection Source Pressures, in: *Proceedings of the 72nd EAGE Conference & Exhibition Incorporating SPE EUROPEC*. Barcelona, Spain, 14 - 17 June 2010, p. F024.

Dost, B., Caccavale, M., Van Eck, T., Kraaijpoel, D., 2013. Report on the expected PGV and PGA values for induced earthquakes in the Groningen area (KNMI Report). KNMI.

Dost, B., Van Eck, T., Haak, H., 2004. Scaling of peak ground acceleration and peak ground velocity recorded in the Netherlands. *Bollettino di Geofisica Teorica ed Appliocata* 45: 153–168.

Douglas, J., Edwards, B., Convertito, V., Sharma, N., Tramelli, A., Kraaijpoel, D., Cabrera, B.M., Maercklin, N., Troise, C., 2013. Predicting Ground Motion from Induced Earthquakes in Geothermal Areas. *Bulletin of the Seismological Society of America* 103: 1875–1897. doi:10.1785/0120120197

Ebel, J.E., Kafka, A.L., 1999. A Monte Carlo approach to seismic hazard analysis. *Bulletin of the Seismological Society of America* 89: 854–866.

Edwards, B., Douglas, J., 2013. Selecting ground-motion models developed for induced seismicity in geothermal areas. *Geophysical Journal International* 195(2): 1314–1322.

Field, E.H., Jordan, T.H., Cornell, C.A., 2003. OpenSHA: A developing community-modeling environment for seismic hazard analysis. *Seismological Research Letters* 74: 406–419.

Gischig, V.S., Wiemer, S., 2013. A stochastic model for induced seismicity based on non-linear pressure diffusion and irreversible permeability enhancement. *Geophysical Journal International* 194: 1229–1249. doi:10.1093/gji/ggt164

Goertz-Allmann, B.P., Wiemer, S., 2013. Geomechanical modelling of induced seismicity source parameters and implications for seismic hazard assessment. *Geophysics* 78: KS25-KS39.

Hanks, T.C., Bakun, W.H., 2008. M-logA Observations for Recent Large Earthquakes. *Bulletin of the Seismological Society of America* 98: 490–494. doi:10.1785/0120070174

Jayaram, N., Baker, J.W., 2009. Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering & Structural Dynamics* 38: 1687–1708. doi:10.1002/eqe.922

Johnston, A.C., Coppersmith, K.J., Kanter, L., Cornell, C. A., 1994. The Earthquakes of Stable Continental Regions. Volume 1: Assessment of Large Earthquake Potential (EPRI Research Report No. TR-102261-V1).

Kijko, A., 2004. Estimation of the Maximum Earthquake Magnitude, M_{MAX}. *Pure and Applied Geophysics* 161: 1655–1681. doi:10.1007/s00024-004-2531-4

Király, E., Gischig, V., Karvounis, D., Wiemer, S., 2014. Validating Models to Forecasting Induced Seismicity Related to Deep Geothermal Energy Projects, in: *Proceedings, 39th Workshop on Geothermal Reservoir Engineering,* Stanford University, Stanford, CA.

Király, E., Zechar, J.D., Gischig, V., Karvounis, D., Heiniger, L., Wiemer, S., 2015. Modeling and Forecasting Induced Seismicity in Deep Geothermal Energy Projects, in: *Proceedings of the World Geothermal Congress.* Melbourne, Australia.

Kostrov, V.V., 1974. Seismic moment and energy of earthquakes, and seismic flow of rocks. *Izv. Acad. Sci. USSR Physics of the Solid Earth* (English Translation) 1: 23 – 44.

Langenbruch, C., Shapiro, S.A., 2010. Decay rate of fluid-induced seismicity after termination of reservoir stimulations. *Geophysics* 75: MA53–MA62. doi:10.1190/1.3506005

Lasocki, S., 2005. Probabilistic analysis of seismic hazard posed by mining induced events, in: *Proceedings of the 6th International Symposium on Rockburst in Mines "Controlling Seismic Risk*". ACG, Perth. 151–156.

Makropoulos, K.C., Burton, P.W., 1983. Seismic risk of circum-Pacific earthquakes I. Strain energy release. *Pure and Applied Geophysics* 121: 247–267.

Matthews, M.V., Ellsworth, W.L., Reasenberg, P.A., 2002. A Brownian model for recurrent earthquakes*. Bulletin of the Seismological Society of America* 92: 2233–2250.

Mena, B., Wiemer, S., Bachmann, C., 2013. Building Robust Models to Forecast the Induced Seismicity Related to Geothermal Reservoir Enhancement. *Bulletin of the Seismological Society of America* 103: 383–393: doi:10.1785/0120120102

Mignan, A., Landtwing, D., Kästli, P., Mena, B., Wiemer, S., 2015. Induced seismicity risk analysis of the 2006 Basel, Switzerland, Enhanced Geothermal System project: Influence of uncertainties on risk mitigation. *Geothermics* 53: 133–146. doi:10.1016/j.geothermics.2014.05.007

Musson, R.M.W., 2012. PSHA Validated by Quasi Observational Means. *Seismological Research Letters* 83: 130–134. doi:10.1785/gssrl.83.1.130

Musson, R.M.W., 1999. The use of Monte Carlo simulations for seismic hazard assessment in the U. K. *Annali di Geofisica* 43: 1–9.

Nederlandse Aardolie Maatschappig B. V. (NAM), 2013. Technical Addendum to the Winningsplan Groningen 2013. Subsidence, Induced Earthquakes and Seismic Hazard Analysis in the Groningen Field.

Ogata, Y., 1999. Seismicity analysis through point-process modeling: A review. *Pure and Applied Geophysics* 155: 471–507.

Pagani, M., Monelli, D., Weatherill, G.A., Garcia, J., 2014a. The OpenQuake-engine Book: Hazard (Global Earthquake Model (GEM) Technical Report). Global Earthquake Model (GEM).

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., Simionato, M., Vigano, D., 2014b. OpenQuake Engine: An Open Hazard (and Risk) Software for the Global Earthquake Model. *Seismological Research Letters* 85: 692–702. doi:10.1785/0220130087

Park, J., Bazzurro, P., Baker, J.W., 2007. Modeling spatial correlation of ground motion Intensity Measures for regional seismic hazard and portfolio loss estimation, in: *Applications of Statistics and Probability in Civil Engineering*. Taylor & Francis.

Petersen, M.D., Cao, T., Campbell, K.W., Frankel, A.D., 2007. Time-independent and time-dependent seismic hazard assessment for the State of California: Uniform California Earthquake Rupture Forecast Model 1.0. *Seismological Research Letters* 78: 99–109.

Reasenberg, P., Jones, L., 1989. Earthquake hazard after a mainshock in California. *Science* 243: 1173–1176.

Shapiro, S.A., Dinske, C., Langenbruch, C., Wenzel, F., 2010. Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations. *The Leading Edge* 29: 304–309.

Utsu, T., 1961. A statistical study on the occurrence of aftershocks. *Geophysical Magazine* 30: 521–605.

Van Eck, T., Goutbeek, F., Haak, H., Dost, B., 2006. Seismic hazard due to small-magnitude, shallow-source, induced earthquakes in The Netherlands. *Engineering Geology* 87: 105–121. doi:10.1016/j.enggeo.2006.06.005

Weatherill, G.A., 2014a. OpenQuake Ground Motion Toolkit - User Guide (Global Earthquake Model (GEM) Technical Report). Global Earthquake Model (GEM).

Weatherill, G.A., 2014b. OpenQuake Hazard Modeller's Toolkit - User Guide (Global Earthquake Model (GEM) Technical Report). Global Earthquake Model (GEM).

Weatherill, G., Burton, P.W., 2010. An alternative approach to probabilistic seismic hazard analysis in the Aegean region using Monte Carlo simulation. *Tectonophysics* 492: 253–278. doi:10.1016/j.tecto.2010.06.022

Wells, D.L., Coppersmith, K.J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America* 84: 974–1002.

Youngs, R.R., Coppersmith, K.J., 1985. Implications of Fault Slip Rates and Earthquake Recurrence Models to Probabilistic Seismic Hazard Estimates. *Bulletin of the Seismological Society of America* 75: 939–964.