Modeling seismic activity on faults for probabilistic hazard assessment

Thomas Chartier

PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Ecole Normale Supérieure

Modélisation de l'activité sismique des failles pour le calcul probabiliste du risque sismique

Soutenue par
Thomas Chartier
Le 31 Octobre 2019

Ecole doctorale n° 560
Sciences de la Terre et de l'Environnement et de Physique de l'Univers, Paris

Spécialité
Terre Environnement

Composition du jury :

Stéphane, Mazzotti Pr, Université Montpellier 2	*Président du jury*
Kuo-Fong, Ma Pr, National Central University	*Rapportrice*
Alice-Agnes, Gabriel Assistant Pr, Munich University	*Examinatrice*
Nicolas, Chamot-Rooke CR, Ecole Normale Supérieure	*Examinateur*
Hélène, Lyon-Caen Dr, Ecole Normale Supérieure	*Directrice de thèse*
Oona, Scotti Dr, Institut de Radioprotection et de Sûreté Nucléaire	*Co-Encadrante de thèse*
Aurélien, Boiselet Dr, Axa	*Invité*

ENS ÉCOLE NORMALE SUPÉRIEURE
PSL RESEARCH UNIVERSITY PARIS
IRSN INSTITUT DE RADIOPROTECTION ET DE SÛRETÉ NUCLÉAIRE
STEP'UP ÉCOLE DOCTORALE
AXA Research Fund
cnrs

Résumé

Les taux de sismicité sont un élément clé du calcul probabiliste de l'aléa et du risque sismique. Une méthodologie innovante est développée pour modéliser les taux de ruptures complexes dans un système de failles (SHERIFS). La flexibilité de la méthodologie permet de la partager et de l'appliquer sur des systèmes de failles différents et de s'en servir comme outil de discussion des hypothèses concernant la sismicité sur les failles. Les taux de sismicité dans la partie Ouest du Rift de Corinthe en Grèce et dans la région de Marmara en Turquie sont modélisés en explorant l'incertitude épistémique sur le scenario de rupture maximale, la distribution en magnitude et fréquence ou encore les conditions de glissement sur la faille. A Marmara, les hypothèses sont pondérées dans un arbre logique en comparant les taux modélisés aux taux calculés à partir des données et des résultats d'un modèle physique basé sur les équations "rate and state". Pour chaque hypothèse de l'arbre logique, le risque sismique à Istanbul, en termes de probabilité d'effondrement d'un immeuble, est calculé pour deux bâtiments de même type construits respectivement suivant le code de construction de 1975 et celui de 1998. Le risque sismique est six fois plus important pour le bâtiment le plus ancien. Parmi les incertitudes explorées, la source d'incertitude plus importante est liée à la distribution en magnitude et fréquence. L'utilisation des données et du modèle physique permet de réduire l'incertitude sur le risque par un facteur 1.6. Une nouvelle méthodologie de désagrégation du risque permet de montrer que le risque à Istanbul est contrôlé en partie par les séismes de magnitude supérieure à 7 sur la Faille Nord Anatolienne et en partie par les séismes de magnitude modérée (4.5 à 6) dans la sismicité de fond, sur des failles non connues, et à une distance du bâtiment inférieure à 10 km.

Abstract

Earthquake rates are a key component of probabilistic seismic hazard and risk assessment. A novel methodology to model earthquake rates in complex ruptures in a fault system (SHERIFS) is developed. The flexibility of the methodology allows sharing it and applying it to various fault systems as well as using it as a tool to discuss the hypotheses concerning the seismicity on faults. The earthquake rates in the Western Corinth Rift in Greece and the Marmara Region in Turkey are modeled while exploring the associated epistemic uncertainty such as on the maximum rupture scenario, the magnitude frequency distribution or the locking condition of the faults. In Marmara, the hypotheses are weighted in a logic tree by comparing them to the rate calculated from the data (earthquake catalog and paleoseismicity) and the results of a "physics-based" model applying the rate and state equations. For each hypothesis in the logic tree, The seismic risk in Istanbul, represented by the probability of collapse of a building, is calculated for two buildings of the same type but one constructed following the 1975 building code and on following the 1998 building code. The seismic risk is six times larger for the older building. Within the explored uncertainties, the one concerning the magnitude frequency distribution is the largest source of uncertainty on the risk of collapse. By using the data and the physics-based model to weight the logic tree, the uncertainty on the risk is reduced by a factor of 1.6. Through a novel risk deaggregation methodology, we identify two seismogenic sources controlling the risk in Istanbul: the large earthquakes (magnitude larger than 7) on the neighboring North Anatolian Fault and the moderate magnitude earthquakes (between 4.5 and 6) occurring in the background zone, on unknown faults, and at a distance shorter than 10 km from the building.

Remerciements / Acknowledgements

This thesis that you are about to read largely exists because of the continuous support of my family, friends and colleague without whom I would never have been able to carry on such endeavor.

En premier lieu, je souhaite remercier mes deux directrices de thèse, Hélène et Oona, qui m'ont permis de me lancer cette thèse. Grâce à leur soutien mais aussi la liberté qu'elles m'ont autorisées dans la mise en place du projet, le sujet de cette thèse correspond parfaitement à mes ambitions de carrière et à mes intérêts scientifiques. Peu d'étudiants en thèse ont la chance d'avoir un sujet de thèse taillé sur mesure. Au cours de la thèse elles m'ont offert un encadrement avec à la fois une grande disponibilité pour la discussion et une grande liberté en me laissant m'engager dans différentes voie de développement scientifique. Elles ont sut rester ouvertes à mes idées mais aussi m'orienter lorsque m'ont entêtement naturel me bloquait dans une seule direction. J'ai pût énormément apprendre de leur rigueur scientifique tout en évoluant dans un environnement amical rendant chaque journée de travail un vrai plaisir. J'espère pouvoir continuer à interagir avec elles, à la fois professionnellement mais aussi personnellement.

Merci aux institutions qui ont supporté cette thèse, l'Ecole Normale Supérieure pour m'avoir permis de partir travailler en Californie, l'Institut de Radioprotection et de Sûreté Nucléaire pour un soutien continu depuis plus de cinq ans et le Fond Axa pour la Recherche pour avoir financer ce projet (merci beaucoup à Aurélien pour ton aide dès la mise en place du projet).

A tous mes collègues, maintenant devenu amis, je dis merci. A ceux qui ont partagé la vue et les plantes du Bureau Banane à l'IRSN : Ludmila, Elif, Arnaud, Flomin, Levent et Rihab. Aux amis du bureau de l'ENS : Lavinia, Arefeh, Eva, Cédric, Eleonora. A toute l'équipe BERSSIN et du laboratoire de géologie, merci! Avec de telles équipes, il a été facile de se motiver tous les jours pour aller travailler sachant que l'entre-aide et l'amitié sera au rendez-vous.

A la ville de Paris, qui est peu à peu devenu ma ville au fil des années. Les longues marches dans ses rues et dans ses tunnels vont beaucoup me manquer.

A tout mes amis, de la surface et des souterrains, Julia, Arnaud, Julien, Martin, PO, Arthur, Simon, Anne, Sonia, Célia, Justine, Lucie, Mylène et tant d'autres qui sont toujours près à aller manger un bout ou boire un coup, et le plus souvent les deux!

Enfin, je voudrais dédier cette thèse à ma grande et belle famille! En particulier à mes parents, qui m'ont toujours encouragé à faire ce qui me faisait plaisir sans pression sur mes résultats et avec toujours un enthousiasme contagieux. A ma tante Denise qui a été depuis toujours et est toujours ma principale interlocutrice pour les discussions scientifiques et qui de plus a été un soutien à la fois logistique et moral pendant toutes mes études! A mon frère et ma soeur, mes tantes et mes oncles, et mes très nombreux cousins et cousines dont l'amour constant a forgé la personne que je suis aujourd'hui.

Contents

List of Figures

List of Tables

Glossaire - Glossary

- **Background Seismicity** - Seismicity that occurs on faults that are not considered in the model.
- **Fault geometries** Note : The following terminology is not a general terminology use across all fields of geoscience but rather in the seismic hazard assessment community and in particular in this study.
  - **Fault system** - Ensemble of faults that are in the same region and are part of the same geodynamic unit.
  - **Fault segment** - Portion of fault defined by the boundaries that are considered to be the limits of the possible ruptures on this fault.
  - **Fault section** - Part of the fault for which a set of paremeter is attributed (slip-rate, dip, ...) but can rupture with neighbouring fault sections.
- **FC** - Fragility Curve - Gives the probability of building dammage for a given ground motion.
- **GMPE** - Ground Motion Prediction Equation - Equation de Prediction du Mouvement Sismique
- **HC** - Hazard Curve - for a range of ground motions, probability of exceeding given ground motion at one site over a period of time.
- **MFD** - Magnitude Frequency Distribution - Distribution en Magnitude et Fréquence
  - **GR** Gutenberg Richter distribution (Gutenberg and Richter [1954]).
  - **TS** Tuned Shape distribution. Tuned to the shape of the distribution of the catalog (see Chapter 2).
- **NAFS** - North Anatolian Fault System - Système de Failles Nord-Anatolien
- **NMS** - Non Main-Shock slip - See Chapter 1
- **Participation rate** - Annual rate of earthquakes rupturing a given fault section. The ruptures can involve other fault sections.
- **PGA** - Peak Ground Acceleration
- **PGV** - Peak Ground Velocity

- **PSA** - Pseudo-Spectral Acceleration
- **PSHA** - Probabilistic seismic hazard assessment - Calcul probabiliste de l'aléa sismique
- **RC** - Reinforced Concrete
- **SHARE** - Seismic Hazard Harmonization in Europe
- **SHERIFS** - Seismic Hazard and Earthquake Rates in Fault Systems
- **UCERF** - Uniform Californian Earthquake Rupture
- **UHS** - Uniform Hazard Spectrum - for a range of PSA of different frequencies, the acceleration that can be exceeded with a set probability over a set period of time (e.g. 10% over 50 years).
- **WCR** - Western Corinth Rift (Greece) Forecast

Introduction

Depuis les plus jeunes heures de la civilisation humaine, les être humains ont dû soit s'adapter à leur environnement soit le domestiquer afin de pouvoir s'établir durablement dans une région. Parmi l'ensemble des défis qu'une civilisation doit surmonter, les catastrophes naturelles font partie des plus difficiles. Nombreux sont les exemples de civilisations ayant été affaiblies ou simplement effacées par un phénomène naturel trop destructeur pour que les êtres humains puissent se relever. En 464 av.JC, un séisme provoqua la destruction du gymnase de Sparte, tuant une grande partie des jeunes guerriers spartiates et donnant l'opportunité aux Hilotes de se révolter contre leurs oppresseurs de Sparte. En l'an 79, Pompéi fut recouvert par les cendres du Vésuve. Plus récemment, en 2005, l'ouragan Katerina frappa la ville de la Nouvelle-Orléans tuant plus de 1800 personnes et fragilisant l'économie de toute une région des États-Unis (Bankston III et al. [2010]). En 2011, une séisme de magnitude 9 frappa le nord du Japon, générant un tsunami qui s'abattît sur la centrale nucléaire de Fukushima causant une série d'évènements qui mena à la fusion du réacteur et à l'évacuation de plus de 80 000 personnes (Kim et al. [2013]).

Malgré les constructions de murs brises-vagues et d'infrastructures plus résistantes au mouvement sismique, les catastrophes naturelles demeurent aujourd'hui une menace difficile à réduire. La principale raison de cette difficulté est la longue période de récurrence entre deux évènements à un endroit donné, en particulier pour les séismes. Les catastrophes naturelles majeures ont typiquement un temps de retour bien supérieure à la durée de vie humaine et à la mémoire collective d'une société. Ces grands temps de récurrence rendent la prise de conscience du risque difficile à assimiler mais rendent également le phénomène physique difficile à étudier.

Pour comprendre et se protéger contre les catastrophe naturelles, il est nécessaire d'en comprendre l'origine. La cause des séismes resta pendant longtemps une énigme souvent attribuée à une action divine. Au XVIII em siècle, une théorie émergea selon laquelle les séismes seraient provoqués par des gaz circulant dans la croûte terrestre. Nous savons aujourd'hui que les séismes sont produits par le relâchement de contraintes accumulées le long d'une faille. Les temps de récurrence des séismes de taille suffisamment grandes pour être endommageant sont de l'ordre de plusieurs centaines d'années à plusieurs milliers d'années, trop longs pour pouvoir étudier un site donné en utilisant l'observation directe de l'évènement.

La sismologie remédie à ce problème en adoptant une approche ergodique, étudiant les séismes au niveau global, pour comprendre quelles propriétés et caractéristiques tous les séismes partagent et ainsi permettre l'élaboration de théories sur les propriétés d'un potentiel futur séisme dans une région donnée. La géologie s'appuie sur les grandes échelles de temps du passé en étudiant les vitesses de déplacement le long des failles générant les séismes à partir de données géomorphologiques ou les traces d'anciens séismes s'étant produits sur une faille à partir de l'étude de tranchées.

Cette thèse présente une méthodologie qui joint informations géologiques et sismologiques pour évaluer le risque sismique dans les régions où des failles actives ont été identifiées.

1 Contexte du projet

En 2014, pendant les études de sûreté nucléaire dans le contexte post-Fukushima, un calculs d'aléa sismique fut conduit pour les 19 centrales nucléaires françaises par l'opérateur des centrales. Je travaillais alors pour l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN) qui a été chargé de l'expertise de ces calculs par l'Autorité de Sûreté Nucléaire (ASN). Ma fonction était de faire l'expertise de la partie probabiliste des calculs présentés par l'opérateur nucléaire. L'une de mes conclusions personnelles fut que pour les sites proches de failles potentiellement actives, les calculs présentés, bien qu'appliquant une approche pouvant être considérée dans son ensemble comme l'état de l'art, ne menait pas à une représentation acceptable des taux annuels de sismicité sur les failles et de l'incertitude associée à ces taux.

Un an plus tôt était présenté le calcul d'aléa sismique pour la Californie, UCERF3 (Uniform Californian Earthquake Rupture Forecast 3) (Field et al. [2014]). Le modèle présenté établit un nouveau standard pour la modélisation des failles pour le calcul d'aléa sismique. Parmi les nouveautés, ce modèle autorise les ruptures sismiques complexes impliquant plusieurs failles tout en approchant le problème à l'échelle du système de faille dans son ensemble et pas seulement au niveau de chaque faille individuellement. De plus, les taux annuels de rupture sont inversés en se basant sur l'ensemble des données disponibles générant ainsi des modèles d'aléa cohérents avec l'état des connaissances à la fois géologiques, géodésiques, sismologiques et paléosismologiques. Malheureusement, l'approche développée pour ce modèle nécessite un niveau de ressources humaines, scientifiques afin de mettre en place des bases de données uniformes et cohérentes, ainsi que des ressources informatiques qui sont hors de portée pour la plupart des études d'aléa sismique dans d'autres régions, notamment en France.

Nous avons mis en place ce projet de thèse après la constatation que le manque de méthodologies et d'outils permettant une représentation des failles dans les modèles d'aléa sismique en accord avec les concepts de la communauté scientifique limitait la prise de décision sur le risque sismique.

Il fut alors clair qu'il était nécessaire de développer une méthodologie et les outils associés à cette méthodologie qui permettent de se rapprocher du standard de UCERF3 dans des régions où moins d'information sur les failles est disponible.

Une seconde conclusion de l'étude post-Fukushima sur les centrales françaises concernait le calcul d'aléa sismique probabiliste en général. Le calcul probabiliste, présenté en détail plus bas, intègre l'ensemble des participations des sources sismiques d'une région et permet une exploration des incertitudes épistémiques associées avec les différentes étapes du calculs. Le résultat est donc difficile à relier aux hypothèses scientifiques sous-jacentes ce qui peut rendre toute prise de décision faite sur la base de ces calculs difficile pour un non spécialiste, en particulier lorsque l'incertitude associée au résultat est forte.

Un second objectif de la thèse fut donc établi: en parallèle de permettre une meilleure modélisation des failles, la méthodologie développée devra également permettre une meilleure communication des incertitudes impactant l'aléa et le risque sismique afin d'aider à la prise de décision.

2 Chaîne de calcul du risque sismique

Différents éléments de calculs s'imbriquent pour obtenir une estimation du risque sismique (1). L'aléa sismique est le mouvement du sol pouvant affecter un enjeu (bâtiment, ville, infrastructure ou vie humaine). La réaction de l'enjeu au mouvement sismique est décrit par sa vulnérabilité. La combinaison de l'aléa et de la vulnérabilité permet de calculer le risque sismique.

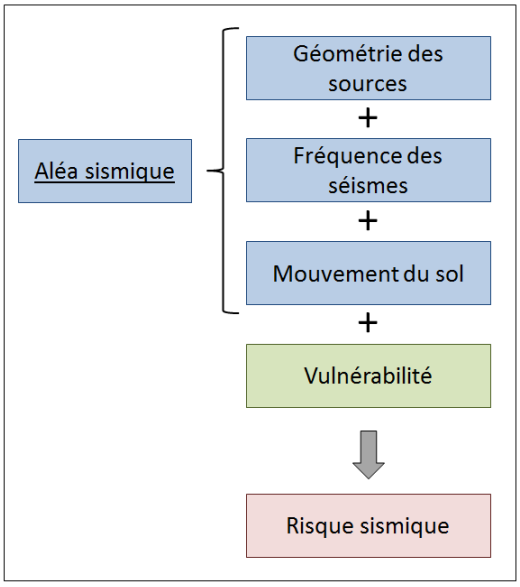

Figure 1: *Chaîne de calcul du risque sismique.*

Le calcul probabiliste de l'aléa sismique peut être divisé en trois étapes : la description de la géométrie des

sources sismiques, le calcul des magnitudes et fréquences des séismes sur chacune des sources, et le calcul du mouvement sismique au site d'intérêt en fonction de la localisation du séisme par rapport au site, sa magnitude et les conditions de sol du site.

Dans cette étude, nous nous concentrerons sur la modélisation des sources sismiques et du calcul de la fréquence des séismes (figure 1). Cependant, nous réaliserons le calcul du risque sismique en intégrant l'intégralité de la chaîne de calcul (figure 1). Pour le calcul du mouvement sismique, nous utiliserons des GMPEs (Ground Motion Prediction Equation) disponibles dans la littérature qui pour des caractéristiques de séismes données à une distance donnée permettent d'obtenir la probabilité de dépasser un certain niveau de mouvement sismique. Pour la vulnérabilité, nous nous appuierons sur des courbes de fragilité publiées qui décrivent la probabilité qu'un bâtiment soit endommagé lorsqu'il est soumis à un mouvement sismique.

3 Problématiques concernant l'intégration des failles dans le calcul d'aléa sismique

3.1 Ruptures complexes

La plupart des études d'aléa sismique (Erdik et al. [2004],Woessner et al. [2015],Sesetyan et al. [2018],...) s'appuient sur des modèles de segmentation des failles permettant une modélisation faille par faille des taux de sismicité mais ne permettant pas de de modéliser les ruptures complexes pouvant rompre plusieurs segments simultanément. Le modèle californien UCERF3 permet de s'affranchir des hypothèses de segmentation. Les taux annuels de ruptures d'un segment unique ou de multi-segments sont calculés par inversion à l'échelle du système de faille entier prenant en compte un grand nombre de paramètres. L'approche utilisée dans UCERF3 a changé l'état de l'art et a montré l'importance de considérer un système de failles comme un ensemble et non pas comme une collection de failles isolées. Cependant, l'implémentation de cette approche est complexe et nécessite une bonne connaissance du contexte tectonique du système de faille.

Une méthodologie plus versatile construisant sur les avancées de UCERF3 est requise pour que l'état de l'art définit par UCERF3 puisse être appliqué à d'autres systèmes de failles.

3.2 Sismicité sur les failles et sismicité de fond

Les modèles d'aléa sismique prenant en compte les failles établissent une magnitude de troncature (Mt) (Woessner et al. [2015], Sesetyan et al. [2018], Chartier et al. [2017a],...). Les séismes de magnitude supérieure à Mt se produisent

sur les failles modélisées et les séismes de magnitude inférieure à Mt se produisent dans la sismicité de fond, une zone sismotectonique définie autour des failles. Selon des études, Mt peut être comprise entre 5.5 et 6.5. Cette approche a deux impacts potentiels sur l'aléa sismique:

- le modèle ne considère pas la possibilité de séismes de magnitude supérieure à Mt sur des failles inconnues, considérées inactives ou simplement non modélisées. Cela peut entraîner une sous-estimation du risque pour les sites qui ne sont pas à proximité immédiate de telles failles.
- le modèle ne considère pas que les séismes de magnitude inférieure à Mt puissent se produire sur les failles. Cela peut entraîner une sous-estimation du risque pour les sites à proximité immédiate des failles.

Les taux annuels des séismes sur les failles sont calculés à partir de la vitesse de glissement de la faille et les taux de sismicité de fond sont calculés en se basant sur le catalogue de sismicité. Il est possible que, dans certaines régions, ces deux sources d'information ne convergent pas et mènent à des taux de sismicité différents.

Afin d'adresser ces problèmes, il est nécessaire de développer une approche innovante de modélisation des taux de sismicité dans la sismicité de fond.

3.3 Contraintes pour la méthodologie à développer

En vue des problématiques exposées précédemment, une nouvelle méthodologie pour les taux de sismicité des failles doit prendre en compte les points suivants:

- suppression des hypothèses de segmentation et ainsi autoriser la possibilité de rupture sismique conjointe de plusieurs segments de failles simultanément.
- une modélisation au niveau du système de faille et non au niveau de chaque segment individuellement.
- l'ensemble des données disponibles concernant un système de failles doit être pris en compte.
- la possibilité d'occurrence de séismes sur des failles non modélisées, c'est à dire dans la sismicité de fond, doit être considérée ainsi que la possibilité d'occurrence de petits séismes sur les failles.
- faciliter l'utilisation de la méthode par des utilisateurs n'étant pas spécialisé dans les méthodes numériques et voulant tester des hypothèses concernant un réseau de faille.
- la méthodologie doit être applicable sur des systèmes de failles variés autant en terme de contexte géodynamique qu'en terme de niveau de connaissance.

4 Problématique de traçabilité des sources du risque sismique et des incertitudes associées

Le calcul du risque sismique prend en compte l'ensemble des sources sismiques d'une région, la fréquence des séismes sur ces sources, le mouvement sismique associé à chaque séisme, ainsi que la réaction du bâtiment face au mouvement sismique. L'ensemble de ces paramètres est intégré pour présenter le résultat sous la forme d'une simple métrique telle que la probabilité d'effondrement d'un bâtiment ou la perte financière probable. Après cette intégration, il est difficile d'identifier les sources contrôlant le risque sismique et l'incertitude, étape essentielle afin d'orienter les pistes de recherche future qui pourront améliorer la compréhension du risque sismique.

De plus, le calcul du risque sismique est le plus souvent réalisé par la communauté des ingénieurs des structures qui s'appuient sur les résultats d'études d'aléa disponibles dans la littérature (Hong and Goda [2006],Iervolino et al. [2018], ...). De ce fait, dans les études de risque sismique, l'accent est en majorité porté sur les techniques de modélisation des structures et très peu sur la modélisation des sources sismotectoniques.

Dans cette étude, les taux de sismicité modélisés sont utilisés pour calculer la probabilité d'effondrement d'un bâtiment. L'impact de chaque hypothèse sur le résultat est analysé afin d'identifier les incertitudes sur les hypothèses d'entrées entraînant les impacts sur le résultat les plus importants. En parallèle, une méthodologie de désagrégation de la probabilité d'effondrement est mise en place afin d'identifier les sources contrôlant le risque.

En adoptant le point de vue du risque sismique, nous souhaitons faire un pas dans la direction de l'amélioration de la communication entre la communauté des géosciences, la communauté des modélisateurs d'aléa sismique et la communauté des ingénieurs. L'illustration des sources d'incertitudes liées aux hypothèses concernant les sources sismotectoniques qui affectent fortement l'estimation du risque sismique peut permettre à la communauté des géosciences d'apprécier quantitativement l'impact de la recherche en géosciences sur la connaissance du risque. Développer des outils d'intégration plus direct de l'aléa et du risque permet également aux communautés des modélisateurs de l'aléa et des ingénieurs d'échanger sur les méthodes de modélisation et d'obtenir une représentation plus réaliste des incertitudes.

Finalement, un résultat d'analyse du risque sismique plus transparent en terme de données utilisées, de méthodes et d'incertitudes associées est une pierre angulaire de la prise de décision concernant les mesures de réduction du risque sismique à mettre en oeuvre.

5 Plan de la thèse

Le **chapitre 1** se compose de trois parties. La première partie est une publication (Chartier et al. [2017b]) présentant l'élaboration de la méthodologie de calcul des taux annuels de rupture dans un système de faille avec une application sur le réseau de faille du Rift de Corinthe. Le code de calcul appliquant la méthodologie fut développé sous le nom de SHERIFS (Seismic Hazard and Earthquake Rates In Fault Systems). La seconde partie du chapitre est une publication (Chartier et al. [2019]) présentant le fonctionnement du code SHERIFS et les évolutions de la méthodologie vis-a-vis de la première partie. La troisième partie argumente l'intérêt de rendre un code de calcul disponible et présente les discussions les évolutions qui ont été possibles grâce au partage de SHERIFS.

Le **chapitre 2** présente une application de SHERIFS pour la faille nord anatolienne en Turquie dans la région de la mer de Marmara. Nombre d'incertitudes pèse sur ce système de faille, parmi lesquelles la condition de bloquage de la faille, la dimension des ruptures possible, la forme de la distribution de la sismicité, et la proportion de séismes pouvant se produire dans la sismicité de fond. Il fut donc possible de tester les capacités de SHERIFS en explorant ces incertitudes dans un arbre logique. Les taux de sismicité modélisés sont comparés aux taux calculés à partir du catalogue de sismicité et des études de paléosismicité disponibles dans la région. Afin d'enrichir la discussion concernant les hypothèses d'entrées du calcul, une modélisation de la sismicité est réalisée avec le code de calcul "physics-based" RSQSim. La comparaison avec les données et les résultats de RSQSim permet d'attribuer des poids à chaque branche de l'arbre logique et ainsi d'obtenir une représentation réaliste de l'incertitude associée aux taux de sismicité dans la région.

Le **chapitre 3** s'oriente vers le calcul du risque sismique et se compose de deux parties. Dans la première partie, l'arbre logique construit dans le chapitre 2 est utilisé pour calculer la probabilité d'effondrement de deux immeubles théoriques de 5 étages à Istanbul. L'impact de chaque hypothèse explorée dans l'arbre logique sur la probabilité d'effondrement est discuté. En utilisant une méthodologie innovante de désagrégation du risque sismique, il est possible d'identifier les sources sismique contrôlant la probabilité d'effondrement et d'observer que ces sources diffèrent selon les hypothèses d'entrées et la résistance du bâtiment. Dans la seconde partie, les résultats préliminaires d'un calcul de perte financière sur un portefeuille de 1000 bâtiments dans la région d'Istanbul sont présentés.

6 Rappel sur les méthodes de calcul de l'aléa sismique

6.1 Calcul déterministe

Dans le calcul déterministe de l'aléa sismique, un scénario sismique est considéré comme ayant une probabilité d'occurrence de 1. Le mouvement sismique est calculé en utilisant une GMPE. Plusieurs approches peuvent être utilisées pour établir ce scénario. Si le site ou les sites d'intérêt sont proches d'une faille active, le scénario peut être définit comme le séisme de magnitude maximal pouvant être envisagé sur cette faille. Si aucune faille active n'est identifiée à proximité du site mais qu'il existe une sismicité diffuse, le scénario est défini par un séisme de magnitude vraisemblable pouvant être généré à proximité du site. Il convient alors d'établir des règles permettant de définir ce qu'est un séisme vraisemblable pour une région.

En France, pour la sûreté nucléaire, c'est cette seconde approche qui est utilisée. Les règles décrivant la manière de définir une magnitude vraisemblable et les marges de sûreté à imposer sont énoncées dans la Règle Fondamentale de Sûreté (RFS - 2001-01).

L'approche déterministe permet de communiquer clairement, à la fois aux ingénieurs et aux décideurs, l'évènement sismique contre lequel il conviendra de se protéger ainsi que les incertitudes épistémiques associées à l'estimation de la magnitude et la localisation de cet évènement. Il est également possible d'explorer les incertitudes épistémiques, pesant sur l'estimation de la magnitude d'un séisme vraisemblable par différent experts ainsi que la GMPE utilisée. En revanche, l'approche déterministe ne permet pas d'explorer les incertitudes aléatoires du mouvement sismique telles que la directivité, ou la chute de contrainte.

6.2 Calcul probabiliste

6.2.1 L'approche Cornell

L'approche probabiliste, développée par Cornell en 1968 sous le nom PSHA (Probabilistic Seismic Hazard Assessment), permet de prendre en considération l'ensemble des sources sismiques et les incertitudes aléatoires que l'approche déterministe néglige. Les incertitudes aléatoires comprennent l'occurrence aléatoire des séismes qui est représentée par le taux annuel d'occurrence d'un séisme de magnitude donnée sur une source sismique, le mouvement sismique généré par un séisme d'une magnitude donnée à une distance donnée.

L'approche PSHA se compose de quatre étapes (2). Tout d'abord, il convient d'identifier la géométrie de l'ensemble des sources sismiques potentielles proches du site d'intérêt. Ces sources peuvent être des failles actives, des points sources représentant l'occurrence d'un séisme, ou bien des zones de sismicité diffuse où la sismicité est

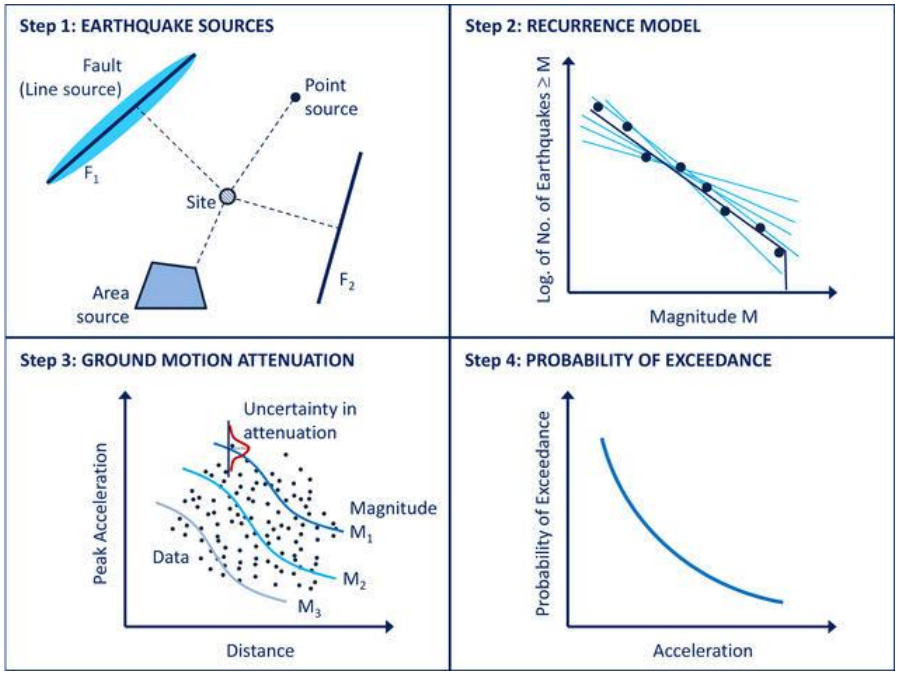

Figure 2: *Étapes du calcul PSHA selon Cornell 1968.*

considérée uniforme spatialement. Dans un second temps, il faut définir la fréquence et la magnitude des séismes pouvant se produire sur chacune des sources. Le troisième élément est la GMPE, permettant d'obtenir la probabilité de dépasser un certain niveau de mouvement sismique en fonction des caractéristiques du séisme (magnitude, distance, cinématique, ...) ainsi que des caractéristiques du sol au droit du site d'intérêt. La quatrième et dernière étape, qui est la combinaison des éléments des trois premières étapes grâce à l'équation 1, permet d'obtenir la courbe l'aléa sismique. La courbe d'aléa sismique décrit la probabilité de dépasser chaque niveau de mouvement sismique pendant une période donnée.

$$P(A > a^*) = \int_M \int_D \lambda_{M,D} \times P(A > a^* | M, D) \tag{1}$$

Où $\lambda_{M,D}$ est le taux annuel de séismes de magnitude M à la distance D; et $P(A > a^*|M, D)$ est la probabilité du mouvement du sol A d'excéder le mouvement cible a^* étant donnée la magnitude M à la distance D. $P(A > a^*|M, D)$ est calculé grâce à une GMPE.

6.2.2 Éléments du calcul probabiliste d'aléa sismique

a) Géométrie des sources

La géométrie des sources sismiques s'appuie à la fois sur les études géologiques et géophysiques de terrain et sur l'étude du catalogue de sismicité. L'identification des failles actives par les géologues permet de localiser là où peut être attendue la sismicité future. L'analyse du catalogue de sismicité permet de définir les ensembles sismotectoniques en

fonction de la densité de la sismicité enregistrée et ainsi définir des sources sismotectoniques pour lesquelles le potentiel sismogène peut être considéré comme uniforme.

b) Récurrence des séismes

Les taux de sismicité des zones sismotectoniques sont estimés à partir du catalogue de sismicité. L'instrumentation sismologique des territoires s'étant développée à partir des années 1960, les catalogues de sismicité instrumentale sont courts comparés à la récurrence des séismes endommageant qui peut être de plusieurs siècles. Des catalogues de sismicité historique sont construits à partir des témoignages de ressentis de séismes ayant eu lieu avant la période instrumentale. Le seuil de détectabilité des séismes s'est amélioré au cours du temps, aussi bien pour les détections instrumentales que pour la survie des témoignages historiques. Pour chaque gamme de magnitude est établie une date à partir de laquelle le catalogue peut être considéré complet. Le nombre de séismes observés pendant la période complète permet de calculer le taux annuel de séismes pour chaque gamme de magnitude.

c) GMPEs

Les GMPES sont établies empiriquement par analyse statistique des mouvements sismiques enregistrés, à partir des bases de données d'accélérogrammes, des relations entre les paramètres du séisme, notamment la magnitude et la distance à la station, et des indicateurs de mouvement sismique issus de l'accélérogramme. Parmi les indicateurs, le plus couramment utilisé est le PGA (Peak Ground Acceleration) qui est la valeur maximale de l'accélération enregistrée à la station lors du séisme. Le PGA étant un indicateur plutôt représentatif du contenu haute fréquence du séisme, des accélérations pseudo-spectrales (PSA for Pseudo-Spectral Acceleration) sont également calculées à partir de l'accélérogramme. Les PSA sont la réponse au mouvement sismique d'oscillateurs harmoniques à une dimension avec un amortissement de 5% de différentes fréquences propres. Le PGV (Peak Ground Velocity) qui est la vitesse maximale enregistrée est également utilisé.

Ces différents indicateurs correspondent à différentes fréquences de réponse de la structure des bâtiments et sont ainsi utiles pour calculer les dégâts potentiels. Par exemple, un bâtiment de 3 étages a une fréquence propre proche de 3Hz, ainsi un calcul de la probabilité d'occurrence d'un PSA de 3Hz sera un indicateur utile pour calculer la probabilité de dégât pour ce bâtiment.

6.2.3 La désagrégation

Le calcul probabiliste agrège la participation à l'aléa sismique de l'ensemble des source sismiques présentent dans une région. Afin d'identifier les sources participant en majorité à la probabilité de dépassement d'un niveau d'aléa sismique donné, il est possible de désagréger le calcul. Pour cela il est nécessaire de séparer les probabilités conditionnelles

de dépassement d'un niveau d'aléa sismique en fonction de la magnitude, de la distance, et/ou de l'emplacement géographique de la source sismique.

Il est également possible de faire la désagrégation en epsilon de la distribution de probabilité de la GMPE qui permet de tracer la participation de mouvements sismiques prédits par la GMPE proches de la moyenne ou bien des extrêmes de la distribution de probabilité.

6.2.4 L'approche par simulation de catalogues de sismicité synthétiques

Il existe des approches alternatives à l'approche Cornell pour calculer l'aléa sismique probabiliste. L'approche "event-based" (Bommer et al. [2002]) utilise les mêmes ingrédients que l'approche Cornell décrite précédemment mais ne s'appuie pas sur le formalisme probabiliste de l'équation 1. Par une approche Monte-Carlo, elle simule un catalogue de sismicité synthétique de plusieurs milliers d'années et calcule le mouvement sismique associé à chaque séisme. En analysant l'historique des mouvements sismiques synthétiques enregistrés à un site donné, il est possible de calculer la probabilité annuelle de dépassement d'un niveau sismique. Cette approche présente l'avantage d'illustrer clairement le mouvement sismique lié à chaque évènement du catalogue et ainsi présenter un résultat plus simple à appréhender par une audience peu familière avec le calcul probabiliste. La simulation d'un catalogue synthétique peut également permettre de modéliser des séquences sismiques de manière simple et explicite tandis que leur probabilisation est plus compliquée à modéliser mais également à communiquer. En revanche, afin d'explorer les incertitudes aléatoires sur l'occurrence des séismes pour chacune des sources sismogéniques et sur le mouvement sismique généré au site pour un séisme donné, il est nécessaire de générer des catalogues de plusieurs millions, voire milliards de séismes. Ceci est extrêmement coûteux en terme de puissance de calcul et de mémoire et mène souvent à négliger l'exploration des incertitudes épistémiques pesant sur la définition des sources.

7 Rappel sur les méthodes de calcul du risque sismique

7.1 Calcul site spécifique

Le calcul du risque sismique pour un site spécifique, pour un site critique tel qu'une central nucléaire ou un barrage, se base sur le calcul d'aléa sismique au site et la désagrégation. Pour une période de retour fixée, en d'autres termes une probabilité fixée, la désagrégation présente les sources sismiques contribuant au dépassement de l'aléa sismique. Une sélection d'accélérogrammes correspondant aux magnitudes et distances participant majoritairement à l'aléa sismique dans la désagrégation est extraite des bases de données (Rizzo et al. [1975]).

Cette sélection d'accélérogrammes est ensuite utilisée par les ingénieurs des structures comme sollicitation

sismique qui, en utilisant des modèles numériques, propagent le mouvement au sein de la structure afin d'évaluer les potentiels endommagements.

Afin de faciliter le calcul, il est également possible de s'appuyer sur des courbes de fragilité. Développées empiriquement ou synthétiquement par des ingénieurs des structures, elles relient un paramètre de mouvement du sol à la probabilité d'endommagement d'un bâtiment. Le paramètre peut être le PGA ou un PSA d'une fréquence correspondant à la fréquence propre du bâtiment.

Utiliser une courbe de fragilité simplifie grandement la mise en place du calcul car aucun calcul supplémentaire de la réaction du bâtiment au mouvement sismique n'est nécessaire. En revanche, le fait d'utiliser un seul indicateur du mouvement sismique entraîne l'omission d'une part de la complexité du mouvement sismique telles que la présence de différents contenus fréquentiels, la durée, ou encore la polarisation.

7.2 Calcul de perte pour un ensemble de sites

Lorsqu'un séisme se produit, il peut endommager un grand nombre de bâtiments simultanément. Pour calculer la probabilité de perte il faut donc considérer la possibilité que plusieurs bâtiments soient endommagé pendant le même évènement sismique. L'approche Cornell (Cornell [1968]) ne permet pas de prendre en compte la probabilité conditionnelle d'endommagement d'un bâtiment sachant qu'un autre bâtiment est endommagé.

Afin de calculer cette probabilité, l'approche "event-based" est utilisée. Un catalogue de sismicité synthétique est simulé à partir du même type de modèle de taux annuel d'occurrence des séismes utilisés dans l'approche Cornell. Pour chaque séisme du catalogue synthétique, un champ de mouvement sismique est simulé grâce à une GMPE. L'incertitude du mouvement sismique prédit par la GMPE peut être explorée en prenant en compte la corrélation spatial du mouvement sismique (Jayaram and Baker [2009]). La probabilité d'endommagement de chaque bâtiment est calculé en utilisant une courbe de fragilité adaptée au type de bâtiment. Différentes métriques peuvent être utilisées pour présenter le risque, telle que le nombre de bâtiments affectés ou effondrés, la perte financière associée au dégâts, ou encore le nombre de victimes. L'analyse statistique du catalogue synthétique de dégât permet de calculer la probabilité du risque et les séismes qui entraînent les pertes les plus importantes ou les plus fréquentes (Hong and Goda [2006]).

Chapter 1

Methodology for modelling earthquake rates in fault systems

Résumé du chapitre

La fréquence annuelle de séismes sur les failles est un élément essentiel du calcul probabiliste de l'aléa et du risque sismique. Dans ce chapitre, nous présentons une méthodologie novatrice pour calculer la fréquence annuelle des séismes dans un réseau de faille à partir de la vitesse de glissement et de la géométrie des failles. Cette méthodologie considère le système de failles dans son ensemble et permet de modéliser les fréquences annuelles de ruptures complexes impliquant plusieurs sections de failles. En plus de la géométrie et de la vitesse de glissement des failles dans le réseau, des hypothèses sur les ruptures possibles et sur la distribution en magnitude et fréquence (MFD) de la sismicité sont requises. La méthodologie permet également de considérer la possibilité que des séismes futurs arrivent sur des failles inconnues, donc dans la sismicité de fond.

La méthodologie est implémentée dans le code SHERIFS (Seismic Hazard and Earthquake Rates in Fault Systems) de telle sorte à être utilisable aisément par des utilisateurs sans expérience de programmation. SHERIFS est assez flexible pour modéliser des systèmes de failles diversifiés et pour explorer une grande variété d'incertitudes épistémiques, notamment concernant les hypothèses de ruptures complexes et de MFD. En comparant les taux modélisés au taux calculés à partir du catalogue de sismicité et aux taux calculés à partir des études de paléosismicité, il est possible de discuter et pondérer les hypothèses d'entrées explorées.

Une application sur le système de failles de la partie occidentale du Rift de Corinthe en Grèce est présentée. Il est observé que malgré le fait que la sismicité du réseau de faille suit une MFD de type Gutenberg-Richter, localement,

les failles peuvent présenter une MFD proche d'un séisme caractéristique. De plus, afin de reproduire les taux de sismicité observés dans le catalogue et dans la paléosismicité, il est nécessaire de considérer possible la rupture conjointe de plusieurs sections de faille durant le même séisme.

Au travers des différents exemples d'application de SHERIFS, nous exposons les difficultés mais également et surtout les avantages de partager un code appliquant une méthodologie novatrice. L'utilisation de SHERIFS par des personnes extérieures à son développement à permis d'améliorer la méthodologie et de faire avancer les discussions concernant un système de failles en comparant avec d'autre systèmes de failles modélisés.

Structure of the chapter

This chapter presents a novel methodology for modelling earthquake rates in faults systems and is composed of three sections.

The first section is an article published in 2017 exposing a proof of concept for an earthquake rate modelling methodology with an application on the Western Corinth Rift in Greece (Chartier et al. [2017b]). It lays down the concepts of the methodology and relies on the fault system parameters established during the PhD thesis of Aurélien Boiselet.

The second section of the chapter is an article published in 2019 presenting the code SHERIFS (Seismic Hazard and Earthquake Rates In Fault Systems) that applies the methodology presented above (Chartier et al. [2019]). This article exposes the details of the algorithm and provides an example of application. The additional features of the code added since the application on the Western Corinth Rift are also presented. The goal of the article is to be a support for any user. The SHERIFS code and a user manual is available online at `https://github.com/tomchartier/SHERIFS`.

The third and final section of the chapter is a discussion about the process of developing and sharing an open-source and user-friendly code such as SHERIFS. In this section, the difficulties of the process are fully exposed as well as the outcomes of the sharing. A large number of scientific discussions were made possible only because the code was available for others to apply in different study areas. The goal of the section is to argue for the open sharing of scientific codes and convince PhD students that the benefits of sharing a code drastically outweighs the time and energy costs.

1.1 Methodology for earthquake rupture rate estimates of fault networks: example for the western Corinth rift, Greece

1.1.1 Abstract

Modeling the seismic potential of active faults is a fundamental step of probabilistic seismic hazard assessment (PSHA). An accurate estimation of the rate of earthquakes on the faults is necessary in order to obtain the probability of exceedance of a given ground motion. Most PSHA studies consider faults as independent structures and neglect the possibility of multiple faults or fault segments rupturing simultaneously (fault-to-fault, FtF, ruptures). The Uniform California Earthquake Rupture Forecast version 3 (UCERF3) model takes into account this possibility by considering a system-level approach rather than an individual-fault-level approach using the geological, seismological and geodetical information to invert the earthquake rates. In many places of the world seismological and geodetical information along fault networks is often not well constrained. There is therefore a need to propose a methodology relying on geological information alone to compute earthquake rates of the faults in the network. In the proposed methodology, a simple distance criteria is used to define FtF ruptures and consider single faults or FtF ruptures as an aleatory uncertainty, similarly to UCERF-3. Rates of earthquakes on faults are then computed following two constraints: the magnitude frequency distribution (MFD) of earthquakes in the fault system as a whole must follow an a priori chosen shape and the rate of earthquakes on each fault is determined by the specific slip rate of each segment depending on the possible FtF ruptures. The modeled earthquake rates are then compared to the available independent data (geodetical, seismological and paleoseismological data) in order to weight different hypothesis explored in a logic tree. The methodology is tested on the western Corinth rift (WCR), Greece, where recent advancements have been made in the understanding of the geological slip rates of the complex network of normal faults which are accommodating the ~ 15 mm.yr^{-1} north–south extension. Modeling results show that geological, seismological and paleoseismological rates of earthquakes cannot be reconciled with only singlefault-rupture scenarios and require hypothesizing a large spectrum of possible FtF rupture sets. In order to fit the imposed regional Gutenberg–Richter (GR) MFD target, some of the slip along certain faults needs to be accommodated either with interseismic creep or as post-seismic processes. Furthermore, computed individual faults' MFDs differ depending on the position of each fault in the system and the possible FtF ruptures associated with the fault. Finally, a comparison of modeled earthquake rupture rates with those deduced from the regional and local earthquake catalog statistics and local paleoseismological data indicates a better fit with the FtF rupture set constructed with a distance criteria based on 5 km rather than 3 km, suggesting a high connectivity of faults in the WCR fault system.

1.1.2 Introduction

The goal of probabilistic seismic hazard assessment (PSHA) is to estimate the probability of exceeding various groundmotion levels at a site (or a map of sites) given the rates of all possible earthquakes. The first step of PSHA following a Cornell–McGuire approach (Cornell [1968]; McGuire [1976]) is the characterization of the seismic sources. For regions where active faults have been identified and their slip rates are known, several methods have been proposed in order to calculate the rate of earthquakes occurring on these faults. The most commonly used methods consider faults as independent structures on which the strong earthquakes are located (e.g., SHARE project in Europe, Woessner et al. [2015]; Yazdani et al. [2016], in Iran; TEM model in Taiwan, Wang et al. [2016]). In these PSHA studies, a background seismicity will generate earthquakes up to a threshold magnitude (Mw) of 6.0 or 6.5, beyond which earthquakes are generated on the faults. The rate of earthquakes for these larger magnitudes is based on geological and paleoseismological records, and the maximum magnitudes depend on the physical dimensions of the fault under consideration. In the resulting model, the rate of lower magnitudes is controlled by seismological information and the rate of stronger magnitudes by geological information. In cases where large historical earthquakes are associated with multiple fault segments, the individual fault segments described by the geologists in the field are regrouped in a larger fault source and a mean slip rate is attributed to the fault source. A specific magnitude frequency distribution (MFD), often Gutenberg– Richter (GR) (Gutenberg and Richter [1944]) or characteristic earthquake (Wesnousky [1986]), describing the mean sliprate-based earthquake rate on the fault is attributed to each fault source. This process requires simplifying fault complexity in terms of geometry and slip rate and does not allow complex ruptures that propagate from one fault source to an adjacent one.

In the past decades, the quality of observations has improved and our understanding of earthquakes has grown (proceedings of the 2017 Fault2SHA meeting). We observe more and more complex earthquake ruptures propagating on several neighboring faults. There is thus a need for hazard models to accurately represent the fault and rupture complexity observed in the field by geologists and to correctly distinguish aleatory from epistemic uncertainties.

Toro et al. [1997] define epistemic uncertainty as “uncertainty that is due to incomplete knowledge and data about the physics of the earthquake process. In principle, epistemic uncertainty can be reduced by the collection of additional information”. Aleatory uncertainty, on the other hand, is an “uncertainty that is inherent to the unpredictable nature of future events”; in this respect a fault-to-fault (FtF) rupture should be treated as an aleatory uncertainty since it is linked to the randomness of the seismic phenomenon.

In order to allow FtF ruptures, the Working Group on California Earthquake Probabilities (WGCEP-2003) for the San Francisco Bay region developed a methodology that explores possible FtF ruptures in a logic tree. Each branch of the logic tree represents a seismic hazard model, and the rate of the corresponding FtF rupture scenario is

obtained by weighting the branches. In the hazard study of the Marmara Region in Turkey, Gülerce and Ocak [2013] used this approach and set the weight of each branch (or rupture scenario) such that the mean seismicity rate modeled by the logic tree fits the recorded seismicity rate around the fault of interest. This method treats the uncertainty of FtF ruptures as an epistemic uncertainty in the PSHA calculation.

The Uniform California Earthquake Rupture Forecast version 3 (UCERF-3) model was developed using a novel methodology that treats all possible combinations of FtF rupture scenarios within the same branch of the logic tree as an aleatory uncertainty (Field et al. [2014]). Per their terminology, faults are divided into smaller sections and all possible section-to-section ruptures are investigated. The possibility of ruptures happening is controlled by a set of geometric and physical rules and the rate of earthquakes is computed using a "grand inversion" of the seismological, geological, paleoseismological and geodetical data available in California. The regional GR MFD of earthquakes of California and the geodetical deformation are used as a target for the total earthquake rupture forecast in each deformation model. This grand inversion relies also on estimates of the creep rate on faults deduced from local deformation data when available.

For many fault networks, only sparse seismological and geodetical data are available, and the geological record is often the most detailed source of information concerning the faults' activity. In such cases, it is necessary to develop a methodology that allows building seismic hazard models relying only on geological data and yet allowing FtF ruptures as an aleatory uncertainty. The sparse geodetical, seismological and paleoseismological data can then be used as a means of comparison to help weighting the different input hypothesis.

In this study we propose such a methodology based on the slip-rate budget of each fault, FtF ruptures hypothesis and assumptions on the shape of the MFD defined for the fault system as a whole. The methodology is developed so as to be flexible and applicable to regions where data on faults, geodesy and seismicity may be sparse. The rate of earthquakes on faults computed using geological information (slip rates) and is then compared to other sources of information such as the regional and local earthquake catalogs and the paleoseismic data in order to weight the different epistemic uncertainty explored in the logic tree. Moreover, it is also known that faults accommodate important amounts of slip in either post-seismic slip or in creep events (e.g., L'Aquilla earthquake 2009; Cheloni et al. [2014]; Napa earthquake 2014; Lienkaemper et al. [2016]). These phenomena, called non-main-shock slip (NMS) later on, are integrated in the slip rates deduced from geological information and should not be converted into earthquake rates when computing seismic hazard. The methodology presented in this study allows part of the geological slip rate to be considered as NMS slip rate.

We use this methodology to generate fault-based hazard models for the western Corinth rift (WCR), Greece, which has been studied for the past decade by the Corinth Rift Laboratory Working Group (CRL-WG) Lyon-Caen et al. [2004];Bernard et al. [2006]; Lambotte et al. [2014]). A large number of active faults have been identified in

this area, and a consensus about their possible geometries and activity rates has been reached within the CRL-WG (Boiselet [2014]). We used this geologic information to test our modeling approach and explore different epistemic uncertainties in a logic tree. Finally, we compared the modeled earthquake rates of each fault with seismological and paleoseismological data in order to weight the hypothesis in the logic tree.

1.1.3 Novel methodology for taking faults into account in PSHA

In most regions of the world the amount of data available to model faults in a PSHA study is often sparse and uncertain. However, the need to consider such data in PSHA is increasing, and the methods to properly incorporate the available geological information in the hazard models are still missing. In this context, we propose building a methodology that allows considering all the available information on faults, allows setting rules to define FtF ruptures and considers single faults or FtF ruptures as an aleatory uncertainty.

Our iterative method allows converting the slip-rate budget of each individual fault rate of earthquakes by requiring that the resulting regional MFD of earthquakes in the whole fault system follows an imposed shape. The MFD shape of each individual fault will thus be a result of the iterative process and not an imposed parameter.

The proposed method is presented here in a nutshell and illustrated in figure 1.1.

1. The necessary input data include the following:
    - A definition of the 3-D geometry of the fault system.
    - An estimate of the geological slip rates of each individual fault that determines the slip-rate budget of the fault.
2. Setting up the methodology requires the following:
    - Choosing suitable scaling laws to estimate the maximum magnitude each fault can host.
    - Assuming the minimum magnitude of earthquakes possible on the faults (Mw 5.0 in this study).
    - Hypothesizing FtF rupture scenarios based on some rules. In this study only a simple distance rule is used to define FtF ruptures. In future developments, more physics-based approaches could be explored. In the example presented in figure 1.1 a, the three faults (fault 1, fault 2 and fault 3) are considered to be sufficiently close to each other; thus, they can either rupture individually (F1, F2, F3) or in FtF ruptures (F1 + F2, F2 + F3 or F1 + F2 + F3).
    - Imposing a shape for the target MFD for the whole fault system. In this study a GR MFD distribution is assumed.

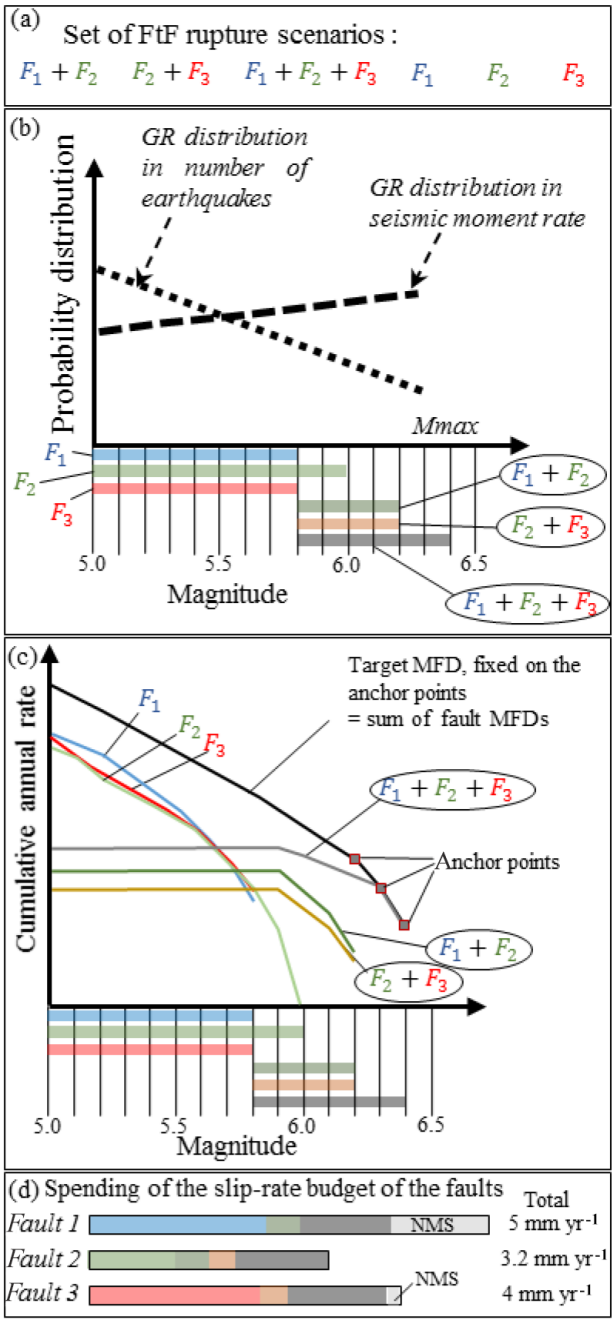

Figure 1.1: *Illustration of the methodology. (a) Set of FtF rupture scenarios. (b) Picking of the magnitude bins and of the sources. (c) Building the target MFD: the black curve is the target MFD anchored at the mean of the three highest magnitude bins (magnitude bin of 0.1). The sum of the resulting MFDs of the six sources has to be equal to the target MFD. (d) Visualization of the partitioning by the iterative methodology of each fault's slip-rate budget (colors correspond to the individual rupture or the FtF rupture; NMS is non-main-shock slip).*

3. Three pre-computational steps are performed to carry out the following:

   - To calculate all possible magnitude bins each fault and FtF rupture scenario can accommodate according to each scaling law considered (figure 1.1b).
   - To calculate the number of incremental quantities of slip rate (dsr) contained in each fault budget. In the

example, fault 1, fault 2 and fault 3 have a slip-rate budget of 5, 3.2 and 4 mm yr^{-1} respectively (figure 1.1d). Therefore, considering a dsr size of 0.01 mm yr^{-1} , the faults budgets will be consumed after 500, 320 and 400 dsr respectively. The sliprate budget of each fault may be spent along individual faults and/or FtF rupture scenarios.

- To convert the target MFD, expressed in terms of rate of earthquakes, into moment rates (figure 1.1b). This target MFD will be used to pick the magnitude bin on which an increment dsr will be spent. Notice that the formulation in terms of moment rate implies that greater magnitudes are more likely to be picked.

4. Iterative process is described as follows:

- First, the bin of magnitude (of width 0.1) where a dsr will be spent is picked according to the target MFD for the whole fault system in terms of moment rate.
- Then, in this bin of magnitude M_i , a seismotectonic source Si (an individual fault or an FtF scenario) that can host this magnitude is picked randomly. The increment of moment rate $d\dot{M}_0$ for this source is calculated following Eq. (1), and the rate of earthquakes increment dre is calculated using Eq. (2).

$$d\dot{M}_0 = \mu \cdot A \cdot dsr \tag{1.1}$$

$$dr_e(M_i) = \frac{d\dot{M}_0}{M_0(M_i)}, \tag{1.2}$$

where $d\dot{M}_0$ is the increment of moment rate for the source Si, μ the shear modulus of the fault, A the area of the source, dsr the increment of slip rate spent, $dr_e(M_i)$ the increment of the rate of magnitude M_i, and $M_0(M_i)$ the seismic moment of a moment magnitude Mw defined by Hanks and Kanamori [1979]:

$$Mw = \frac{2}{3} log(M_0) - 10.7. \tag{1.3}$$

- At each iteration, the slip-rate budget of the faults participating in the scenario accommodating the earthquakes of the three highest magnitude bins (0.3 being the range of uncertainties in the scaling laws used to assess the maximum magnitude) is checked:
- If there is still a portion of the slip-rate budget to be spent, the dre calculated is added to the rate of earthquakes of magnitude M_i for the source Si.
- If one of the faults of the FtF rupture generating the largest earthquake has exhausted its sliprate budget, the final rates of the highest magnitude earthquakes are reached. Then, knowing the shape of the imposed target MFD, the target rate at the fault system level for all magnitudes bins is known (figure 1.1c). At this stage, an additional check is performed:

- if adding the dre calculated for magnitude M_i onto the source Si leads to the exceeding of the target MFD for this magnitude, then this dre is not added to the source Si and the increment dsr of this computation step is considered as NMS slip;
- if adding the dre to the source Si does not lead to the exceeding of the target MFD, this dre is added to the source Si.
- The increment of slip rate dsr is then removed from the slip-rate budget of the fault or the faults involved in source Si.
- If the slip-rate budget of each fault is exhausted, the fault and the corresponding FtF rupture scenarios the fault is involved in are removed and cannot be picked anymore in subsequent iterations of the computation.
- These steps are repeated until all the slip-rate budgets of all the faults in the system are spent either on single-fault ruptures, FtF ruptures or NMS slips.

The output of this process is an earthquake rupture rate for different magnitudes for each fault and FtF rupture scenario in the model, considered as aleatory uncertainty. We also record how the slip-rate budget of each fault is partitioned between the different FtF ruptures and how much NMS slip was needed on each fault in order to fit the target MFD shape (here GR MFD) with a given set of FtF rupture scenarios (figure 1.1d). In the example, fault 1 spends 43 % of its budget on single-fault ruptures (blue color), 7 % on F1 + F2 ruptures (dark green), 23 % on F1 + F2 + F3 ruptures (dark grey) and 27 % on the NMS slips (light grey). On the other hand, 100 % of the slip-rate budget of the slower moving fault (i.e., fault 2) is converted into earthquake rates (0 % NMS) and limits the rate of the largest magnitude earthquakes (F1 + F2 + F3) (see Annex 1).

A simplified example of the application of this methodology based on only two faults is presented in the Supplement to this paper (see Annex 1). This example illustrates step by step the way in which the proposed methodology allows the transformation of slip-rate budgets of faults into earthquake rates.

Post-processing then includes the following:

- Exploring the epistemic uncertainties: Many assumptions have to be made when setting up the methodology (scaling law, FtF rupture set, faults parameters, etc.), and the different possible hypothesis should be explored in a logic tree.
- Reality checks: The last step of the methodology involves comparing the modeled earthquake rates with independent data, such as the seismicity rates deduced from the catalog and paleoearthquake rates deduced from trench studies. Each branch of the logic tree is then weighted according to its performance with these independent data.

In this study, we applied the proposed methodology to a well-documented rifting zone in the Aegean Sea (figure 1.2).

1.1.4 Application to the western Corinth rift fault system

The east–west striking Corinth rift is the most seismically active structure in Europe with several earthquakes larger than Mw 5.5 recorded in the historical times as well as in the instrumental period (e.g., Jackson et al. [1982]; Papazachos and Papazachou [2003];Makropoulos et al. [2012]). The Corinth Rift Laboratory was set up in 2001 in the western and most seismically active part of the rift (Lyon-Caen et al. [2004]) with the goals of understanding the rifting process and providing key elements for the seismic hazard assessment of the region.

The geodetical deformation measured by Global Positioning System (GPS) shows a highly localized opening of the Corinth rift at a rate of 10 mm yr^{-1} in the east and 15 mm yr^{-1} in the west (Avallone et al. [2004]) over a width of less than 20 km, inducing a high strain rate. This deformation is accommodated by a complex network of both north- and south-dipping normal faults. Geological studies of these faults have shown that the north-dipping faults located on the southern coast have a higher slip rate than the south-dipping northern faults, giving the rift its asymmetrical structure. In the south, the Peloponnese is uplifted by the activity of these faults (Armijo et al. [1996]; Ford et al. [2013]), and in the north the coast line is subsiding.

The western Corinth rift fault slip rates were inferred from the displacement of geologic markers in the field or from seismic profiles on each individual fault, with the exception of the two blind faults identified by their recent seismic activity (1995 fault, Bernard et al. [1997] and Pyrgos fault, Sokos et al. [2012]) and for which the microseismicity recorded close to the fault was transformed into the slip rate on the fault plane. These latter slip rates are therefore subject to a very large uncertainty. The estimated geological extension rate expressed by the sum of the horizontal projection of the geological slip rates of the faults is in the range of 3 to 6 mm per year, which is three times less than the geodetical extension rate. Given this disagreement, the WCR is a good candidate to test whether the earthquake rates calculated using our methodology, which relies only on geological information, can account for the occurrence of the large earthquakes that have been documented in the region (Albini et al. [2017]).

The WCR fault system has been described in detail by Boiselet [2014], who defined a model for the fault system, including geometries and slip rates for each fault (figure 1.2, table 1.1) and a set of possible FtF ruptures (hereafter model B14). The B14 model proposes a set of FtF rupture scenarios (table 1.2) assuming that two neighboring faults can make up an FtF scenario only if they are less than 3 km apart. In this paper, we also explore a logic tree branch for an alternative rupture set (see figure 1.3) with higher fault connectivity (B14_hc), where faults can break together if their fault traces are separated by 5 km or less, therefore allowing a wider spectrum of possible FtF rupture

Fault name	ID	Length (km)	dip	Seismogenic depth (km)		Slip rate mm.yr^{-1}			Mmax		Time-frame of the data
				Upper	Lower	Min	Mean	Max	WC94	Le10	
Psathopyrgos fault	f1	8.5	60	0	6	4.6	5	5.5	5.8	5.7	2 kyr
Neos Erineos fault	f2	11.4	55	0	7	2.3	3.2	4.1	6.0	5.9	3–4 kyr
Aigion fault	f3	8.6	60	0	7	3.5	4	4.6	5.8	5.8	50–60 kyr
East Helike fault	f4	14.5	55	0	7	3	3.5	4	6.1	6.0	10–12 kyr
West Helike fault	f5	11.2	55	0	7	0.5	0.9	1.4	6.0	5.9	800 kyr
Trizonia fault	f6	10.6	65	0	7	1.3	1.4	1.5	5.9	6.0	800 kyr
West Channel fault	f7	10.8	45	0	2.5	0.4	0.45	0.5	5.6	5.5	240–400 kyr
South Eratini fault	f8	12	45	0	6.5	0.6	1	1.4	6.0	6.0	800 kyr
East Channel fault	f9	22	45	0	4.5	1	1.4	1.8	5.7	5.7	1500 kyr
North Eratini fault	f10	11.5	60	0	6	2.4	4	5.6	5.9	5.8	12 kyr
Marathias fault	f11	17.4	60	0	6.5	1.39	1.4	1.41	6.1	6.0	400 kyr
1995 fault	f12	14	35	8	12	0.5	3.2	7	6.0	6.0	5 yr
Pyrgos fault	f13	11	35	6	11	0.5	3.2	7	6.1	6.0	5 yr

Table 1.1: Fault characteristics in Boiselet [2014]. Mmax values calculated using the equations for normal faults using the rupture area. WC94 refers to Wells and Coppersmith [1994] and Le10 to Leonard [2010] scaling laws.

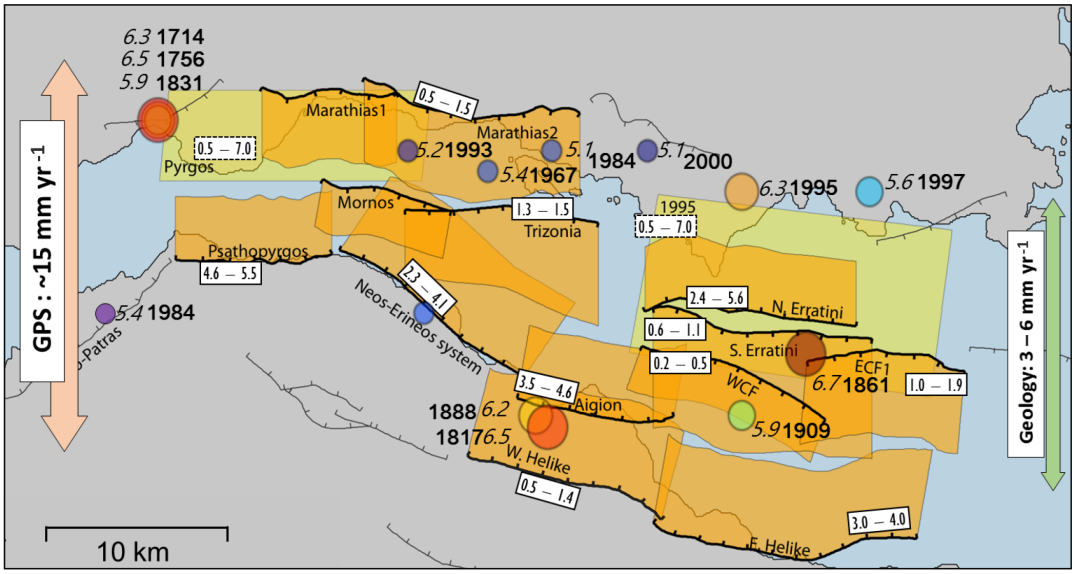

Figure 1.2: *Map of the active faults of the western part of the Corinth rift (modified from Boiselet [2014]). The orange polygons are the surface projection of the active faults. The yellow polygons are the surface projection of the blind faults (Pyrgos fault and 1995 fault). Earthquakes of the catalog during the complete period are represented by circles with color and size depending on the magnitude. The date and magnitude of earthquake are indicated. The minimum and maximum values (mm* yr^{-1} *) of the slip rates of the faults are indicated in the white boxes. The green arrow shows an approximation of the rift extension calculated by projecting horizontally the faults' slip rate, and the pink arrow shows the extensional rate of the rift measured by GPS.*

scenarios (additional scenarios in bold in table 1.2). For a comparison with classical fault PSHA studies, we explore a branch with only a simple fault rupture called B14_s. In this branch no FtF rupture is allowed.

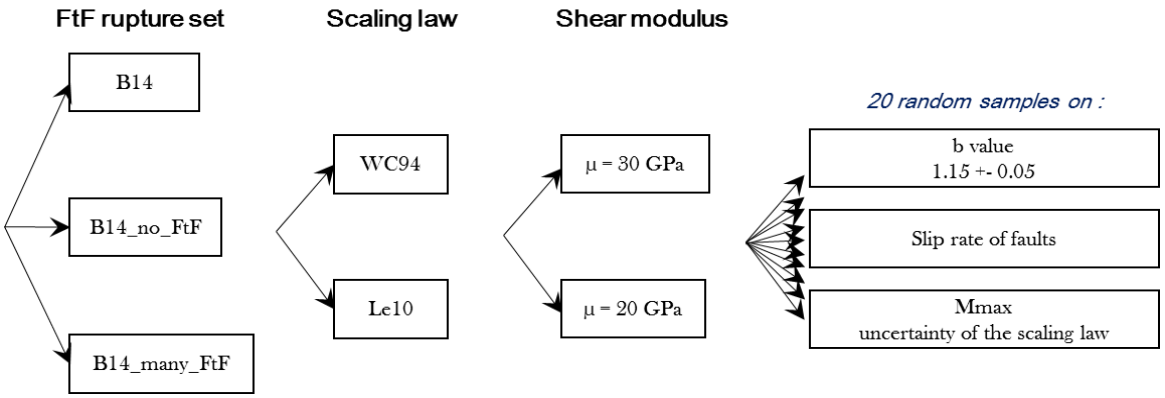

Figure 1.3: *Logic tree explored for this study.*

The target MFD is chosen to be in the shape of a GR MFD with a b value of 1.15 ± 0.05, which is a typical value for extensional systems (Schorlemmer et al. [2005]).

In this study we explore other epistemic uncertainties having a potential impact on the modeled earthquake

FtF sets	Faults involved in the scenario					Mmax	
						WC94	Le10
B14, B14_hc, B14_s	All the single-fault ruptures					see table 1.1	
3 km	f3	f2				6.2	6.2
distance	f3	f2	f1			6.4	6.2
criteria	f2	f1				6.2	6.0
	f4	f5				6.2	6.0
B14	f1	f13				6.2	6.2
B14_hc	f4	f12				6.4	6.2
	f4	f8				6.4	6.2
	f4	f8	f5			6.5	6.4
	f4	f8	f9			6.5	6.4
	f8	f9				6.2	6.0
5 km	f11	f6				6.3	6.3
distance	f11	f6	f1			6.4	6.5
criteria	f11	f6	f2			6.5	6.5
	f11	f6	f2	f1		6.6	6.5
B14_hc	f3	f5				6.2	6.2
	f3	f7				5.8	5.8
	f3	f9	f7			6.2	6.1
	f3	f8	f9	f7		6.4	6.5
	f3	f4	f2	f1		6.5	6.6
	f4	f7				6.2	6.2
	f4	f8	f7			6.4	6.5
	f4	f8	f9	f7		6.5	6.5
	f8	f10				6.3	6.3
	f3	f6	f2	f1		6.5	6.5
	f3	f6	f2	f1		6.5	6.5
	f3	f12				6.2	6.3
	f3	f4				6.3	6.3
	f8	f9	f7			6.3	6.2
	f3	f4	f5	f2	f1	6.6	6.7

Table 1.2: Rupture scenarios considered in each model. Branch B14_s considers only the single-fault ruptures. Branch B14 considers the single-fault ruptures and the FtF ruptures with a 3 km distance criteria. Branch B14_hc considers the single-fault ruptures and the FtF ruptures with a 3 and a 5 km distance criteria. Mmax values are calculated using the equations for normal faults based on the rupture area. WC94: Wells and Coppersmith [1994]; Le10: Leonard [2010]

rates (figure 1.3) in addition to the different FtF rupture sets previously described; two scaling laws (Wells and Coppersmith [1994]; Leonard [2010]) have been used to calculate the maximum magnitude based on rupture area for normal faults, and two values of the shear modulus μ have been used: 30 GPa, a commonly used value in hazard studies, and 20 GPa to represent the low-shear wave velocity in the WCR region recently estimated based on ambient noise tomography (Giannopoulos et al. [2017]). For each branch, 20 random samples are drawn from triangular distributions in order to explore the epistemic uncertainty affecting the b value of the target MFD (1.15 $\pm$ 0.05), the slip rate of the faults and the uncertainty within the scaling law.

1.1.5 Modeled earthquake rupture rates and comparison with independent data

Using our method, the modeled rate of earthquakes for the WCR is then compared to the rate of earthquakes observed in the catalog. The seismicity catalog considered in this study (figure 1.3) is the SHEEC catalog (Stucchi et al. [2013]; Grünthal et al. [2013]) developed in the framework of the SHARE project updated for six historical earthquakes (Albini et al. [2017]) and three instrumental earthquakes (based on Baker et al. [1997]; and P. Bernard from the 3HAZ Corinth project, personal communication, 2017). The updates and their implication on the catalog are summarized in table 1.3. We propagate the earthquake magnitude uncertainties in the estimate of seismic moment rate and earthquake rate calculations by randomly sampling the magnitude of each earthquake within their uncertainties (Stucchi et al. [2013]; Albini et al. [2017]) and by using two hypotheses of completeness. In table 1.4 the time of completeness of the catalog for Greece calculated by the SHARE project (Stucchi et al. [2013]) and the times calculated by Boiselet [2014] using the approach proposed by Stepp [1972] at the scale of the Corinth rift region are reported.

A first reality check is performed to compare the modeled and the catalog seismological moment rates. The seismological moment rate is calculated directly using the rates of earthquake of each magnitude in the catalog based on the moment magnitude relation (Eq. 3) The seismic moment rates in models B14 and B14_hc are in good agreement with the seismic moment rate deduced from the catalog, whereas B14_s predicts a higher seismic moment rate (figure 1.4a). This comparison provides a better confidence in the models where FtF ruptures are possible than in the B14_s model. In the single rupture model (B14_s), 90 % to nearly 100 % of the geological slip rate is converted into seismic moment rate with only less than 10 % interpreted as NMS slip rate. On the other hand, when FtF ruptures are possible (B14 and B14_hc), 25 % of the geological slip-rate budget of the faults is interpreted as NMS slip (figure 1.4b).

A second reality check consists in comparing modeled and catalog MFDs (figure 1.4c). The B14_s model does not manage to reproduce the rate of earthquakes deduced from the catalog, as it predicts a higher rate of Mw 5 earthquakes and no earthquakes of Mw 6.3 and above. On the other hand, we observe a good agreement of the MFDs of models B14 and B14_hc with the catalog. B14 reproduces the cumulative earthquake rate for Mw of 5.6 to 6.1 well, whereas model B14_hc reproduces the cumulative rate of earthquakes of Mw 5.0 to 5.5 better.

Date	Type of update	Old parameters	New parameters	Special implication for the catalog
14 May 1748	Magnitude	Mw = 6.4 ± 0.25	Mw = 5.9 (5.4–5.9)	Not in the complete period for this range of magnitudes
23 Aug 1817	Magnitude	Mw = 6.6 ± 0.25	Mw = 6.5 (6.0–6.5)	
26 Dec 1861	Location	(38.22, 22.139)	(38.28, 22.24)	Not associated with Aigion fault
9 Sep 1888	Magnitude	Mw = 6.3 ± 0.4	Mw = 6.2 (5.7–6.2)	
25 Aug 1889	Location and magnitude	(38.25, 22.08) Mw=6.24 ± 0.25	(38.50, 21.33) Mw=6.4 (6.4–6.6)	Earthquake outside the WCR
3 Mar 1965	Depth and magnitude	Depth = 10 km Mw = 6.5	Depth = 55 km Mw = 6.8	Earthquake associated with the subduction zone, not with the WCR fault system
15 Jun 1995	Location and magnitude	(38.37, 22.15) Mw=5.8	(38.36, 22.20) Mw=6.3	
5 Nov 1997	Location	(22.28, 38.41)	(22.28, 38.36)	

Table 1.3: Earthquakes updated in the historical and instrumental catalogs of the western Corinth rift.

SHARE project Woessner et al. [2015]		Boiselet [2014]	
Magnitude range	Date of completness	Magnitude range	Date of completness
4.1–5.1	1970	5.0–5.4	1958
5.1–5.7	1900	5.5–6.0	1904
5.7–6.5	1650	6.0–6.5	1725
$\geq$ 6.5	1450	6.5–7.0	1725

Table 1.4: Completeness hypotheses explored in this study.

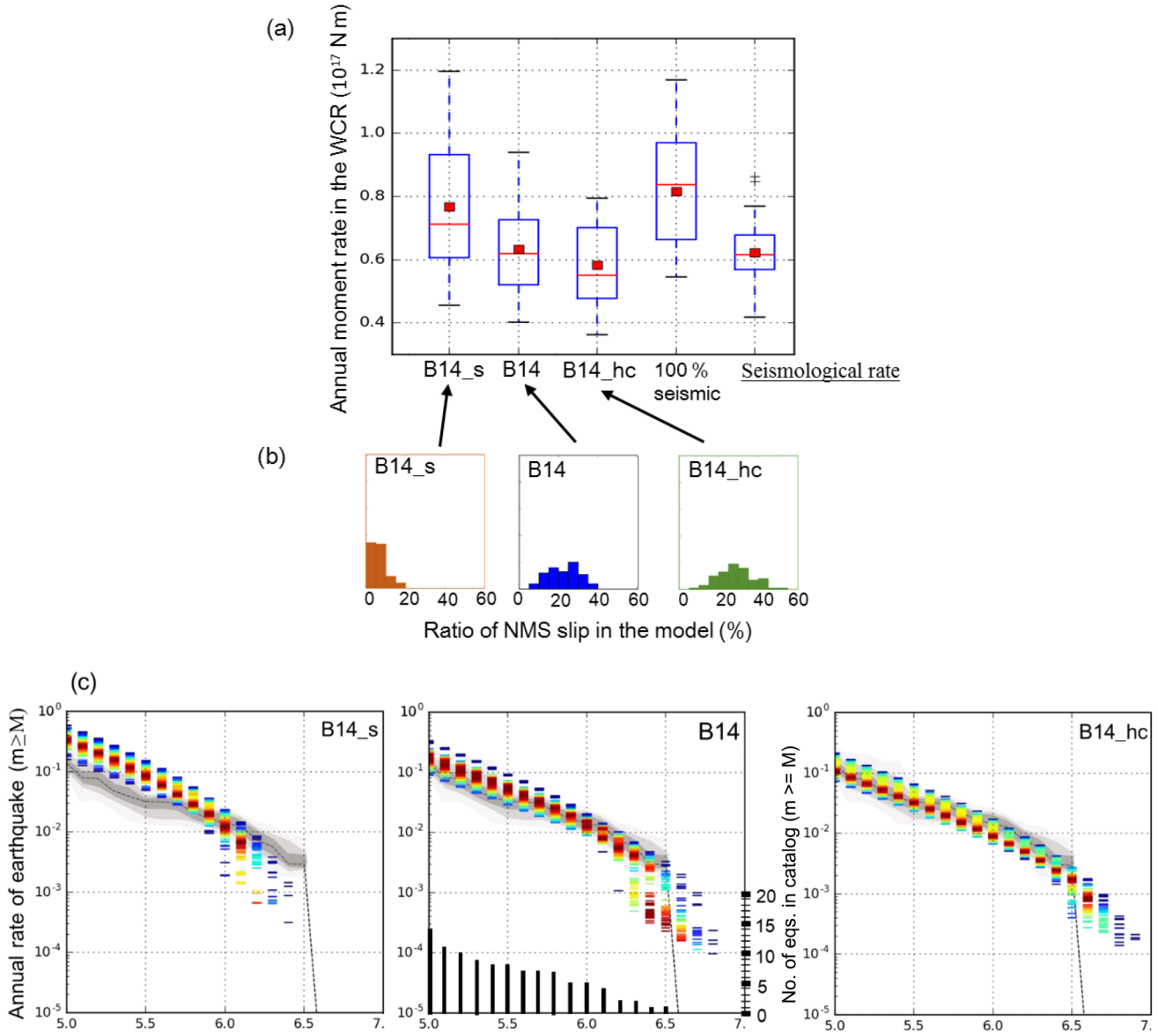

Figure 1.4: *Modeled seismicity for the WCR fault network and comparison to the seismicity rate based on the earthquake catalog of the complete period. (a) Comparison between the modeled moment rates for each FtF scenario set and the seismological rate calculated from the earthquake catalog. Each box represents the SD around a mean and median value represented by a red square and a red line respectively. From left to right: the three first boxes are for each hypothesis scenario set in the logic tree, the fourth box shows the moment rate assuming 100 % of the slip rate of faults is converted into seismic moment, and the fifth box shows the moment rate calculated from the earthquake catalog. (b) Distribution of the ratio of NMS slip resulting from the three deformation models. (c) Comparison between the modeled GR MFD deduced from geological data for the whole fault system and that deduced from the WCR catalog. The models are represented as a colored density function with the red colors for the rates predicted by the higher number of models. The cumulative rates calculated from the catalog are shown as a grey density function. The cumulative number of earthquakes in the catalog is indicated by black bars in the central figure.*

Slip-rate budget repartition

The way the slip-rate budget is spent between FtF ruptures, single-fault ruptures and the NMS slip ratio of the fault depends on the slip rate of the fault and the FtF ruptures the fault is involved in. Slow slipping faults that are involved in large FtF rupture scenarios (Neos Erineos or West Helike) have the majority of their slip-rate budget consumed by these large FtF ruptures (figure 1.5). On the contrary, the fast slipping faults that are involved in few FtF ruptures scenarios (1995, Pyrgos, North Eratini) spend their budget predominantly on single-fault ruptures producing a high number of small to medium magnitude earthquakes, which leads to easily exceeding the GR regional target and thus implies a higher proportion of NMS slip rate on these faults.

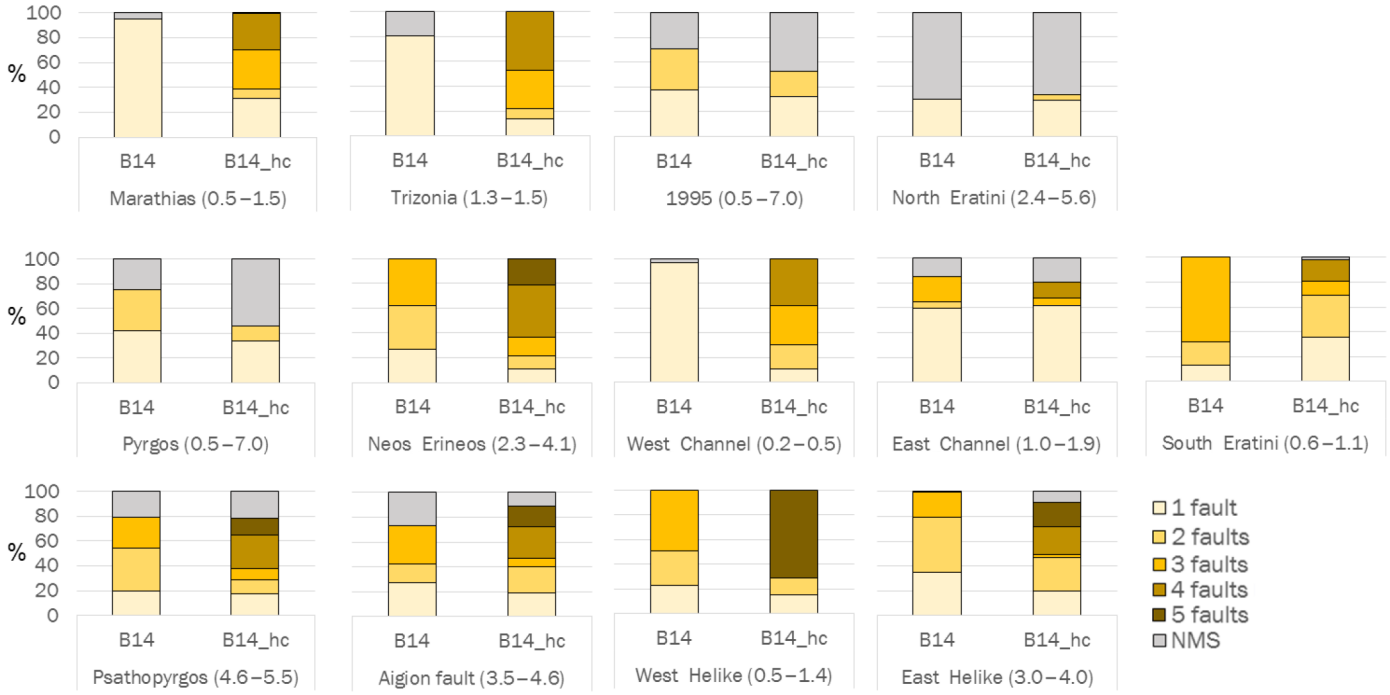

Figure 1.5: *Visualization of the way the slip-rate budget of each fault is spent. The color depends on the number of faults involved in the FtF rupture. Minimum and maximum values of the slip rate on each fault is shown in brackets in* $mm\ yr^{-1}$.

Models B14 and B14_hc have a similar mean 25 % ratio of NMS slip ratio (figure 1.4), but this ratio is not distributed in the same way between the different faults in each model. However, an important NMS proportion on the blind faults (Pyrgos and 1995-fault) and the offshore North Eratini fault is found for both models. There are three main factors that can induce this result: either the FtF sets are not realistic, the slip rates explored on those faults are not realistic and do not include enough complex ruptures with these faults, or there is a mechanism of NMS slip such as creep or slow slip events happening on theses faults.

Earthquake rupture rate on the Aigion fault

We now choose to focus our interest on the Aigion fault. Since this fault is one of the most active faults of the WCR and crosses the city of Aigion, it represents a major source of seismic hazard and risk for the region.

The earthquake rate modeled on the Aigion fault depends of the FtF rupture set allowed in the model (figure 1.6). The resulting MFD of the Aigion fault has the shape of a GR MFD for model B14 and B14_s, with a steeper slope for the B14_s model. In the B14_hc, the MFD computed for the Aigion fault resembles a characteristic earthquake distribution of magnitude close to Mw 6.0, similar to the maximum magnitude of earthquakes rupturing only the Aigion fault. It is worth noting that the larger magnitude earthquakes in figure 1.6b and c involve not only the Aigion fault but also the neighboring faults participating in the FtF ruptures (figure 1.5, table 1.2).

Using the paleoseismological data presented by Pantosti et al. [2004], it is possible to propose rates of large magnitude earthquakes on the Aigion fault (figure 1.6). This paleorate (rate of paleoearthquakes) is subject to large uncertainties but can be used to validate or invalidate the different FtF rupture set hypothesis. In the B14_s model where faults only break on their own, the Aigion fault is not able to accommodate the paleoearthquake magnitudes. In the B14 model, where fault rupture is only allowed between faults separated by 3 km or less, the modeled earthquake rates are lower than the rates inferred from the paleoseismological study. In the B14_hc model, where FtF ruptures are allowed for faults separated by 5 km or less, the modeled earthquake rates agree well with the paleorate, within the margin of uncertainty.

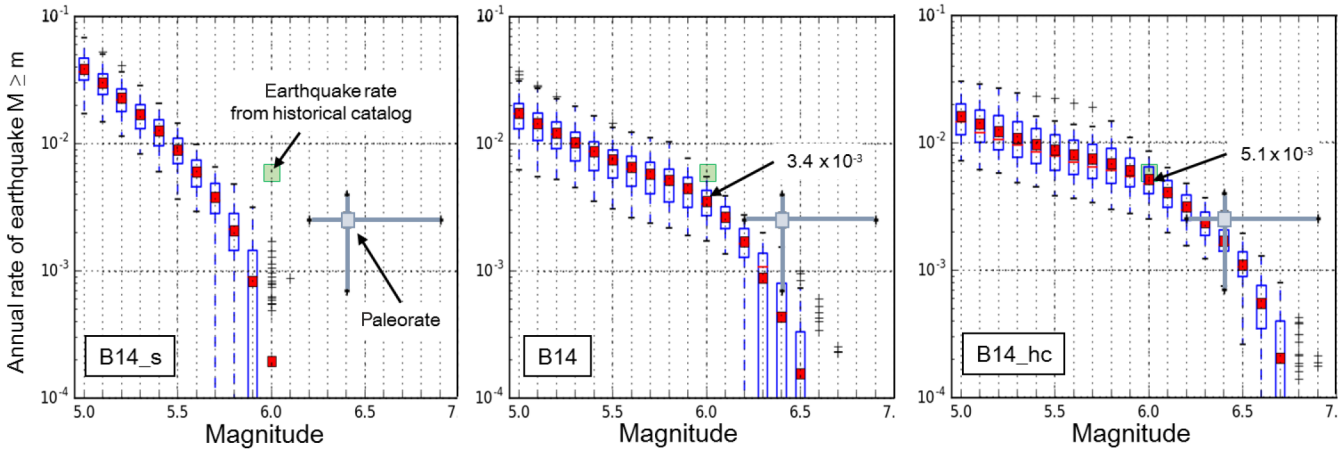

Figure 1.6: *Rate of earthquakes occurring on the Aigion fault for each FtF rupture set. Variability resulting from the exploration of the logic tree is illustrated by the blue boxes. The annual rates of Mw ≥ 6 earthquakes on Aigion fault are indicated for the B14 and B14_hc models. The grey square represents the paleorate (rate of paleoearthquakes) interpreted from Pantosti et al. [2004] with its uncertainties. The green box represents the rate of earthquakes greater than Mw 6 on the Aigion fault inferred from the historical catalog.*

According to the recent reappraisal of the historical seismicity (Albini et al. [2017]), the Aigion fault is most

likely the source of the 1817 Mw 6.5 (6.0–6.5) and the 1888 Mw 6.2 (5.7–6.2) earthquakes. This leads to estimates of annual rates of Mw $>$ 6 earthquakes on the Aigion fault of 0.005 to 0.007 (figure 1.6) depending on the date of completeness of the catalog used (table 1.4). The model B14_s does not manage to reproduce the great magnitudes earthquakes observed in the catalog. The annual rates for earthquakes Mw $>$ 6 of 0.0034 and 0.0051 predicted by models B14 and B14_hc respectively are statistically compatible with the rate inferred from the catalog.

Weighting the logic tree branches

The comparison with independent local data allows suggesting weights for the different FtF rupture set hypothesis (figure 1.3) for hazard calculation.

The B14_s branch, where faults can only rupture independently, fits neither the annual moment rate nor the earthquakes rate of the catalog of the region, nor the paleoearthquake magnitude on the Aigion fault (figure 1.4 and figure 1.6). We conclude that this branch should not be used for a hazard calculation in the western Corinth rift.

Between the two branches where FtF ruptures are possible, B14_hc manages to match the earthquake rates of the catalog for a range of magnitudes where statistics are stronger (14 earthquakes Mw 5.0 and above) compared to the B14 model (matching only 4 earthquakes of Mw 6.0 and above in the catalog) (figure 1.4). The B14_hc branch matches the Aigion fault earthquake rates inferred from the paleoseismology and the historical catalog better than the B14 model (figure 1.6). The agreement with the earthquake rate in the regional catalog and the better reproduction of the Aigion fault data of the B14_hc model lead us to propose a stronger weight for this model compared to the B14 model for the estimate of hazard for Aigion city.

1.1.6 Conclusion and perspectives

The methodology presented in this study uses a system-level approach rather than an individual-fault-level approach to estimate the rate of earthquakes on faults based on the geological data collected for each fault and allowing FtF rupture in the hazard model as an aleatory uncertainty. The application of the methodology to the WCR fault network shows that, in order to match a GR MFD for the whole fault system, part of the fault slip rates have to be spent as a non-main-shock slip. The way the fault slip rate is partitioned among single or FtF ruptures and the resulting shape of the individual fault MFD depends on the location of the fault in the network and the fault's characteristics. The earthquake rates modeled using the geological data on the faults are compared with the local earthquake catalog and paleoseismic data in order to weight the different epistemic hypothesis. In the case of the WCR, and for future seismic hazard assessment for the city of Aigion, these reality checks suggest attributing a stronger weight to the branch allowing FtF ruptures between faults with the 5 km distance criteria (B14_hc), a lower weight to that based on the 3

km criteria (B14) and a null weight to the model where only single-fault ruptures are allowed (B14_s).

The fault network used for the application concerns only the western part of the Corinth rift fault network. Integrating the rest of the network in the model could modify the final outcome and should be explored in future developments.

More reality checks will be implemented in the future in order to weight the different uncertainties of the logic tree based on the results of the ongoing microseismicity studies in the WCR (i.e., use the possible presence of repeater earthquakes on the Aigion fault to validate NMS slip ratio, Duverger et al. [2015]).

The methodology presented in this article can be applied to other fault systems in different tectonic environments. In order to implement this approach, the geometries and slip rates of the faults have to be known within uncertainties, FtF rupture scenarios sets have to be defined, and the shape of the regional MFD needs to be assumed or inferred from the regional catalog. If for the WCR the GR distribution seems suitable, it has been shown that a Youngs and Coppersmith distribution (Youngs and Coppersmith [1985]) can be more appropriate for other fault systems (e.g., by Hecker et al. [2013]). Given the flexibility of our methodology any other target MFD can be easily implemented in the methodology.

The earthquake rupture rate calculated using this methodology is very sensitive to the choice of possible FtF rupture scenarios. The comparison with the earthquake catalog and local data, such as the paleoseismological data, can provide guidance to the strength of each hypothesis. Nevertheless, the choice based on distance between faults should be supported by more physical approaches in the future such as Coulomb stress modeling (Toda et al. [2005]) and/or dynamic modeling of ruptures (Durand et al. [2016]).

The methodology at this stage does not consider the background seismicity. The example of the dense WCR fault system allowed setting aside this issue in order to test our methodology and focus on the FtF ruptures. Future developments of the methodology need to allow part of the modeled seismicity rate to be in the background. If performing hazard calculation for a region wider than the fault system itself, it is necessary to combine the models built with this methodology with classical area sources.

The following section is an article (Chartier et al. [2019]) presenting the evolution of the methodology since Chartier et al. [2017b] and the code (SHERIFS) that applies the methodology. The aim of this following article is to be a support for a user of the code, explaining in detail the algorithm rather than focusing on a specific application.

1.2 SHERIFS – Open-Source Code for Computing Earthquake Rates in Fault Systems and Constructing Hazard Models

1.2.1 Abstract

Modelling the seismic potential of active faults and their associated epistemic uncertainties is a fundamental step of probabilistic seismic hazard assessment (PSHA). SHERIFS (Seismic Hazard and Earthquake Rate In Fault Systems) is an open-source python code that builds hazard models including earthquake ruptures involving several fault sections or Fault-to-Fault (FtF) ruptures. It contains user-friendly tools to calculate the annual rate of FtF ruptures in a fault system based on the slip rate estimates and accounting for associated background seismicity.

SHERIFS applies a forward incremental approach following three rules: (1) the FtF ruptures allowed in the fault system are defined as input by the user and explored randomly, (2) the magnitude frequency distribution (MFD) of the modelled seismicity in the fault system must follow an imposed shape, (3) the slip-rate budget attributed to each fault section must be preserved in the calculation if the first two rules allow it. Indeed, in some cases, a fraction of the slip rate budget must be considered as being spent in Non-Main-Shock events (NMS) such as creep or post-seismic slip. Background seismicity rates are defined by the hazard modeler as the ratio of on-fault/off-fault seismicity for different ranges of magnitude.

Given a coherent set of input hypotheses, SHERIFS allows end-users to build the seismic hazard fault model thanks to an interactive user-friendly interface. It aims to help interactions between field data collectors and hazard modelers to explore and weight epistemic uncertainties affecting the input hypotheses. To do so, SHERIFS includes tools to compare modeled earthquake rates to the available local data (earthquake catalog and paleoseismological data). This comparison can be used to weigh different hypotheses explored in a logic tree and discard the hypotheses that are not in agreement with the data. SHERIFS's outputs are in a format that can be used directly as inputs for PSHA in the OpenQuake Engine (Pagani et al. [2014]).

1.2.2 Introduction

Developing a hazard model for a region where active faults have been described requires to convert the geological and geophysical data into 3D fault geometries and slip rates and then into earthquake rates. Most commonly seismic hazard modelers have to develop their own approach to manage the available data. Each modeler has a different approach for translating slip rates into earthquake rates on faults which sometimes does not allow a straightforward understanding of the resulting hazard model. Some assumptions have to be made when building a hazard model and, in the ideal case, the data collectors are involved in the modeling process to insure the working hypotheses are acceptable.

In order to help data collectors build hazard models, some tools are available (FiSH, OpenQuake toolkits). The FiSH (Pace et al. [2016]) modeling approach provides estimates of earthquake rates and time dependent probability of earthquakes in the period of interest for the hazard assessment using the data available on a fault. This MATLAB code can easily be operated by a user with limited coding skills. FiSH can be used to explore a range of epistemic uncertainties affecting the earthquake rates of the modeled fault. Since the approach in FiSH focuses on one fault only, multi-fault ruptures can only be modeled by considering average parameters for the whole structure. The OpenQuake toolkits allow building hazard models using as input shapefiles containing information on the faults but considering as well a fault by fault approach and not a fault system approach.

Thus, in most hazard models (e.g. Woessner et al. [2015]; TEM Wang et al. [2016]), small faults are usually grouped into large structures in order to allow for larger magnitude earthquakes. Moreover, background earthquakes are handled using a truncated approach in which earthquakes with a magnitude lower or equal to M6.4 occur only in the background zone with a rate defined by the rate in the earthquake catalog whereas magnitudes higher than M 6.4 are located on the faults with a rate defined using the average slip-rate of the fault. This approach allows integrating faults within area zones while avoiding double counting. However, it can lead to discontinuities in the regional MFD since the rate of low magnitude earthquakes and the rate of high magnitude earthquakes are calculated from different sources of information (Danciu et al. [2018]). Last but not least, magnitudes stronger than M6.4 may potentially occur in the background as well, especially in regions of slow tectonic deformation and in the presence of concealed faults (e.g. Dartfield earthquake, Hornblow et al. [2014]).

It has been known for a long time that several fault sections of a fault or of several faults can rupture during a single FtF rupture event (e.g. Knuepfer [1989], Barka and Kadinsky-Cade [1988] and Wesnousky [1988]). It is therefore paramount to have modelling procedures that allow a large set of possible FtF ruptures to occur in an aleatory fashion in the hazard model while reflecting the individual slip-rate of each section.

The Californian model UCERF3 (Field et al. [2014]) tackles this issue and allows FtF ruptures to occur in the model by relaxing the segmentation assumption and treating the problem at the faults system level. The annual rate of FtF ruptures is inverted in a "Grand Inversion" process considering the large amount of data available in California. UCERF-3 allows all possible ruptures in the fault system that follow a set of plausibility rules. The individual slip-rates of each section, the earthquake rates at the level of California and at the level of the paleo-seismic sites are some of the constraints used to invert for the rate of each rupture. This model also allows small earthquakes to happen on faults and large earthquakes to happen in the area sources by defining off-fault seismicity without appealing to the artificially truncated distribution commonly used.

In this paper we propose an alternative approach to calculate the earthquake rates in a fault system that preserves the main philosophy of UCERF-3. The aim of the proposed approach, named SHERIFS (Seismic Hazard

and Earthquake Rates In Fault Systems), is to provide a user-friendly tool with a graphical interface that allows data collectors and fault modelers to build hazard models in a forward processing scheme with an exploration of the epistemic uncertainties. The approach consists in the conversion of the slip-rate budget of each fault of the system into rupture rates for all possible single fault and FtF ruptures. SHERIFS relies on a forward modeling incremental method with a simple lay out of the assumptions about the input model which leads to an easy construction of the logic tree. The only required input data are an estimate of the geometry and slip-rate of the modeled faults.

The theoretical background of the methodology is presented in Chartier et al. (2017).

In this paper we present how this methodology has been implemented in SHERIFS, an open source tool coded in the Python language. The objective of SHERIFS is not only to provide the optimal hazard model but rather to allow the exploration of a large range of epistemic uncertainties and we recommend using it as such. By providing an easy-to-use tool for modeling faults in hazard models, we wish to promote discussions and feedbacks between data-providers and hazard modelers, one of the main objectives of the Fault2SHA ESC working group (`https://Fault2SHA.net/`). SHERIFS includes a set of tools allowing to compare the modelled earthquake rates to the local earthquake catalog or paleoseismological data and to discuss the different input hypotheses. Thus, SHERIFS can be useful not only to hazard modelers but also to geologists as a means to test the coherency of the rupture hypotheses that best describe the seismic potential of a fault system.

This tool creates hazard models for fault systems that can be used as inputs for OpenQuake (Pagani et al. [2014]) but since most PSHA computing codes use similar entries, the output of SHERIFS could easily be adapted to provide inputs for other codes as well (e.g. OpenSHA, Field et al. [2003]; CRISIS, Ordaz et al. [2017])

1.2.3 General methodology applied in SHERIFS

The SHERIFS code allows end-users to build the fault model thanks to an interactive user-friendly interface. The logic tree structure and the input files can be easily modified for a re-run if different parameters need to be tested.

This section of the article synthesizes the main steps of the methodology implemented in SHERIFS as presented in Chartier et al. [2017b]. This methodology requires as input data fault section geometries and slip-rate estimates for all the faults of the fault system in order to calculate the earthquake rupture rates.

The SHERIFS iterative methodology requires first establishing possible FtF earthquake rupture scenarios in the fault system and fixing the shape of the target MFD for the entire fault system. At each iteration, a magnitude is randomly picked based on the target MFD and an earthquake rupture scenario able to host such magnitude is selected. For this scenario, an increment of the slip-rate budget of the involved faults is converted into earthquake rupture rate

for the considered magnitude. At the beginning of the iterative process, the absolute value of the target MFD is not known, only the shape is imposed. In SHERIFS we explore different criteria to set the MFD target value that are detailed below. The iterative computation continues until the slip-rate budget of the faults is exhausted.

Required information and input working hypotheses

At first, the limits of the fault system have to be set. If the specific extent of the fault system is clearly limited in the field, we suggest using the enclosing boundary. In some applications, the fault system can be much larger than the area of interest for the hazard study. In such a case, the part of the system considered should be large enough to make sure that an extension of the system would not affect the rupture rates inside the zone of interest. If the fault network of a region is composed of distinctive sub-regions where it is believed that no rupture can go through, we suggest treating them as different fault systems.

Within the defined fault system, the geometry (trace, dip, and upper and lower seismogenic depth), the slip-rate (with uncertainties) and the kinematics (normal, strike-slip, reverse or the rake) of the fault sections must be provided. For partly creeping faults, we suggest the user to correct the geological/geodetic slip-rate input to reflect the "seismic" slip-rate that should be considered in SHERIFS. Alternatively, the shear modulus and/or the geometry of the creeping sections can also be adapted in order to model only the locked part of the fault.

We suggest defining typical lengths of fault sections to be on the order of the seismogenic depth. Faults with a longer length should be cut in smaller sections (that can rupture together). The sections should be defined as small as necessary in order to account for local information. For example, a change in slip-rate, rake or dip should be considered for defining the sections. The SHERIFS approach is based on the assumption that the seismicity of the fault system taken as a whole follows the shape of a known MFD (i.e. Gutenberg and Richter [1944] hereafter GR or Youngs and Coppersmith [1985], hereafter YC). If the local seismicity data doesn't show clearly a distinct shape, SHERIFS allows the user to explore several hypotheses in a logic tree.

a) Hypothesis on the FtF ruptures One of the most important working hypothesis of the SHERIFS methodology concerns the definition of possible FtF ruptures allowed in a given fault system. SHERIFS doesn't impose the way to choose the set of FtF ruptures, and only requires a list. The user is free to use rules (as done in UCERF3) or rely on physical simulations to define possible FtF ruptures. As the rules to define FtF ruptures are subject to discussion (i.e. Schwartz [2018]), SHERIFS allows to explore different sets of possible FtF ruptures.

In order to implement in SHERIFS single and FtF ruptures following the OpenQuake formalism, a single fault rupture is defined in the input files with either the "simple" or "complex fault" typology, depending on the user's choice,

and can host magnitudes starting from the minimal magnitude up to the maximum magnitude that the single fault can accommodate (scaling law- dependent value). The "simple fault" typology is defined by the fault trace, the dip and the upper and lower seismogenic depth and the "complex fault" typology allows more details in the fault geometry such as listricity or seismogenic depth variable along strike. A FtF rupture, on the other hand, is defined with the "characteristic fault source" typology, hosting only the greater magnitudes activating the entire fault surface included in the FtF rupture.

b) Sampling the model space of uncertainties The slip rate of the faults, the magnitude scaling parameters and the b value are usually associated with important uncertainties. In SHERIFS, uncertainty bounds are considered and explored based on the user-defined number of random samples. For each branch of the logic tree, SHERIFS will create a number of models equal to the number of random samples. For each model, a slip-rate value is picked uniformly within the uncertainty bounds of the slip-rate attributed to each fault, the parameters of the scaling law are picked independently following a Gaussian distribution within their error bounds; and a b-value is picked within the box-car range of b-values considered by the user. The first sample provided in the output file (model number 1) is always the mean value of the slip-rate, scaling law and b-value parameters.

In order to explore slip-rate value uncertainties of neighboring sections, the user can choose between a "correlated" or "random" sampling method. If the "random" option is chosen, slip-rate values are selected in the distribution without consideration of the values sampled in neighboring fault sections. Therefore, a section of the fault can be sampled in the lower part of its distribution when its neighboring section of the same fault can be sampled in the upper part of this distribution. If the "correlated" option is chosen, sections that can rupture together in FtF scenarios will always be sampled in the same part of their distribution. In practice, the distribution of possible slip-rate on each fault is divided in four quarters and if two faults can rupture together, their slip-rates will be sampled in the same quarter of the distribution for this model. We suggest that the "correlated" option be used when working with long fault systems where long faults have to be discretized in smaller sections. In the case where two very different opinions concerning the slip-rate of the faults are available, for example slip rates based on geological or geodetic considerations, we suggest building two different models rather than exploring these options in one single distribution.

Concerning the magnitude scaling relationships, the user can choose a set of published options (Wells and Coppersmith [1994], Leonard [2010], Thingbaijam et al. [2017]) and explore them in a logic tree. For some options, it is possible to choose which dimension of the fault to use (length or area). A new branch of the logic tree will be created for each chosen option. Within this branch the uncertainty affecting the estimate of the maximum magnitude is explored by random sampling of the parameters of each scaling relationship.

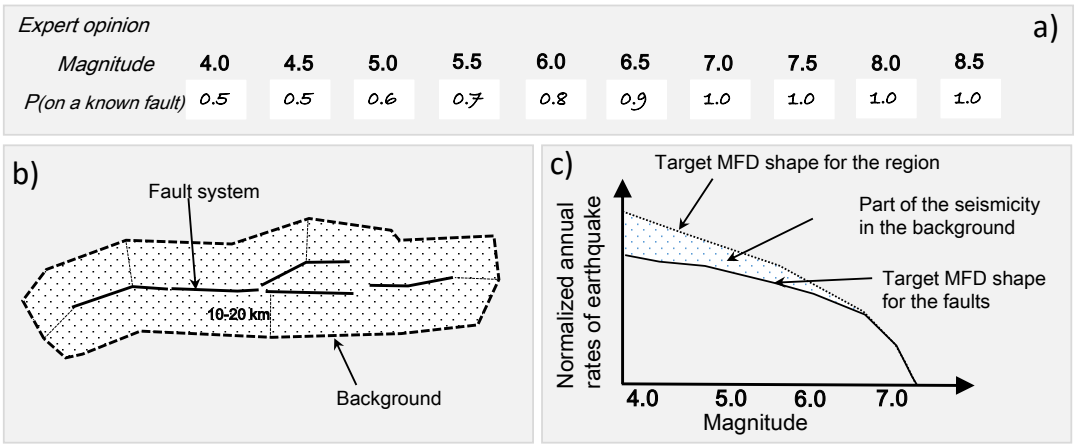

Figure 1.7: *Modeling of the background seismicity in seismic hazard and earthquake rate in fault systems (SHERIFS). (a) Expert opinion of the probability of a future earthquake to occur on a modeled fault for magnitude bin. (b) Example of background geometry. (c) Target magnitude–frequency distribution (MFD) modified to take into account the background seismicity.*

c) Setting the background Fault-based PSHA models usually take into account the possibility of an earthquake occurring on an unknown fault by defining a background zone around the faults. The seismicity rate of this background zone is usually deduced from the catalog and follows a GR distribution between the minimum magnitude Mmin and a truncated magnitude Mt (usually 5.8< Mt <6.5). The earthquakes with a magnitude larger than Mt are confined to the modelled faults and their rates are defined by the slip-rate and geometry of the faults.

In SHERIFS, earthquake rates are deduced from the geological information and are partitioned between background and faults (figure 1.7a). In order to define the background zone geometry the user can rely on simple distance criteria (for example, 10 or 20 km distance from the faults) to draw the background zone around the fault system (figure 1.7b). The background geometry and properties should be defined specifically for each model considered for the fault system. The seismicity rate of the background is defined by the user by setting the ratio, R, of the seismicity occurring on the modelled faults versus in the background. In the graphical interface this ratio has to be set for each magnitude M>Mmin of the distribution and is likely to increase with magnitude. To help set this ratio the user may rely on analysis of the distance between earthquakes and modeled faults. In SHERIFS, the MFD for the whole system has been defined to follow a target shape. The ratio R defined by the user for each magnitude bin is used to deduce the target shape for the faults. At the end of the iterative process, when the absolute MFD value of the fault system is known, the MFD of the background is then calculated using this ratio (figure 1.7c).

Since R is very difficult to determine, several background hypotheses are meant to be explored. This straight forward definition of the background has been introduced in SHERIFS since the version used in Chartier et al. [2017b]. We hope to encourage the discussion about the expert opinion on the matter of the background seismicity and its effect

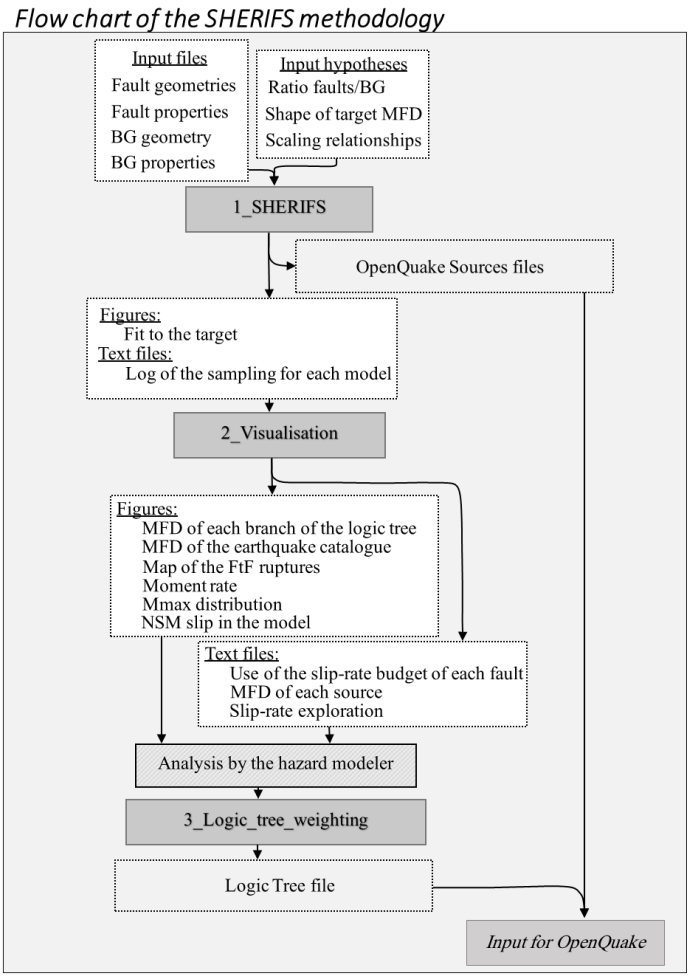

Figure 1.8: *Flowchart of the overall SHERIFS methodology. BG, Background; FtF, fault-to-fault; NMS, non-mainshock.*

on the seismic hazard assessment.

Workflow of the SHERIFS tool

SHERIFS workflow includes three tools that should be used sequentially (figure 1.8). The first tool sets up and runs the SHERIFS calculation and the logic tree exploring the epistemic uncertainties with the help of a graphical interface. The second tool allows visualizing the hazard model and comparing the modeled earthquake rates to available data. After a critical analysis of this comparison, the third tool allows the end-user setting the weights of the logic tree branches. The final outputs can then be directly used to perform probabilistic seismic hazard assessment.

Once the necessary information are in the input files and the python file1_SHERIFS.py has been filled with information concerning the run, SHERIFS opens a series of windows using a graphical interface. The purpose of these windows is to help the user build the logic tree and the input files necessary for the PSHA calculation. The detail of each

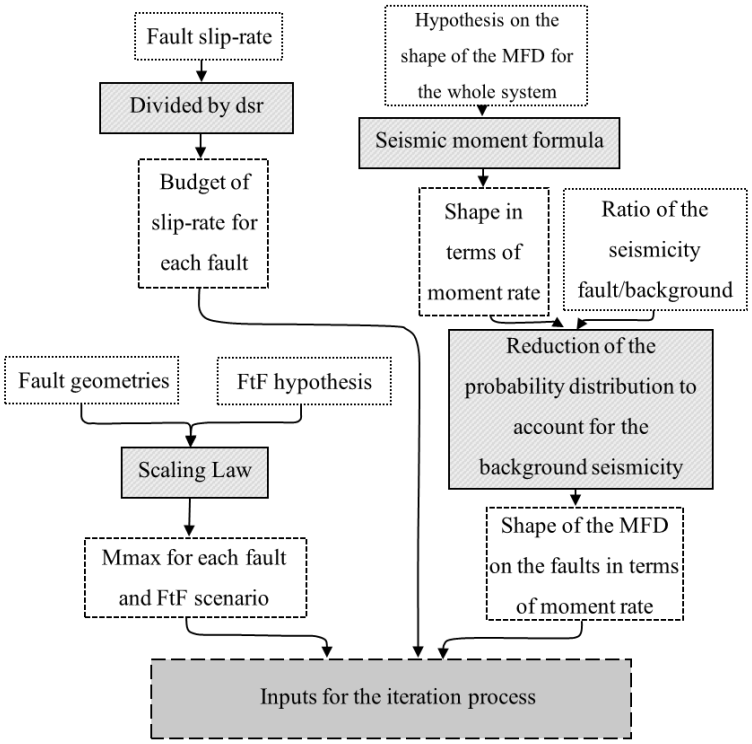

Figure 1.9: *SHERIFS data processing workflow. Dsr, sub-divided in slip rate.*

of the windows is explained in the user manual (provided with the code: `www.github.com/tomchartier/SHERIFS`). It is worth noting that the use of the graphical interface is not required by SHERIFS; an advanced user with a good knowledge of the required information and files structure can run SHERIFS independently of the graphical interface. It is also possible for the user to modify some hypotheses of the calculation or add or delete a logic tree branch and rerun the calculation.

The workflow for the calculation of the earthquake rates is described in figure 1.9 and figure 1.10.

The first steps consist in formatting the inputs provided by the user in order to run SHERIFS (figure 1.9). The slip-rate budget of each fault is sub-divided in slip rate increments (dsr). The maximum magnitude is deduced using the geometry of each section of the maximum FtF rupture scenario and the scaling law. The shape of the target MFD is then corrected for the proportion of the seismicity assumed to occur in the background and then converted in terms of moment rate.

SHERIFS iterative methodology spends the slip-rate budget of each fault until the target MFD is reached and the slip-rate budgets of each fault are exhausted (figure 1.10). In some cases, the target is reached before the budget of all faults is exhausted; the remaining slip-rate budget is then considered as (Non-Main Shock (NMS) slip and not converted into earthquake rates. The proportion of NMS slip on a specific fault of the system is an output of the methodology resulting from the combination of faults parameters, the set of FtF scenarios explored and the imposed

shape of the regional MFD.

More specifically, at each iterative loop, one magnitude bin of size 0.1 is randomly picked using the MFD defined in terms of moment rate and a source (single fault or FtF rupture) able to host such magnitude is selected (figure 1.10). If the target is not already set or not reached for this bin of magnitude, the rate of this magnitude for the picked source is increased and the slip-rate budget of the fault or faults involved in the source is reduced by one dsr increment; if the target is reached, the dsr increment is considered as NMS slip.

The selection of fault or FtF scenario that can host the picked magnitude is realized with faults preferentially picked according to the ratio of remaining slip rate budget over the slip-rate initial budget. This ad hoc rule has been set up in order to help isolated faults that don't break in any large scenarios spend their slip rate budget.

The absolute value of the target MFD is set when the system reaches one of the following:

1 - the budget of at least one of the faults participating in the largest magnitude earthquakes is exhausted. The target is then set because it is not possible to change the rate of earthquakes of the three largest magnitude bins, since the budget of one of the faults participating in the largest magnitude FtF ruptures is exhausted. Using the shape of the regional target MFD, the absolute value of the target earthquake rate for all magnitude bins can then be set (figure 1.11a). The choice of the three largest bins was made in order to allow flexibility around the target shape for the largest bins and avoid that a single odd rupture scenario in the last magnitude bin limits the rates of the whole system.

2 - the rate of the two largest magnitude bins is limited by the budget of a fault being exhausted but the third largest one is not. The annual rate on the latter bin is then limited to twice the rate of the mean of the two largest bins in order to limit the differences of earthquake rates between the last magnitude bins and keep the consistency of the target MFD shape set (figure 1.11b).

3 - the seismic moment required in order to fit the target shape and the rate of the largest magnitude earthquakes is equal to the remaining seismic moment in the system at a given moment of the iterative process. This ad hoc rule then sets the target in order to ensure that the rate of the smaller magnitude earthquakes will also reach the set target (figure 1.11c). Rule 2 and 3 were newly introduced since the version of the code used in Chartier et al. [2017b] in order to better model the target shape of certain fault systems proposed by some beta users. Additional novelties introduced in the latest version of SHERIFS are discussed in the electronic supplement (see Annex 1).

SHERIFS will check if the resulting MFD of the fault system matches the shape of the target imposed by the user. In the event that the shape doesn't match, which occurs when the slip-rate increment is not small enough, then the model is rerun with a slip-rate increment divided by two. After three re-runs, the model is accepted but a warning is displayed.

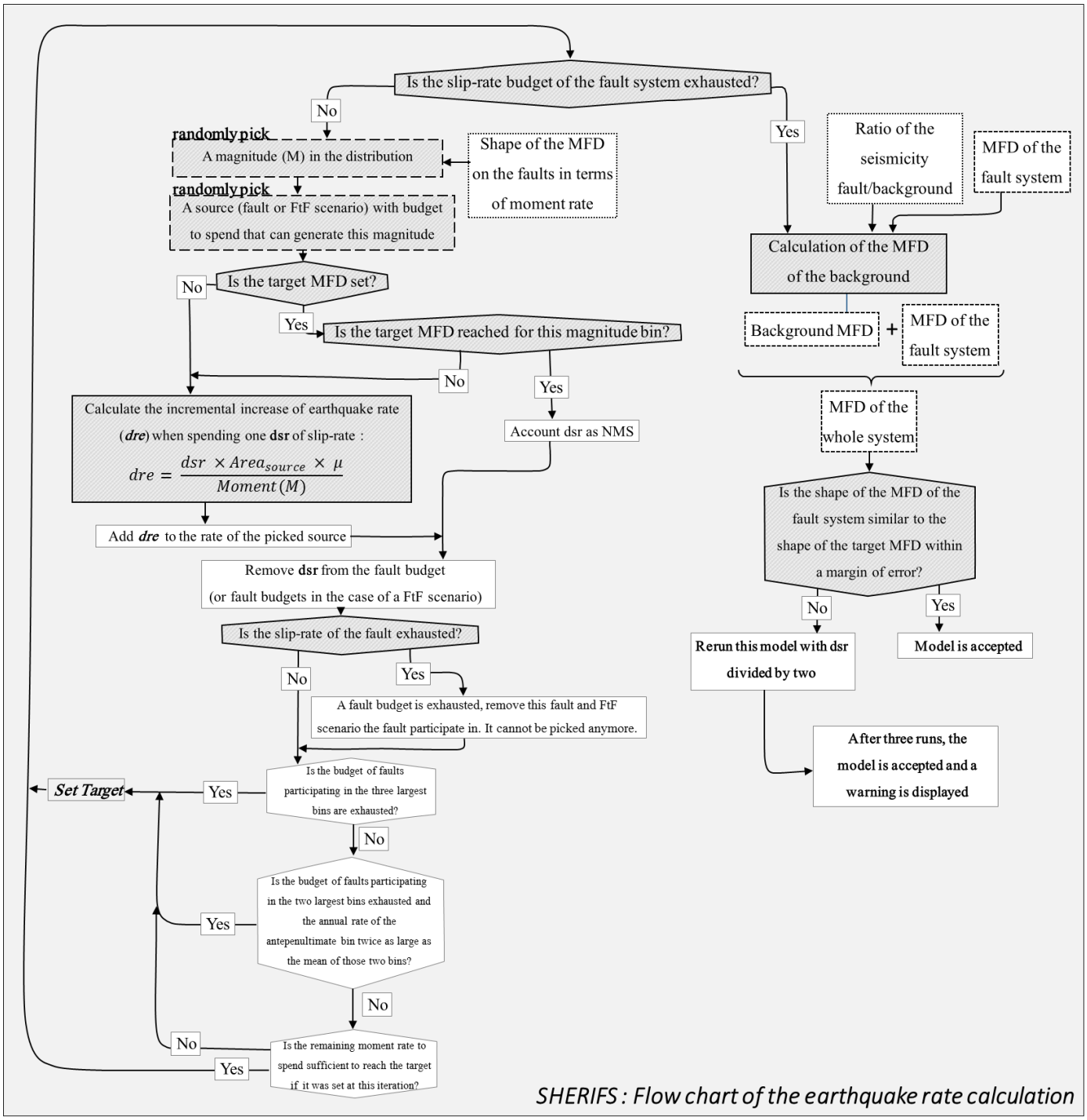

Figure 1.10: *SHERIFS iterative process workflow to compute earthquake rates on faults.*

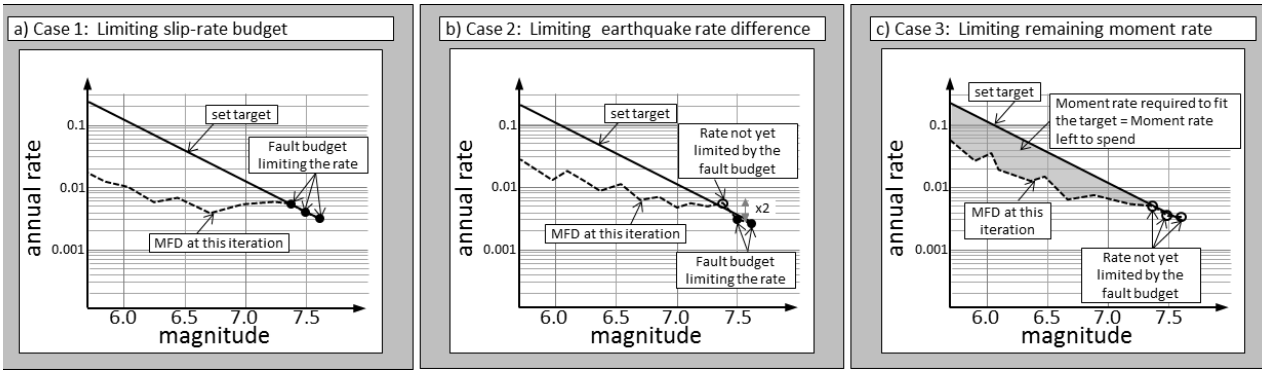

Figure 1.11: *The three cases of setting the target in the iterative process. (a) Case 1: limiting slip-rate budget, (b) case 2: limiting earthquake rate difference, and (c) case 3: limiting remaining moment rate.*

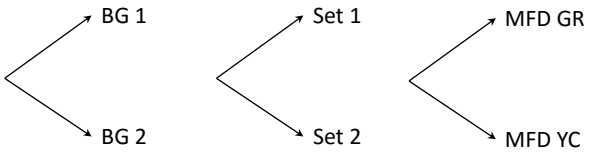

Figure 1.12: *Logic tree for the example calculation in which BG1 and BG2 are the two background hypotheses, set 1 and set 2 are the two rupture scenario set hypotheses, and MFD GR (Gutenberg– Richter) and MFD YC (Youngs and Coppersmith) are the two MFD shape hypotheses.*

It is important to point out that SHERIFS is not modelling earthquake ruptures as an earthquake simulator would do but only converting slip-rate on faults into earthquake rates. Each spending of a slip-rate increment is not the modeling of a true earthquake but is converted into rate of earthquakes that can be allowed on a specified fault.

Non Main Shock slip

Non Main Shock slip is the proportion of slip-rate budget of a fault section that couldn't be transformed into earthquake rates within the SHERIFS procedure. A high NMS proportion signifies that the input slip-rate for this fault section is not compatible with the set of scenarios, the target MFD and the input slip-rate of the neighboring faults given the assumptions and rules of SHERIFS. In this case, NMS can be regarded as an error or misfit of the model. Models that predict a very high NMS, typically over 30 to 40%, need to be reconsidered in detail to understand the origin of this NMS. Similarly, the high NMS possibly calculated for some fault sections needs to be understood. The reason for this NMS can be diverse: the set of scenario involving the fault section or/and the fault slip rate set in input for this section are not realistic or the target MFD is not suitable for the fault system. Large NMS slip can also results from miscalculations in the code, when using a too large slip-rate increment for example. NMS slip differs from what could

be strictly called a misfit or error by the fact that it can be explained by physical phenomena and should therefore not necessarily be reduced to zero. NMS can reflect that part of the slip-rate deduced from geology or geodetic inversions which can be dissipated in non-seismic processes such as after-slip, creep or slow slip events but cannot be taken as a definitive proof of the existence of theses phenomena in a specific fault system as it could reflect some limitations of the SHERIFS approach.

1.2.4 Model visualization

This tool allows extracting and visualizing numerous results that can be easily compared with the seismicity catalog of the region and plotted on a map. The user needs to specify in the python file the name of the run and the data files to use as well as the spatial limits of the maps. The user can select the type of information to plot, very useful when running a large fault model. The user manual describes in more details the use of this tool.

The produced figures include the distribution of MFDs of the whole model and of each individual branch of the logic tree. These MFDs can be compared to the MFD of the local earthquake catalog. The local catalog is extracted using the geometry of the background defined by the user. The uncertainties of both the completeness period of the catalog and the magnitude of the earthquakes can be explored. A user who wishes to study the MFD of a region smaller than the whole system can define a sub-area. The code will extract the MFD of the faults within this sub-area and the MFD of the catalog for this sub-area. This tool also plots the MFD of individual faults and fault sections and compares the results to the rate of historical or paleo-earthquakes when this information is available for single faults.

The user can use this tool to extract statistics concerning the amount of NMS slip in the model, the maximum magnitude for each branch, the repartition of the moment budget between single fault ruptures and complex FtF ruptures.

Maps of each FtF rupture scenario are generated allowing the user to check that the input scenarios were correctly set up.

1.2.5 Logic Tree Weighting

Once the content of the logic tree has been analyzed and compared to the local data (catalog and paleo-seismicity), the user needs to attribute weights to each branch of the logic tree. This tool opens a graphical interface to help building the logic tree file used as an input for the PSHA calculation.

SHERIFS aims to promote the discussion of the uncertainties affecting the earthquake rates on faults. The Logic Tree Weighting tool should not be run before a proper discussion of each branch of the logic tree.

1.2.6 Applications of SHERIFS

Past applications of SHERIFS

The proof of concept of the SHERIFS methodology was realized on the Western Corinth Rift (WCR) fault system (Chartier et al. [2017b]). In this methodological exercise, a Gutenberg-Richter FMD target shape was considered and three sets of FtF scenarios where tested: B14_s that considered only single fault scenarios; B14 that considered FtF ruptures with a distance criteria of 3km between two faults; and B14_hc that considered a higher connectivity within the fault-system and allowed FtF ruptures separated by 5 km or less.

SHERIFS was run with these hypotheses and sanity checks where performed.

The comparison of the fault model results with the seismicity revealed that the B14_s branch of the logic tree implied seismicity rates that were incompatible with the rates computed from the catalog. Moreover, the B4_s branch didn't produce the large earthquakes observed in the historical catalog. Finally, comparison of the modeled rates with paleoseismic data available for the Aigion fault, clearly indicated that the B14_s branch cannot reproduce the rate of large earthquakes deduced from paleoseismic data. On the other hand, for the B14 and the B14_hc branches, where FtF ruptures were allowed, the agreement between the modeled rates and the instrumental/historical seismicity and paleoseismic rate improved, with the 5 km criteria used in B14_hc explaining better the data than the 3 km distance criteria used in B14. The discussion of these results lead to the conclusion that the model B14_s should be discarded from the logic tree and the model B14_hc could have a stronger weight than B14.

The SHERIFS methodology is presently being tested on different types of fault systems in France, Italy, Israel, Spain, Ecuador, India and Turkey. A comparison exercise was also performed between SHERIFS and alternative methodology that are currently being develop within the FAULT2SHA ESC working group (`https://Fault2SHA.net/`) using a fault system in northern Italy (Visini et al. [2019]).

1.2.7 Example of a SHERIFS calculation

We present here a simplified example of application of SHERIFS for the five southern faults of the WCR (Greece) with a discussion of the resulting models. We hope this can be used as an illustration of the type of discussion that can be supported by SHERIFS. Two key features of SHERIFS have been used: a hazard model has been generated and a fundamental discussion on the input hypotheses has taken place. All the calculations files for this example can be found with the SHERIFS code at `https://github.com/tomchartier/SHERIFS`. Using the SHERIFS graphical interface, the model was run (please refer to the user manual for a detailed description of the user interface). For this example, the logic tree presented in figure 1.12 is used with 20 random samples on Mmax, b value (between 0.95 and

Model Name	Fault Name	Dip	Oriented	Mechanism	Seismogenic depth (km)		Slip-rate ($mm.yr^{-1}$)			Shear Modulus (GPa)
					Upper	Lower	Min	Mean	Max	
Example_Model	F1	60	N	N	0	6	4.8	5.0	5.2	30
Example_Model	F2	55	N	N	0	7	3.0	3.2	3.4	30
Example_Model	F3	60	N	N	0	7	3.8	4.0	4.2	30
Example_Model	F5	60	N	N	0	7	3.3	3.5	3.7	30

Table 1.5: Faults Parameters Used for the Example Calculation.

1.05) and the slip rates of the faults (table 1.5). For this run, an increment of slip rate dsr of 0.001 mm/yr was chosen.

In the branch BG1, 100% of the seismicity of all magnitudes is considered to be modeled on the faults. In the hypothesis BG2, 60%, 70%, 80%, 90%, 95%, and 100% of the seismicity is considered to be on the faults for magnitude 4.0, 4.5, 5.0, 5.5, 6.0, and Mw 6.5 and more respectively. In the set of rupture scenarios Set 1, faults F1, F2 and F3 can rupture together in scenarios of two or three faults and fault F5 can only rupture alone. In the branch Set 2, all four faults can rupture together. Finally, two hypotheses of target MFD are explored: GR and modified Youngs and Coppersmith YC (see supplementary material for the complete formulation of the modified YC equation, see Annex 1).

After running the model visualization tool, it is possible to compare the modelled earthquake rates to those of the catalog for each hypothesis of the logic tree.

SHERIFS is a tool that allows to discuss how each branch and each hypothesis of the logic tree performs against the earthquake catalog and paleo earthquake data. This analysis also allows us to understand which hypotheses on this simple fault system are able to reconcile the geologic data (slip-rate, geometry of the faults) with the seismological and paleoseismic data.

Models using a target MFD shape GR match the rate of the catalog better than the YC MFD (figure 1.13). Given the large discrepancy between the modelled rates and the catalog rates for the YC hypothesis, this branch can have a weight of 0 in the logic tree. Between the Set 1 and Set 2, Set 1 leads to slightly higher rates than Set 2 which matches the rate of the catalog better. In this respect, the branch Set 1 can be weighted stronger than the branch Set 2 in the logic tree. The modelled rate can be compared at the fault level with the rate calculated from paleoseismology data (figure 1.14). Here for Fault 3, where some trenches have been conducted, we compare the rate of all earthquake ruptures that involved Fault 3 with the rate calculated from the paleoseismology study.

While both rupture sets hypotheses agree with the rate of paleo earthquakes, within the range of uncertainty,

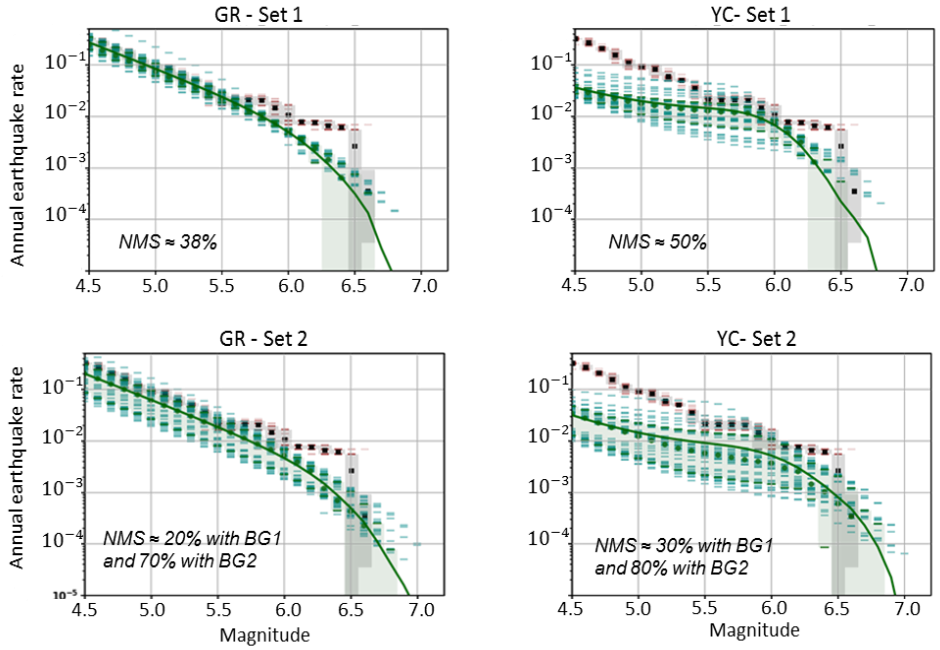

Figure 1.13: *For each hypothesis of MFD shape and rupture scenario set, comparison between the modeled MFD of the whole logic tree and the earthquake rate calculated from the catalog. Dashed green lines are the MFD of each individual model, the solid green line is the mean MFD, and green patches represent the uncertainty (16–84 percentiles). The dotted black line is the rate from the catalog with uncertainties. The ratio of NMS is indicated for each background hypothesis.*

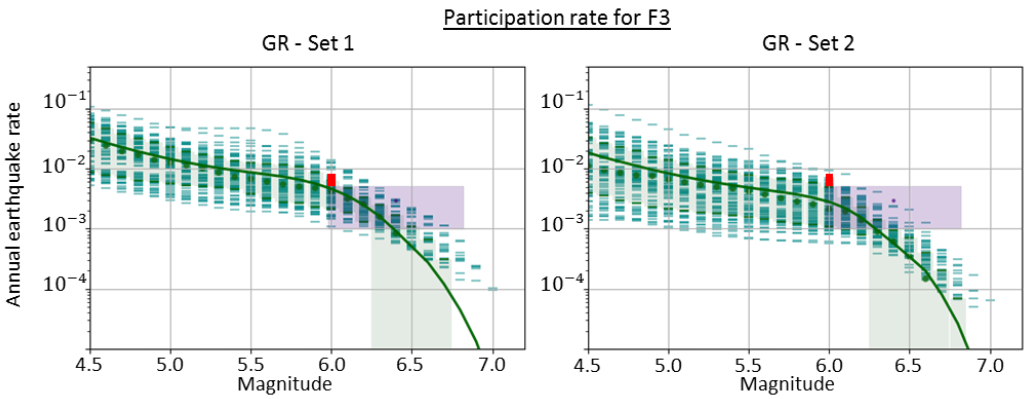

Figure 1.14: *Each green dotted line is the sum of the rates of all different ruptures passing through this fault. The solid green line is the mean rate. The purple dot is the paleoearthquake rate and the purple box is the associated uncertainty. The red dot is the historical earthquake rate and the red box is the associated uncertainty.*

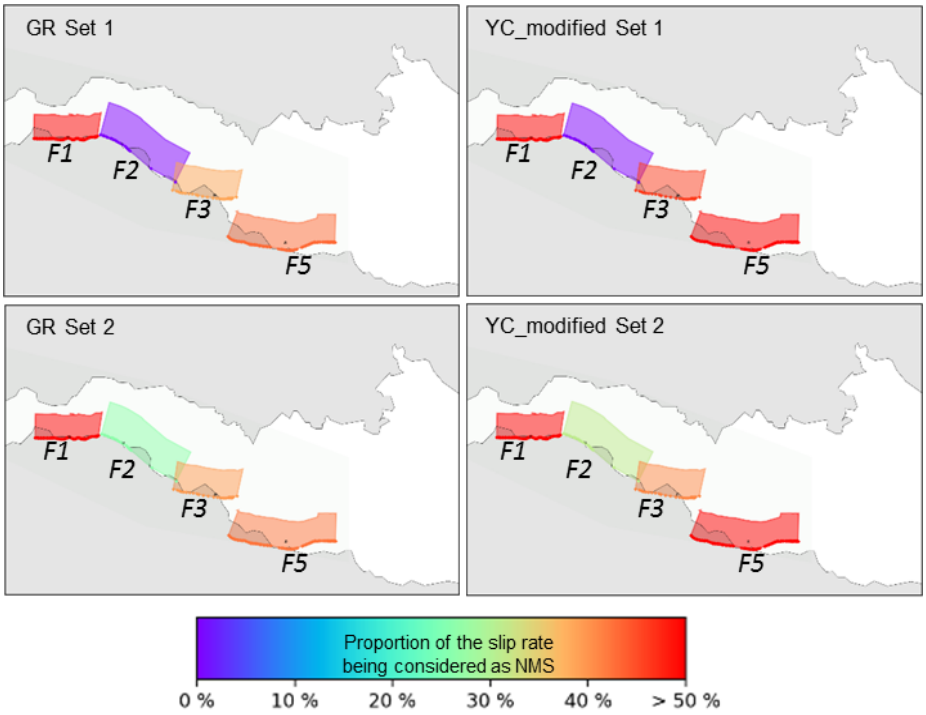

Figure 1.15: *Map of the NMS slip ratio for each branch of the logic tree.*

the Set 2 hypothesis leads to an average annual rate closer to the paleoseismic rate and therefore could be weighted stronger than Set 1. Interestingly even though the regional target MFD is of the shape of a GR, the participation rates of the fault can, and often do, differ from the GR distribution. Similar observations have been made with different system level approaches such as integer programming (Geist and Parsons [2018]) and simulated annealing (Field et al. [2014]).

On the NMS slip map (figure 1.15), we can identify, which faults have trouble spending their slip rate budget given the set of hypothesis. This can help identifying some possible issues with the model input hypotheses.

A NMS ratio of more the 30% for a fault might indicate some modeling issues for a particular model. In this example, all branches of the logic tree lead to large NMS slip on most of the faults of the system. This is due to the fact that this example system is a subset of the faults of the western Corinth Rift fault system and not including the other faults of the system and therefore the possible FtF ruptures with those faults leads to some difficulties to accommodate the input slip-rate as seismic slip-rate in the SHERIFS approach. In other words, identifying the faults that belong to a fault system is paramount.

1.2.8 Limitations of SHERIFS

Although SHERIFS allows considering faults systems as a whole and modeling complex earthquake ruptures in seismic hazard assessment models, it remains a simplification of the natural systems and several limitations have to be noted before any use of this methodology. Although, SHERIFS explores a wide range of epistemic uncertainties, a full exploration of the epistemic uncertainty would require comparison with other fault modeling approaches (Geist and Parsons [2018], Field et al. [2014], Visini et al. [2019]).

SHERIFS is not able to consider complex slip distributions along strike for a given rupture. In the present version, the methodology considers that the spending of slip-rate budget is uniform for all faults participating in a given rupture. SHERIFS being an open-source code, it can be modified at will and we encourage any user willing to develop and test this feature to do so.

When considering only a part of a fault system, SHERIFS will not resolve the rates at the edge of the system correctly because it is likely that ruptures with faults that lie outside of the modelled area were omitted. Users should verify that the sites of interest are located in the center of the modelled area, where the hazard assessment is not sensitive to changes in the definition of the limits of the modelled area, ensuring that all sources potentially contributing to the hazard for the site are correctly modelled.

Testing of SHERIFS on different fault systems has led to important improvements regarding the versatility of the code. Although many cases can now be supported by SHERIFS, it is not impossible that some applications may require further adaptations of the code. Hence, we suggest that SHERIFS be used with care and that the results be always checked by the end-user.

In order to allow the code to be applicable to very diverse fault systems and hypotheses, a few controls on the input files have been implemented. It is the duty of the user to make sure the input files are consistent with the hypotheses wanted for the calculation. SHERIFS provides output files and figures to help judging the validity of the results. The output figures of SHERIFS are not in any case exhaustive and we encourage any user that requires more specific outputs to take part in the development of python routines producing these outputs and to share these developments with the community under the same GNU Lesser General Public License.

1.2.9 Conclusion and perspectives

SHERIFS is an open-source python tool to calculate the rate of earthquakes in a fault system while relaxing segmentation hypotheses. It relies on three basic ingredients: an assumption on the shape of the regional MFD, hypotheses on the possible FtF ruptures and preservation of slip rate variability along faults, if needed. SHERIFS's objective is to provide

data collectors and fault modelers with a friendly and easy-to-use interface to explore coherency of input hypotheses and the impact of uncertainties in data and input hypotheses on the resulting seismic hazard.

Output files from SHERIFS are formatted to be directly usable as input files for the PSHA calculation. Future developments will include tools to analyze the hazard outputs with complex logic trees.

We hope the seismic hazard models generated using SHERIFS can represent more accurately the vision of the data providers.

Data and resources SHERIFS can be accessed free of charge on (`https://github.com/tomchartier/SHERIFS`) and can be used under the license (GNU Lesser General Public License). This open source code can be freely modified by the user as long as the modified code is also available with the same license. Questions about SHERIFS can be asked on the google group `https://groups.google.com/forum/#!forum/sherifs`. This google group aims to be a platform to share the questions related to the use of SHERIFS for the modelling of fault systems. We hope to be able to answer the questions and to improve the points where users have identified short-comings of SHERIFS.

1.3 On sharing SHERIFS

This section presents my experience of sharing the SHERIFS code, the difficulties that came with it but also the numerous scientific payoffs. Hopefully this section can help convincing other PhD students to spend the time and energy required for sharing their codes.

1.3.1 The struggles of sharing an open-source code

Building an operational scientific calculation code from the ground up is not an easy task, especially for someone who is not a trained programmer as it is the case for most PhD students in earth sciences. After having adapted the theoretical background to be numerically compatible and tested different parts of the code, the code is often only usable by the few persons having participated closely to its development.

Choosing a programming language

The choice of the programming language needs to consider four aspects : the popularity of the language in the community, the easiness to install program in the language, the easiness to code in the language for the main developer and the computing efficiency.

For SHERIFS, the language filling the four boxes was python. Python is the most commonly used language at the time, it is easy to install on any platform, development in python is intuitive and it is efficient enough to calculate earthquake rates on faults. For other projects, Matlab could be more suitable if the popularity of Matlab in the community out-weights the difficulties to get access to Matlab (e.g. licence costs). If the computing efficiency is at the essence, Fortran or C could be a better choice.

Making a "user friendly" code

Depending on the target audience, the meaning of the term "user-friendly" can vary widely. In the case of SHERIFS, users can be seismologists, hazard modelers or geologists. Some users might be dazzled by using computer codes and command lines. In order to help theses users to get a head start, I set up a user interface for SHERIFS.

The main factor affecting user-friendliness is the availability of a complete user manual. The user manual should be detailed enough that a new user can use the code to its full extent on its own which means that writing the user manual takes a considerable amount of time.

Beta testing

Assessing the usability of the code by outside parties requires beta-testers. Finding beta-testers is a relatively easy task if the code offers practical solution to a problem. For the case of SHERIFS, many geologists and seismologists in the Fault2SHA Working Group (`http://fault2sha.net/`) were interested in modelling the seismic activity in the fault system they are studying. Before SHERIFS only the FiSH (`http://fish-code.com/`) modeling code, that considers faults individually and not as fault systems was available (Pace et al. [2016]).

Within the Fault2SHA working group, we organized a training workshop with a dozen participants. Before the workshop, the code was sent to the participants so they could try to run the test making sure every component of the code would work during the training. SHERIFS runs with python and relies on common python libraries and should be easy to install on every platform (Linux, Mac or Windows). However, some aspects of the code, in particular the writing and reading of text files was different depending on the platform.

The workshop took time and resources for the organization but allowed direct interactions between developer and users. Through these interactions, changes were made in the code to permit the usability by everyone on every platform. Discussion on the meaning of each term used in the manual about fault modelling clarified the different points of views. After this beta testing, changes were made to both the code and the user manual that made the code more easy to use as intended.

1.4 The pay-offs of sharing an open-source code

1.4.1 A more adaptive and reliable code.

The SHERIFS code was developed for an application to the fault system of the Western Corinth Rift, a normal fault system of moderate size but with high slip-rate values. Even though the SHERIFS approach is in theory applicable for any fault system, in practice, the code needed to be adapted to correctly handle larger systems.

The main beta tester of SHERIFS has been Octavi Gomez-Novell with an application to the Betics fault systems in South Eastern Spain (Gómez-Novell et al. [2018]). In this fault system and for certain hypothesis on the set of ruptures that are possible, different constrains for setting the target MFD where added. We also added tests to guarantee that the target was respected in the resulting modelled MFD. This exercise made the code more versatile and more reliable.

1.4.2 Discussing input hypotheses

The main gain of SHERIFS has been the increased ability for discussing scientific hypotheses concerning fault systems. Having a versatile and user-friendly code allowed different users to apply the methodology on their fault system and to compare the modelled earthquake rates with the available data. Using a common language, we were able to discuss some modelling questions that were shared in several fault systems. I will present here some of these, still on-going, discussions.

Rupture scenarios

Octavi Gomez Novel (Gómez-Novell et al. [2018]) applied SHERIFS on the 400 km long Betics Fault system, South Eastern Spain (figure 1.16), dominated by SW-NE left-lateral strike slip faults, some of which are oblique reverse faults. They discussed the uncertainty concerning the possible multi-fault ruptures in the system and developed a logic tree exploring different hypotheses on the ruptures allowed in the system. The comparison of the modelled rate (earthquake catalog and paleoseismicity) with the data allowed to narrow the number of possible hypotheses, eliminating the branch where the faults rupture only independently and the branch where the rupture can propagate across the whole system (figure 1.17).

During Etienne Marti's master thesis (Marti [2018]), we were able to test rupture hypotheses brought forward by the paleoseismic records along the left lateral Dead Sea Fault. After compiling the available paleosismic data, Lefevre et al 2018 suggested that in the Dead Sea Fault System (DSFS), earthquake magnitudes might be limited to magnitude less than 7. in the North while they might reach magnitude 7.6 in the South. Marti tested different sets of rutpure scenarios and concluded that in order to spend the slip-rate budget of the faults, earthquakes larger than magnitude 7.5 are required in the North as well as in the South (figure 1.18). In this exercise, SHERIFS was used as a tool to test paleoseismic hypotheses.

Magnitude Frequency Distribution

In the DSFS, two MFD hypotheses were explored in order to model the seismicity rate, the GR hypotheses and a modified characteristic MFD (Chartier et al. [2019]). Only the modified characteristic MFD was able to reproduce both the catalog seismicity rates and the rates of paleo-earthquakes along the fault (figure 1.19).

The divergence from the GR MFD is also observed for the NAFS (chapter 2) and the San Andreas Fault System (SAFS) (Youngs and Coppersmith [1985]). Page and Felzer [2015] argue that this observation for the SAFS is due to an observation bias. In other applications of SHERIFS, in normal fault system of the Western Corinth Rift (see

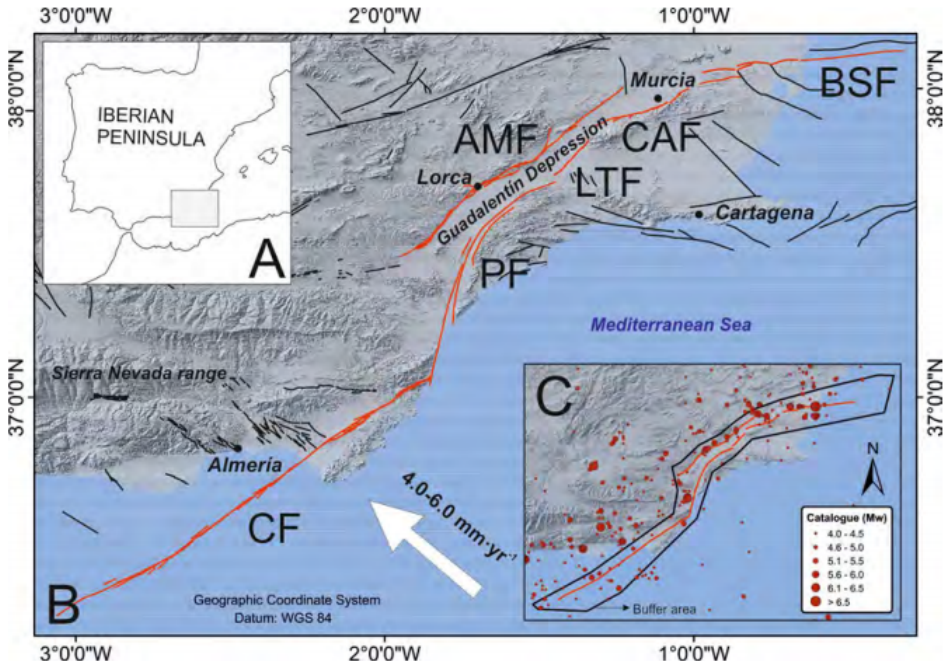

Figure 1.16: *Location of the study area within the Iberian Peninsula. B. Tectonic scheme of the Eastern Betics Shear Zone (EBSZ). Faults considered in the present study are depicted in red. CF: Carboneras fault; PF: Palomares fault; LTF: Los Tollos fault; CAF: Carrascoy fault; BSF: Bajo Segura fault; AMF: Alhama de Murcia fault. C. Seismicity map inferred from the earthquake catalogue used in this study (IGN-UPM working group, 2013), which covers from year 1048 to 2011. The buffer area used to extract the seismicity is indicated. The white arrow represent the converge between the African and the Iberia plates. (modified from Gomez-Novell et al submitted)*

section 1), in strike slip fault system of the Betics (Gomez Novel et al submitted), and in the strike slip fault system of the Armoric Region of France (Jacottin [2019]), the GR MFD hypothesis managed to reproduce the local data. Additional data in these regions might change this observation. After analysis of the paleo-seismic records and the earthquake catalog of several faults in New Zealand, Stirling and Gerstenberger [2018] concluded that the data is too sparse to be conclusive and this uncertainty on the faults MFD should be explored in hazard models. In chapter 3, we show that this uncertainty on the MFD target is the leading source of uncertainty on the estimation of the probability of collapse for a building in Istanbul.

Background seismicity

Hypotheses on the background seismicity modelling are not usually explored in seismic hazard assessment studies, mostly due to a lack of modelling tools. The system level approach implemented in SHERIFS allows testing a range of possibilities for the ratio of earthquakes in the background and on the faults in a very simple way. In chapter 3, we show how the hypotheses on background seismicity can affect the estimation of the seismic risk in Istanbul. In Istanbul, the seismicity rates on the faults reflect the slip-rate of the fault since they are calculated locally based on geological

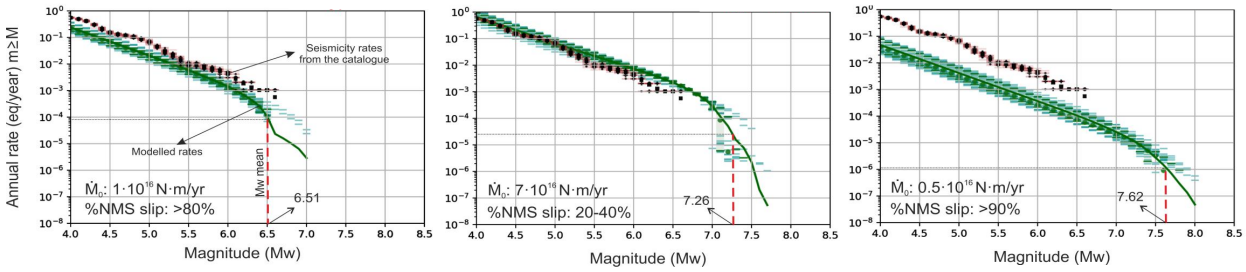

Figure 1.17: *Results of the consistency check obtained for three fault-rupture hypotheses for the Betics Fault System, from left to right : single faults ruptures, intermediate ruptures, the whole system ruptures. Each green MFD corresponds to the Gutenberg-Richter distribution (GR) for the whole fault system, modelled from geological data. The green line is the mean of the cumulative annual rates of the samples explored for every magnitude. The garnet MFD corresponds to the seismicity rates extracted from the catalogue. Each MFD is represented as a coloured density function, where the most intense colours correspond to the cumulative annual rates predicted by a higher number of samples in the modelling, and so, they condition the position of the mean. Note that the mean of the maximum magnitudes from the 20 samples explored for each hypothesis is indicated. The seismic moment rate modelled for the whole fault system is also indicated for every hypothesis. It is fundamental to clearly state that differences in this rate for the different hypotheses are due to the fact that each hypothesis implies a different amount of aseismic slip or Non-main-shock slip (NMS). (modified from Gómez-Novell et al. [2018])*

and geodetic information. The background is added to this seismicity following the ratio set as input for SHERIFS.

In her Masters thesis (Jacottin [2019]), Marie Jacottin modelled the seismicity rates on the 350 km long South Armorican Fault System, Bretagne, Western France. Since no slip rate estimates were available for this slow moving fault, the slip rates were based on a strain map calculated from the GPS displacement field. Therefore hypotheses on how the slip-rate is distributed between the faults and the background were made (figure 1.20).

The modelled seismicity rates can then be compared with the recorded seismicity and those hypotheses on the background seismicity that reproduce the data can be weighted more strongly in a logic tree.

1.5 Conclusion

While the work of sharing a code can be tedious, especially for research oriented PhD students, the availability of the code brings a community together and allows discussion on modelling issues that are shared in different study regions. Enabling the use of a methodology in a given region by the local experts often leads to interesting scientific discussions.

Finally, this move towards open sciences helps sharing the hazard models, therefore increasing the number

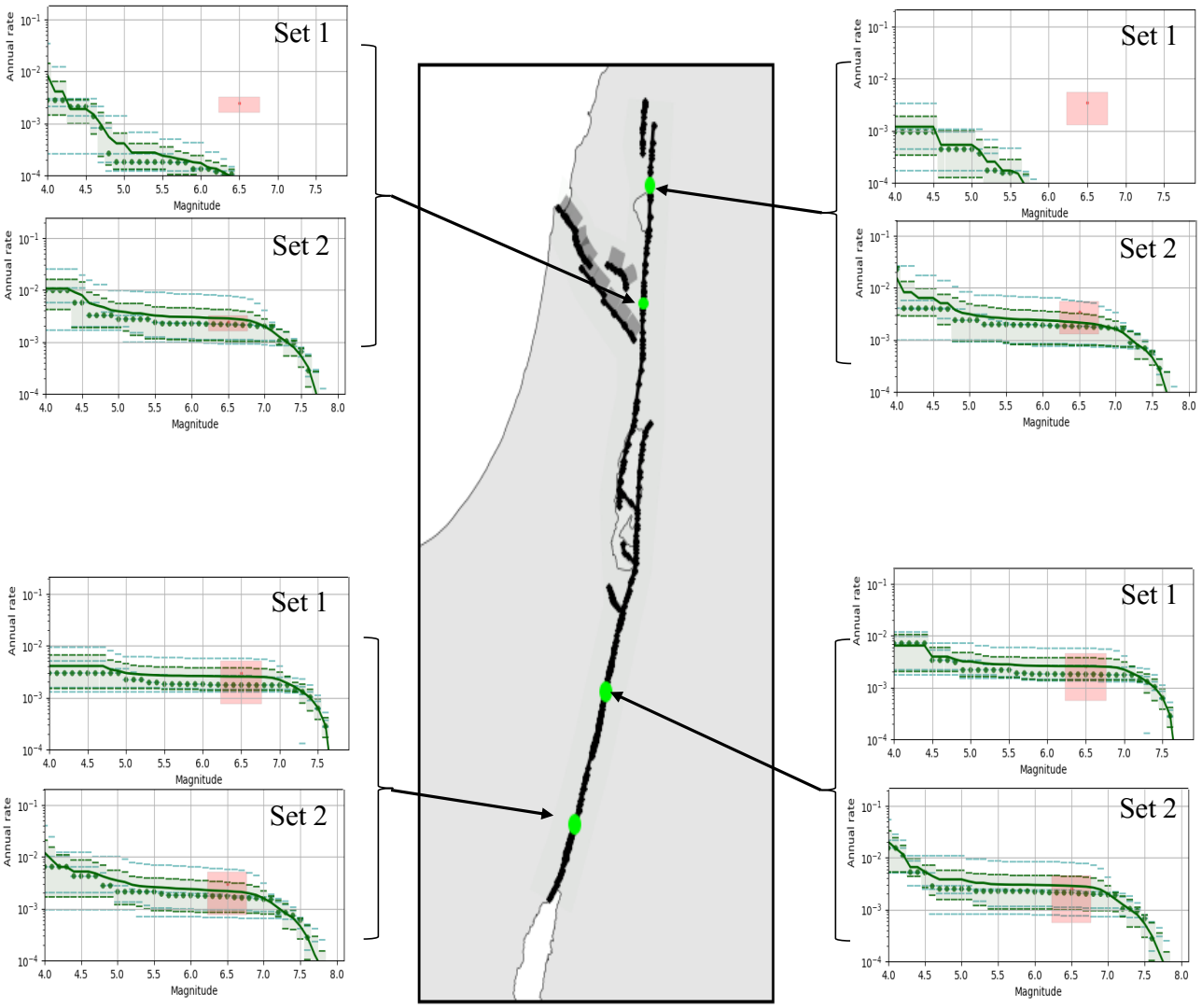

Figure 1.18: *Comparison between the participation rates modelled with SHERIFS (in green, solid line is the mean value, the dashed line are each model of the logic tree) and the Lefevre et al. [2018] paleoearthquake rates (in red) at four sites along the fault. Modified from Marti [2018]. In Set 1 the ruptures in the north are smaller than magnitude 7 and can reach magnitude 7.6 in the south. In Set 2, the ruptures can be larger than magnitude 7.6 both in the north and in the south.*

of people able to review and discuss the models. The open source hazard and risk calculation software OpenQuake (Pagani et al. [2014]), has greatly improved the sharing of hazard models and has allowed easy comparison between hazard estimates which in turn led to a better understanding of the epistemic uncertainties that affect the models. We hope that SHERIFS can be a step in the same direction and improve the discussion around modelling the earthquake rates in fault systems.

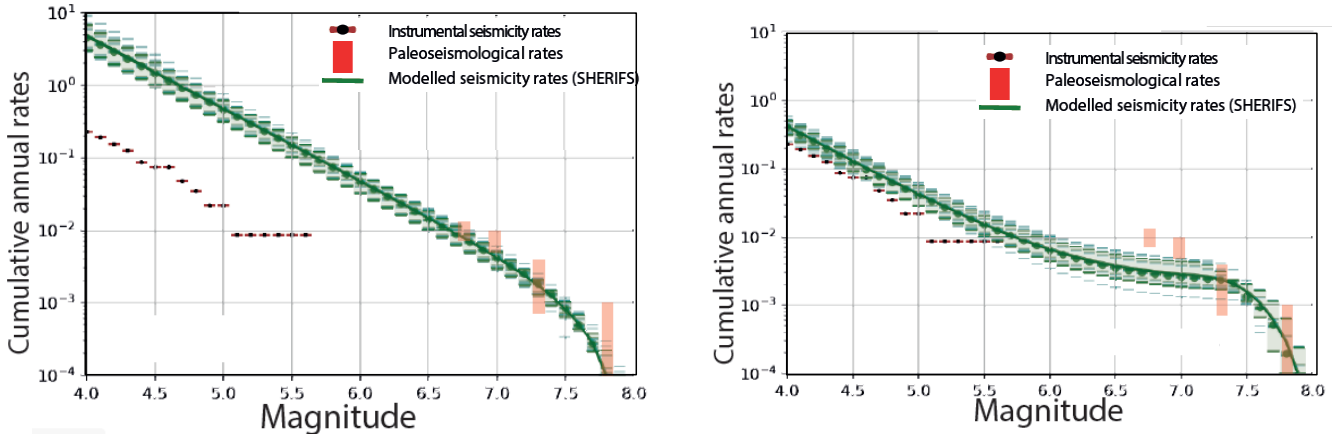

Figure 1.19: *Comparison between the MFD modelled with SHERIFS (in green) and the MFD calculated from the earthquake catalog (in black) and the paleoearthquake rates calculated from Lefevre et al. [2018] (in red). On the left, the MFD when using the GR MFD hypothesis, on the right, the MFD using the modified characteristic MFD hypothesis. Modified from Etienne Marti's Master thesis (Marti [2018]).*

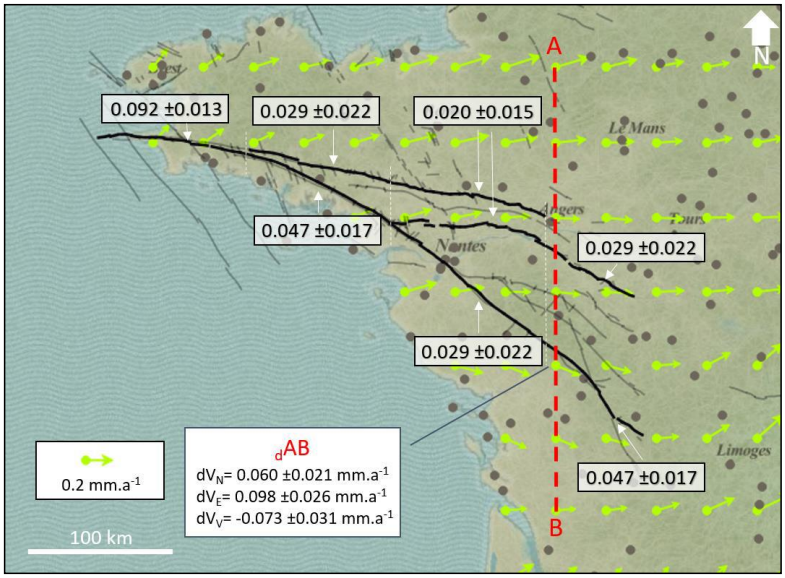

Figure 1.20: *Deformation map calculated for Western France from the GPS data. The differential velocity between the point A and B is indicated for the three components. The slip-rate attributed to the fault sections by sharing the A-B differential between parallel sections is indicated in the box next to the fault (in black). Modified from Jacottin [2019]*

Chapter 2

Modelling the earthquake rates in the Marmara Region and the associated uncertainties

Résumé du chapitre

Dans ce chapitre, nous utilisons la méthodologie SHERIFS pour calculer les taux de sismicité dans la région de la Mer de Marmara dans le but d'estimer le risque sismique pesant sur la ville d'Istanbul en Turquie. Istanbul se trouve à une vingtaine de kilomètres de la faille Nord Anatolienne qui a une vitesse de glissement proche de 20 mm/an. La faille Nord Anatolienne est localisée sous la Mer de Marmara sur une longueur d'environ 200 km; son étude est donc un challenge. Au travers d'un arbre logique, nous explorons un ensemble d'incertitudes ayant le potentiel d'affecter le taux de sismicité modélisé.

Quatre hypothèses concernant le blocage de la faille dans la partie occidentale de la Mer de Marmara sont explorées: une hypothèse où la faille n'est pas bloquée et n'accumule donc pas d'énergie élastique sur cette section, une hypothèse où la faille n'accumule que 50% du glissement en énergie élastique, une hypothèse où la faille est bloquée sur les 5 premiers km et glisse dans sa partie plus profonde, et enfin une hypothèse où la faille est totalement bloquée et accumule 100% du glissement en énergie élastique.

La MFD calculée à partir du catalogue de sismicité ne suit pas une distribution du type Gutenberg-Richter (GR): la fréquence d'occurrence des séismes de magnitude supérieure à 7 est plus forte que ce que la décroissance

logarithmique d'une courbe du type GR prédirait. La réalité d'une telle distribution pour un système de faille est débattue dans la communauté scientifique, certains argumentant que cette observation est biaisée par la courte période de complétude du catalogue comparée à la longueur du cycle sismique. Nous établissons donc deux branches dans l'arbre logique: une des branches utilise une MFD suivant une distribution du type GR et l'autre une MFD de la forme de la MFD du catalogue.

Deux hypothèses sont considérées pour la rupture la plus large possible dans le système de failles. La première considère un séisme de l'ordre de magnitude 7.7, rompant les deux ruptures de 1999 simultanément. La seconde envisage que des ruptures plus grandes, générant des séismes de magnitude 8 sont possibles dans le système, cette hypothèse ayant été considérée par d'autres études.

Nous utilisons les capacités de SHERIFS pour modéliser trois hypothèses différentes pour la sismicité de fond, se produisant sur des failles non connues ou non considérées dans le calcul. Ces hypothèses vont d'une considération que la grande majorité des séismes se produisent sur les failles connues, y compris les séismes de magnitudes modérées (de 4.5 à 5.5), à une considération que la majorité de magnitudes modérées se produisent sur des failles non connues et qu'il est possible que des séismes de magnitude 7 se produisent sur des failles considérées inactives ou non connues.

Enfin, pour chaque branche de l'arbre logique, 10 tirages aléatoires sont effectués sur les vitesses de glissement et les paramètres de la loi d'échelle reliant la dimension des ruptures à la magnitude.

Afin de pondérer chaque hypothèse de l'arbre logique, les taux de sismicité sont comparés aux taux calculés à partir du catalogue de sismicité et des études de paleosismicité. Ces données, étant un échantillon court du cycle sismique de la faille, ne permettent pas de choisir clairement une hypothèse entre les deux possibilités de MFD et les deux possibilités de rupture maximale. Pour compléter cette discussion, nous modélisons un catalogue de sismicité synthétique avec le code RSQSim qui simule un catalogue de séismes à partir des lois de friction "rate ans state". Ce catalogue de 10 000 ans ne présente pas de séismes plus grand que magnitude 7.7 et sa MFD diverge de la GR MFD.

Basé sur ces observations, il est alors possible d'attribuer un poids plus important pour ces deux hypothèses dans l'arbre logique.Le résultat de cette étude est une estimation des taux de sismicité dans le réseau de faille et de leurs incertitudes représentant clairement l'état des connaissances actuel.

2.1 Introduction

The North Anatolian Fault System (NAFS) runs through the North of Turkey for a distance of more than 1500 km (Figure 2.1) accommodating the westward extrusion of the Anatolian plate with a right lateral motion of around 25 mm/yr (Reilinger et al. [2006]). Along the western portion of the NAFS, this right lateral motion is partitioned between to fault branches with the northern branch accommodating most of the motion (Reilinger et al. [2006]). This northern branch of the NAFS crosses the Sea of Marmara 20 km south of the city of Istanbul (Figure 2.2). The last earthquakes having occurred in the Marmara Sea, severely damaging the city of Istanbul, date back to 1766 and 1894 (Ambraseys [2002]). Growing from a population of less than one million in 1950, Istanbul is now a city of more than fifteen million. The 1999 Izmit earthquake that ruptured the fault eastward of the Marmara Sea has been a reminder of the hazard faced by Istanbul. The NAFS is expected to be the major source of seismic hazard looming on the megalopolis. Motivated by better constraining the seismic hazard, the two decades following the 1999 earthquake have been rich in geological, geophysical and geodetic studies of the Marmara Region (Ergintav et al. [2014]). These studies have brought new light on the structure (Armijo et al. [1999], Le Pichon et al. [2001], Le Pichon et al. [2003] ...) and dynamics (Hergert and Heidbach [2010], Ergintav et al. [2014], Bohnhoff et al. [2017] ...) of the NAFS and raised important questions such as the locking condition of the fault and the possibility of a large rupture in the Marmara Sea.

Building on the findings of these studies, several Earthquake Rate Forecast (ERF) models and hazard maps have been developed (Erdik et al. [2004], Gülerce et al. [2017], Murru et al. [2016],Sesetyan et al. [2018], Demircioğlu et al. [2018]). While these studies improve our understanding on the hazard for Istanbul, none of them fully explore the epistemic uncertainties in the parametrization of the ERF parameters that have the potential to affect the seismic hazard in Istanbul.

The NAFS being offshore on its portion closer to Istanbul, satellite-based geodesy (GPS and INSAR) are not able to resolve accurately the locking condition of the fault which remains debated (Klein et al. [2017]). The possibility of creep in the Marmara Sea might have an impact on the seismic hazard for Istanbul and should be explored in the seismic hazard assessment.

Past studies (Erdik et al. [2004], Demircioğlu et al. [2018]) have used the classic approach to model faults and background seismicity in seismic hazard. The rate of earthquakes in the background is based on the local seismic catalogue up to a threshold magnitude (M5.5) and the rate of larger magnitude earthquakes are based on the slip-rate of the faults. This approach can lead to several limitations in Probabilistic Seismic Hazard Assessment (PSHA) and eventually in the seismic risk assessment. First, this approach lacks the consideration that earthquakes with a magnitude larger than the threshold magnitude can occur in the background as it has been observed in July 2019 in California where a magnitude 7.1 earthquake occurred on a fault that was not modelled as active in the UCERF3

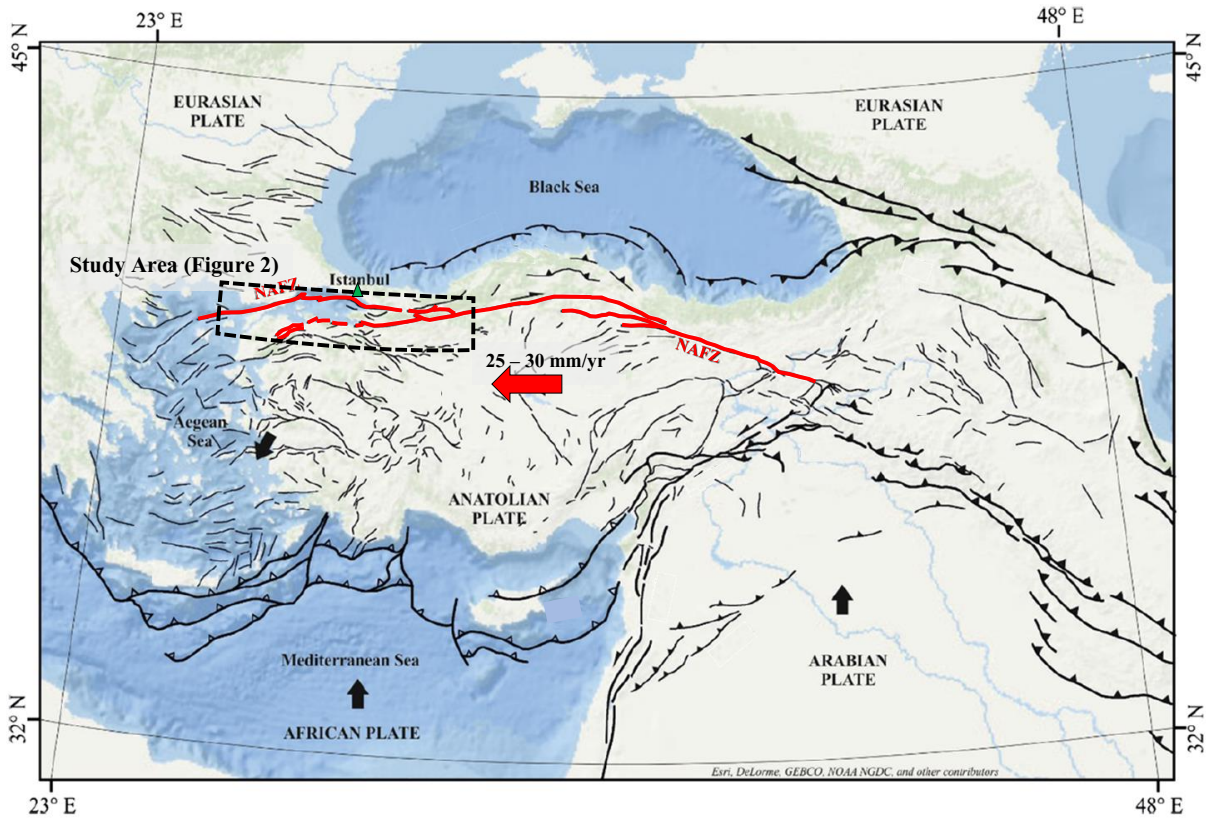

Figure 2.1: *Regional tectonic setting. Modified from Duman et al. [2018]. The active faults are in black except for the North Anatolian Fault Zone (NAFZ) which is in red. The black and red arrows show the plate motion relative to a fixed Eurasia.*

model (Brandenberg et al. [2019],Field et al. [2014]). However, the possibility of such earthquake occurring in the background was considered (Field et al. [2014]). Second, this approach when used with a spatially uniform distribution of the seismicity, cannot consider that intermediate magnitude earthquakes (lower than the threshold magnitude) can be more likely to occur on the fault than on any given point of the background.

In this study, we will use the recently developed SHERIFS code (Chartier et al. [2019]) that allows exploring these uncertainties using a logic tree approach. Furthermore, SHERIFS considers a system level approach, in which different rupture scenarios are explored in an aleatory manner, relaxing fault segmentation and defining complex multi-faults ruptures.

Through the SHERIFS approach we can also address two issues (1) the uncertainty on the size of the largest rupture that may occur along the NAFS and (2) the shape of the Magnitude Frequency Distribution (MFD).

The geometry of the network could potentially host larger ruptures than the one observed in the historical time. For example, the change of azimuth in the geometry of the fault in front on the Princes Island (Figure 2.2) that was not crossed by the 1999 Izmit earthquake could be crossed by a future earthquake as it has been shown by Oglesby

et al. [2008] and Aochi and Ulrich [2015] using dynamic rupture modelling. Furthermore, ruptures that run through more complex fault geometries have been observed in fault systems (e.g. the 2016 Kaikoura earthquake, Klinger et al. [2018]).

It has been long observed that the number of earthquakes in a region decreases with the magnitude of the earthquake. Gutenberg and Richter [1954] established the logarithmic decrease of the number of earthquake with the magnitude (noted as GR law hereafter). However, when analyzing the MFD of the earthquake catalog in the Marmara region we can observe a larger number of earthquakes of magnitude 7.0 and above than predicted by the GR law. Since there are only 6 earthquakes larger than magnitude 7.0 in the catalog, the long term representativity of this divergence from the GR law can be debated. Several studies have discussed this aspect of earthquake statistics along the San Andreas strike slip fault systems; either arguing for the divergence (Schwartz and Coppersmith [1984]) or the GR law (Page and Felzer [2015]). Recently, Stirling and Gerstenberger [2018] have analyzed several fault zones in New Zealand and have argued for the systematic exploration of the uncertainty on the shape of the MFD when modelling faults in seismic hazard assessment. We explore this uncertainty in this study.

The earthquake rates modelled with SHERIFS using each combination of uncertainties will be compared to the earthquake rates calculated from the earthquake catalog and the paleoseismic records in order to give a score to each model. However, for some hypotheses, the comparison with the data is not sufficient for rating one hypothesis against another. We tackle this issue by modelling a synthetic earthquake catalog using the earthquake simulator RSQSim (Richards-Dinger and Dieterich [2012]) along the fault system. By analyzing the statistics of the synthetic catalog, we are able to discuss the physical validity of the different hypotheses explored and preferentially weight the branches of the logic tree that are more compatible with physically-based models.

Because the aim of this study is to better understand the seismic risk in Istanbul, in chapter 3, the uncertainties concerning the modelling of the earthquake rates in the fault system are propagated in the calculation of the probability of collapse of a building in 50 years. The impact of each uncertainty on the uncertainty of the estimation of the probability of collapse is quantified and discussed.

2.2 The Western North Anatolian Fault System

Fault traces and some of the slip rate estimates along the North Anatolian fault system have been recently updated in the Emre et al. [2018] and references within. In this study, we rely on their map (available online at `http://www.mta.gov.tr/v3.0/hizmetler/yenilenmis-diri-fay-haritalari`, last accessed August 2019) to digitize the fault traces (Figure 2.2).

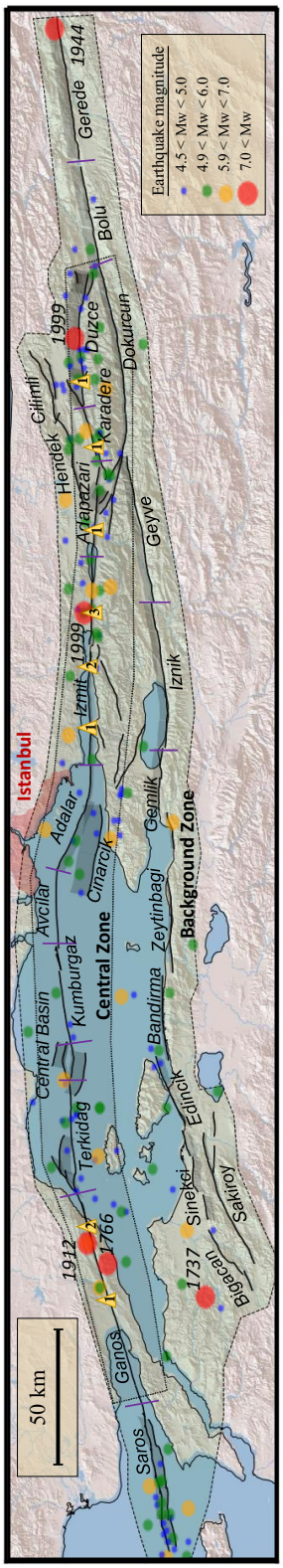

Figure 2.2: *Fault system and the earthquake catalog used in this study (Emre et al. [2018]). The faults sections are delimited by the purple lines and the name of the fault sections are indicated in black. Only the earthquake of magnitude greater than magnitude 4.5 and during the complete period (see Table 2.2) are plotted. The year of occurrence of the largest earthquakes (M>7) are indicated. The yellow triangles are the location of paleoseimic studies.*

The faults parameters of the main faults used in this study are presented in Table 2.1. Slip rates are taken from Hergert and Heidbach [2010] who inverted the GPS velocity field in a geomechanical model in order to calculate the slip-rate of the fault network. These slip-rates are in general agreements with the slip-rate calculated by Flerit et al. [2004] and Ergintav et al. [2014]. Since they agree with the long-term geological slip-rate of the faults, these values of slip-rate have been preferred to those estimated by Reilinger et al. [2006] calculated at a larger scale and that tend to overestimate the slip-rate compared to the geological estimate (Hergert and Heidbach [2010]). Uncertainties on the slip-rate value (Table 2.1) are taken into account by a random exploration in a uniform distribution between the minimal slip-rate value and the maximal slip-rate value set from the literature (Hergert and Heidbach [2010],Klein et al. [2017]) and weighted in order to give a central weight to the mean value. It can be noted that the range of the slip-rate explored in this study does not include values that vastly differ from the geological estimate (23 mm/yr according to Le Pichon et al. [2003]), since they integrate the deformation on a wider area than the fault fault itself.

The scientific community has been debating over the possibility of the NAF been creeping in the Western Marmara region, along the Terkidag, Central basin, Kumburgaz and Acvilar sections of the fault (table 2.1, figure 2.2). Creep has been observed on other strike-slip faults in the world (San Andreas Fault – Nason [1973], Longitudinal Fault Taiwan – Hsu and Bürgmann [2006]) and is expected to release an important part of the tectonic load of the fault without producing major earthquakes hence possibly reducing the seismic hazard in the region. The large scientific efforts of the past two decades have led to conclude that the Eastern part of the fault in the Marmara Sea is locked (Diao et al. [2016]). However, the distance from the fault to the shore in the western part of the sea remains a challenge for traditional land-based or satellite-based geodetic instruments. While novel sea-floor geodesy experiments seem to suggest that the fault is locked at the surface at least at some places (Lange et al. [2019]), these conclusion are still under discussion (Yamamoto et al. [2019]). Relocated microseismicity and the identification of repeater earthquakes suggest that the lower part of the fault might be creeping (Schmittbuhl et al. [2016b],Schmittbuhl et al. [2016a]). Based on the latest studies of the fault in this area, we propose to explore four models of locking condition for the NAF that represent the current state of knowledge. The "creep" hypothesis considers the fault as fully creeping, thus reducing its contribution to the seismic slip-rate budget to 0. The "partial creep" hypothesis considers the fault releasing half of its slip-rate as creep and half as seismic moment, reducing its slip-rate by half. The "deep creep" hypothesis considers that the fault is fully locked for the first 5 kilometers and creeping below, and the "fully locked" hypothesis considers that 100% of the fault slip-rate can be released as earthquakes.

For the faults within the Armutlu peninsula located between the northern strand and the southern strand of the NAF, south of the Izmit section and north of the Iznik section (figure 2.2), we are not able to find slip-rate estimates. However, the velocity field shows very little deformation (Ergintav et al. [2014]) within the Armutlu peninsula. Based on this observation, we assume the slip-rate of these faults to be less than 1 mm/yr. It is worth noting that these faults are relatively far from Istanbul in comparison to the NAF and their participation in the seismic hazard affecting

Fault Name	Rake	Creep hypothesis	Seismogenic depth (km)	Min slip-rate (mm/yr)	Mean Slip-rate (mm/yr)	Max slip-rate (mm/yr)
Ganos	180	Fully Locked	15	17	18	19
Terkidag	180	Creeping	15	0.	0.	0.
		Partial Creep	15	7.5	8.	8.5
		Deep Creep	5	17	18	19
		Fully Locked	15	17	18	19
Central Basin	180	Creeping	15	0.	0.	0.
		Partial Creep	15	7.5	8.	8.5
		Deep Creep	5	17	18	19
		Fully Locked	15	17	18	19
Kumburgaz	180	Creeping	15	0.	0.	0.
		Partial Creep	15	7.5	8.	8.5
		Deep Creep	15	17	18	19
		Fully Locked	15	17	18	19
Acvilar	180	Creeping	15	7.5	8.	8.5
		Partial Creep	15	7.5	8.	8.5
		Deep Creep	15	17	18	19
		Fully Locked	15	17	18	19
Adalar	-168	Fully Locked	10	12.2	14.3	16.5
Cinarcik	-147	Fully Locked	15	2.2	3.6	5
Izmit	180	Fully Locked	15	18	20	22
Adaparazi	180	Fully Locked	15	16	18	20
Karadere	180	Fully Locked	15	11.8	15	18
Duzce	180	Fully Locked	15	11.8	15	18.2

Table 2.1: Properties of the faults of the central zone. The full properties table is available in the annexes.

the city is expected to be negligible.

In this study, we combined two catalogs : the earthquake catalog from Kadirioğlu et al. [2018] homogenized in Mw for the period 1900-2012 and the catalog SHEEC (Stucchi et al. [2013]) for the period 1000-1899. The completeness period used in this study are based on these two studies and presented in Table 2.2.

Completeness from	4.0 - 4.7	4.8 - 5.2	5.3 - 5.7	5.8 - 6.2	6.3 - 6.7	6.8 - 7.2	7.3 - 7.7	7.8+
SHARE (Woessner et al. [2015])	1987	1952	1900	1850	1750	1700	1700	1700
Kadirioğlu et al. [2018]	1992	1977	1937	1850	1750	1700	1700	1700

Table 2.2: Completeness time used in this study.

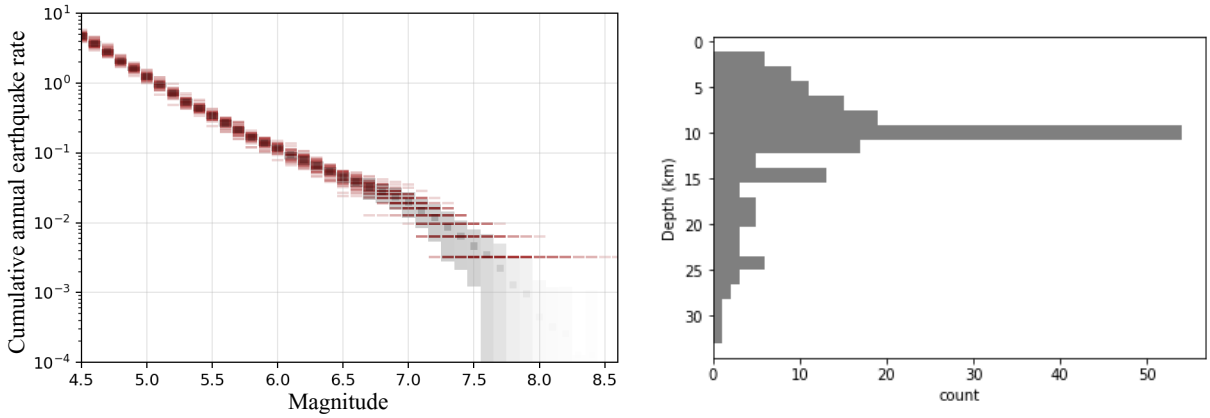

Figure 2.3: *Left : Magnitude Frequency Distribution for the study area calculated from the catalog and the completeness periods presented in Table 2.2. The red dashes are individual Monte Carlo samples on the earthquake magnitude uncertainties and the completeness period. The mean rate value is indicated by black squares and the 16th and 84th percentiles are indicated by the grey area. Right : Depth distribution of earthquakes of magnitude 4.0 and above since 1992 in the Kadirioğlu et al. [2018] catalog and corresponding to the background zone shown in Figure 2.2.*

In this study, based on the analysis of the depth distribution of earthquakes in the instrumental catalog (Kadirioğlu et al. [2018]) and of the rupture inversion of the Izmit earthquake, we have limited the seismogenic thickness of the fault to a depth of 15km. 80% of the earthquakes are in the first 15km of the crust in the region of interest (Figure 2.3). The Izmit rupture inversions show that most of the slip during the earthquake was contained in the first 15km of the crust (Reilinger et al. [2000]).

Dikbaş et al. [2018] present a review of the paleoseismic studies that have been carried on the Izmit, Adapazari, Karadere and Duzce segments of the fault. They propose an interpretation of the rupture history of these segments. Based on the number of ruptures observed in one location and the length of the observation period, we estimate the annual rate of earthquakes with a magnitude larger or equal to 7.2 +- 0.1 for different sections of these segments. The location of these segments are presented in figure 2.2 and the estimated rates are presented in table 2.3. At the West of the Marmara Sea, for the Ganos segment, the paleoearthquake rates have been calculated from Rockwell et al. [2009] and Meghraoui et al. [2012]. Both studies lead to similar rates for earthquakes larger than 7.2 (table 2.3). The paleoearthquake rates are calculated by dividing the number of events by the observation time, the uncertainty reflects the statistical uncertainty due to the small number of observation assuming a poissonian process.

Site	Reference	Number of events	Observation time	Annual rate	Uncertainty
Ganos 1	Rockwell et al. [2009]	4	1000	0.00357	0.001
Ganos 2	Meghraoui et al. [2012]	5	1600	0.003	0.001
Izmit 1	Klinger et al. [2003]	4	1600	0.0025	0.0013
Izmit 2	Dikbaş et al. [2018]	5	1600	0.0032	0.0014
Izmit 3	Dikbaş et al. [2018]	5	1600	0.0032	0.0014
Adaparazi 1	Dikbaş et al. [2018]	3	1600	0.0018	0.001
Karadere 1	Dikbaş and Akyüz [2011]	3	1600	0.0018	0.001
Duzce 1	Pantosti et al. [2008]	4	1600	0.0025	0.0013

Table 2.3: Annual rate of M 7.3+ earthquakes based on the paleoseismicity.

2.3 Earthquake rate modelling

In this section, we describe two approaches for calculating earthquake rates in fault systems. The first approach is SHERIFS (Chartier et al. [2019]), a statistical approach that converts the slip-rate of the faults into earthquake rates built to respect a given shape of MFD at the system level. The second approach is RSQSim that generates a long catalog of synthetic ruptures using the loading rates and the rate and state equation (Richards-Dinger and Dieterich [2012]). SHERIFS allows to explore a wide range of epistemic uncertainties on the input hypotheses and RSQSim allows to discuss which of these hypotheses are physically plausible.

2.3.1 The statistical approach: SHERIFS

Core principle and main input hypothesis

SHERIFS uses an iterative budget spending approach of the slip-rate of the fault to calculate the annual rate of occurrence of each rupture of a predefined set of ruptures. In an iterative manner, SHERIFS randomly selects user-defined rupture scenarios for which involved faults have slip-rates budget to spend. The random selection is done in order to ensure that the resulting system level MFD has the shape imposed as input (b value in the case of a Gutenberg-Richter MFD, for example). It is important to recall that SHERIFS doesn't simulate earthquakes but only converts slip-rates into earthquake rates. For this reason, SHERIFS is computationally light which allows aleatory (combination of ruptures) and epistemic (logic tree) uncertainties concerning the fault system to be easily explored.

SHERIFS takes as input the geometry and slip-rate of the faults, the set of multi-fault ruptures that can be expected in the fault network, and the shape of the MFD defined at the fault system level. Before the calculation, the

actual value of the MFD and the shape of the MFD of each individual fault are not known. They will be deduced from the fault slip-rate budget and the other hypotheses. Depending on the combination of input hypotheses and faults parameters, SHERIFS can consider part of the slip-rate budget of some fault as Non Main Shock (NMS) slip in order to respect the target MFD shape. A NMS of more than 30% is most likely to be an indication that the combination of input hypotheses used doesn't agree with the faults parameters in the SHERIFS framework and that they should be reconsidered.

Background seismicity

One uncertainty that SHERIFS allows to explore is the proportion of seismicity that can occur in the background, on faults that are unknown or not considered as active in the model. In most PSHA, this is taken into account by a background zone with a GR MFD truncated at a given Mt.

In SHERIFS, it is possible to define a-priori the proportion of earthquakes that can be expected on the faults and the proportion in the background for each range of magnitude. In order to assess these proportions, we analyze the distance of each earthquake to the closest fault in the model. Considering the poor knowledge on the epicentral location of the historical earthquakes, we only consider the instrumental catalog after 1970. Since we are interested in the hazard in Istanbul, we only consider the spatial distribution of earthquakes around the sea of Marmara (the central zone in Figure 2.2).

In order to represent the epistemic uncertainty associated with the proportion of seismicity considered on the faults versus that on the background, we set up three branches of the logic tree. The proportion of earthquakes considered to occur on the faults for each branch is presented in Table 2.4. Studies of the NAFS off-fault deformation might bring some light on these value in the future (Şengör and Zabcı [2019]).

Magnitude	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	8.0+
Background 1	0.8	0.8	0.8	0.8	0.8	0.9	1.0	1.0	1.0
Background 2	0.7	0.7	0.7	0.7	0.7	0.8	0.9	1.0	1.0
Background 3	0.4	0.4	0.4	0.45	0.5	0.6	0.8	1.0	1.0

Table 2.4: Ratio of earthquakes assumed to be on the faults for each background hypothesis.

MFD

The annual rate of earthquakes obtained by analysis of the Kadirioğlu et al. [2018] catalog in the Central Zone (Figure 2.2) indicates a MFD diverging from a GR MFD for magnitudes larger than 6.5 (figure 2.3). The rates are obtained

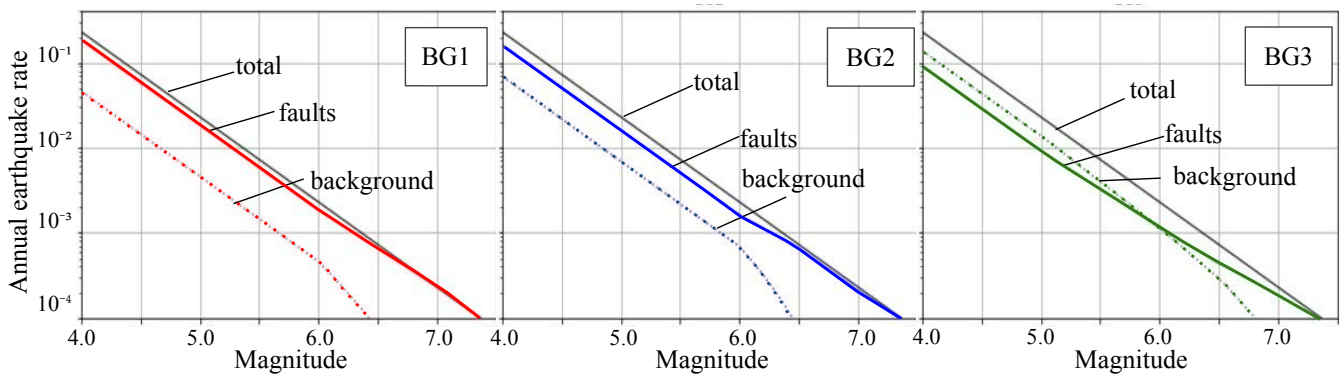

Figure 2.4: *Distribution of earthquakes between the background seismicity and the faults for each background hypothesis.*

while exploring the uncertainties on the magnitude of earthquakes, on the completeness period length using a Monte Carlo approach and the uncertainties linked to the low number of earthquakes for the larger magnitudes. Despite this large number of uncertainties explored, the resulting MFD significantly differs from the GR MFD shape.

Like any earthquake catalog, the Turkish catalog is short in comparison with the seismic cycle and it can be argued that the observed shape is a result of the incomplete sample of the phenomenon and that the long term MFD should follow a GR distribution. This debate on the catalog MFD has been present concerning the Californian earthquake catalog as well (Page and Felzer [2015],Parsons et al. [2018]).

In order to reflect the lack of consensus in the community, we explore two hypotheses for the target MFD : one where the target MFD follows a GR truncated between a minimum magnitude and a maximum magnitude, one where the target MFD follows a shape tuned to that of the rates deduced from the catalog. The tuned shape MFD is described by the following equation and is composed of two parts, both defined by a double truncated GR with a b value deduced from by the lower magnitude part of the catalog ($4.5 < \mathrm{M} < 6.0$) that follows an exponential decrease.

$$P_i(M) = \begin{cases} \frac{e^{-\beta\times(M-0.05-M_{min})}}{1-e^{-\beta\times(6.8-M_{min})}} - \frac{e^{-\beta\times(M+0.05-M_{min})}}{1-e^{-\beta\times(6.8-M_{min})}}, & \text{if } M < 6.6 \\ P_i(M=6.6)/3, & \text{if } M = 6.7 \text{ or } 6.6 \\ P_i(M=6.3), & \text{if } M = 6.8 \\ P_i(M=6.8) \times (10^{(b\times(M-6.8))}), & \text{if } M > 6.8 \end{cases}$$

where P_i is the density function depending on M; b describes the linear decrease with M; $\beta = b \times \log 10$; and M_{min} is the minimal magnitude considered in the calculation.

Set of rupture scenario

The historical catalog doesn't record earthquakes larger than 7.5 since 1700. Based on the statistics of the earthquake catalog, Bohnhoff et al. [2016] consider this value of 7.5 to be the largest possible magnitude in the Marmara region

while earthquakes can reach up to magnitude 8.0 in the Eastern part of the NAFS. A recent review of paleoseismological studies and historical earthquake studies (Dikbaş et al. [2018]) suggests that the largest earthquake having occurred in the region could have been a rupture including both the 1999 Izmit earthquake rupture area and the 1999 Duzce earthquake rupture area resulting in a magnitude 7.7 earthquake. Based on this assumption, we define a Set 1 of more than 300 Fault to Fault (FtF) ruptures allowing a large diversity of ruptures to be possible in the system. Figure 2.5 a) exposes the largest ruptures possible in Set 1.

The resolution of the earthquake records diminishing as we consider older events, it is possible that larger earthquakes occurred in the NAFS but may not have been observed neither in the historical times nor in the paleoseismic records. Furthermore, since most bends in the system could be crossed by a rupture without jumping a large distance or large changes in azimuth, we need to imagine that larger earthquakes might be possible. This hypothesis has been considered in previous studies (Murru et al. [2016], Mignan et al. [2015]). We thus build a Set 2 of FtF ruptures to explore this possibility in which the largest possible ruptures correspond to a magnitude 8.0 earthquake. Figure 2.5 b) exposes the largest ruptures possible in Set 2.

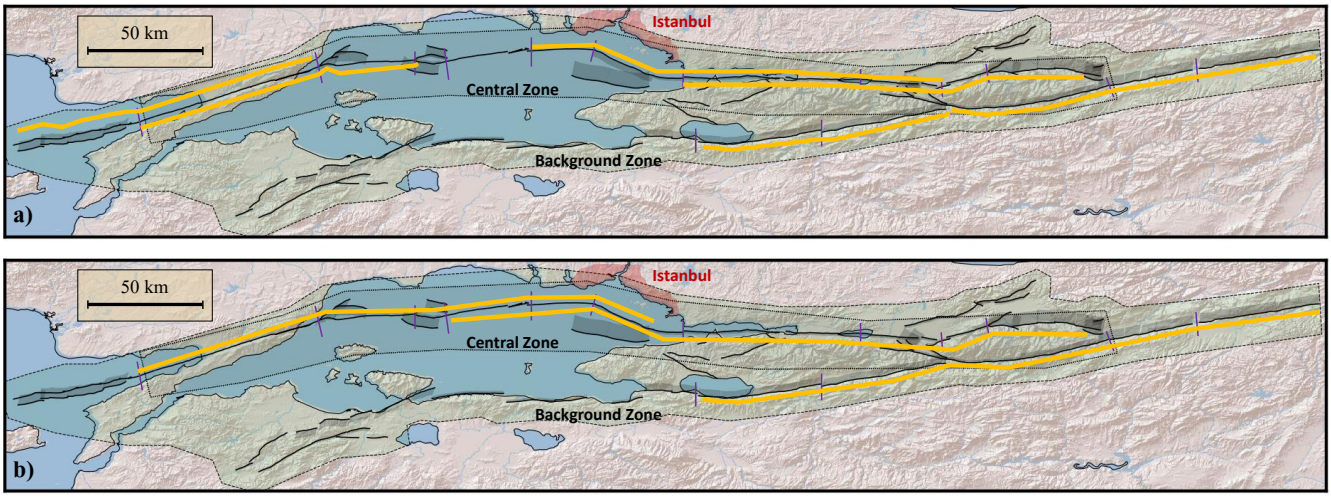

Figure 2.5: *Largest ruptures included in the models using the Set 1 (a) and the Set 2 (b).*

Locking condition of the NAFS in the Western Marmara region

As exposed during the presentation of the NAFS earlier, there is uncertainty concerning the locking condition of the NAF is the western Marmara Sea. Four hypotheses of locking conditions ranging from fully creeping to fully locked are explored in a logic tree to represent the current state of knowledge. The fault parameters for each hypothesis are presented in table 2.1.

The logic tree

SHERIFS is run with the hypotheses of each of the branch of the logic tree (Figure 2.6). We then compare the modelled annual earthquake rates with the seismicity rates from the earthquake catalog and the paleoseismic records.

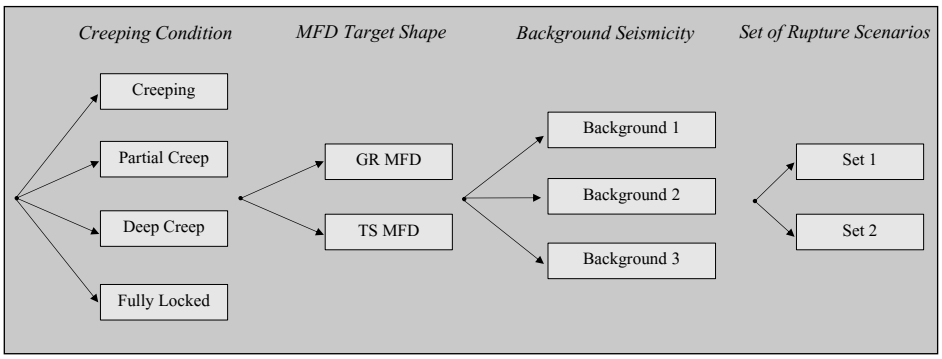

Figure 2.6: *Logic tree explored in this study. For each branch, the scaling law parameters and the slip-rate uncertainties are explored through 10 random samples.*

Results : Earthquake rates modelled by SHERIFS

As we have not considered the entirety of the NAFS but only the faults in the region of Istanbul (figure 2.1), the faults located at the limit of our model are not able to rupture with the sections of the fault we didn't consider. This can involve some modelling issues which will result in a significant portion of the fault slip-rate being spent as NMS slip.

In this study, we only modelled part of the NAF (figure 2.1), therefore some boundary effects might occurs on the faults that would ruptures with fault outside of the modelled zone. With the aim of modelling the seismic risk in Istanbul, far from the system boundaries, we compare only the rates for a region smaller than the whole modelled system centered on Istanbul (Figure 2.2) that is not affected by boundary effects. In this zone, we extract the earthquakes from the catalog and the modelled rate of ruptures. For ruptures that are only partly located in the zone, the rate is corrected by taking into account the proportion of the fault in the zone.

When compared with the rates calculated from the earthquake catalog for the central zone (figure 2.2), the rates modelled using SHERIFS are slightly higher than the catalog (figure 2.7). However, some combinations of hypotheses lead to better fit to the rate from the catalog than others. The branches using the GR target shape lead to a rate of earthquakes of magnitude between 4.5 and 6.5 that is larger than the observed rate for all models (figure 2.7). The tuned shape (TS) hypothesis is in better agreement with the catalog for this range of magnitudes. Both MFD hypotheses reproduce the rate of large earthquakes equally well. The branches using the Set 2 of rupture scenarios, allowing ruptures up to magnitude 8, have lower earthquake rates than the branches using Set 1 and obtain a better fit with the rates from the catalog. The hypothesis on the creep hypothesis has only a small effect on the earthquake rates

at the level of the central zone. The four hypotheses lead to similar rates (figure 2.7). In conclusion, the branches using the combination of the TS MFD and the Set 2 of ruptures lead to the best fits with the rates from the earthquake catalog.

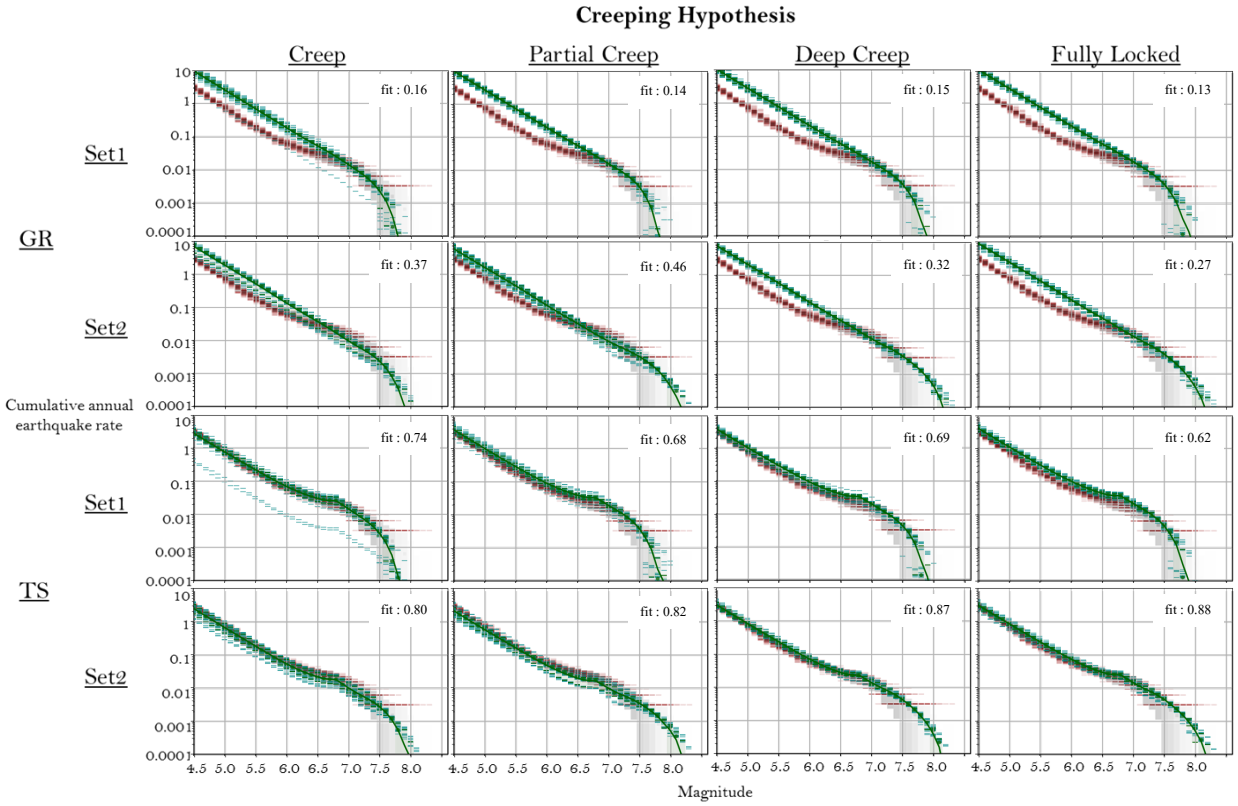

Figure 2.7: *Comparison between the rates calculated from the earthquake catalog (in red with uncertainties in grey) and the model rates (in green) for the central zone of the fault system, close to Istanbul (figure 2.2). The dashed green lines are the MFD for each individual model of the logic tree. The dotted green line are the 16th and 84th percentiles of the distribution and the continuous green line is the mean value of the distribution. Each figure shows the match between model and data for each branch of the logic tree, organized as a table. A fit of 1. is a perfect fit and a value close to 0. expresses a very poor fit.*

In Figure 2.8, we compare the participation rate of fault sections where an estimate of the rate of earthquakes is available based on paleoseismic studies. The participation rate is the rate of all the ruptures for which the given fault section is involved in. For most sites, the rate of earthquakes modelled is in general agreement with the rate deduced from the paleoseismicity for all branches of the logic tree.

At the Ganos 1 site (Figure 2.9), the branch "Creep", assuming a complete creep in the west of the Marmara Sea, leads to rupture rates significantly lower than the paleoseismic rates. For this branch of the logic tree, the Ganos fault spends around 46% of its slip-rate budget as Non Main Shock (NMS) slip. The proportion of NMS slip is the proportion of slip-rate budget that couldn't be spent in seismic moment rates in SHERIFS. A large NMS (>20-30%)

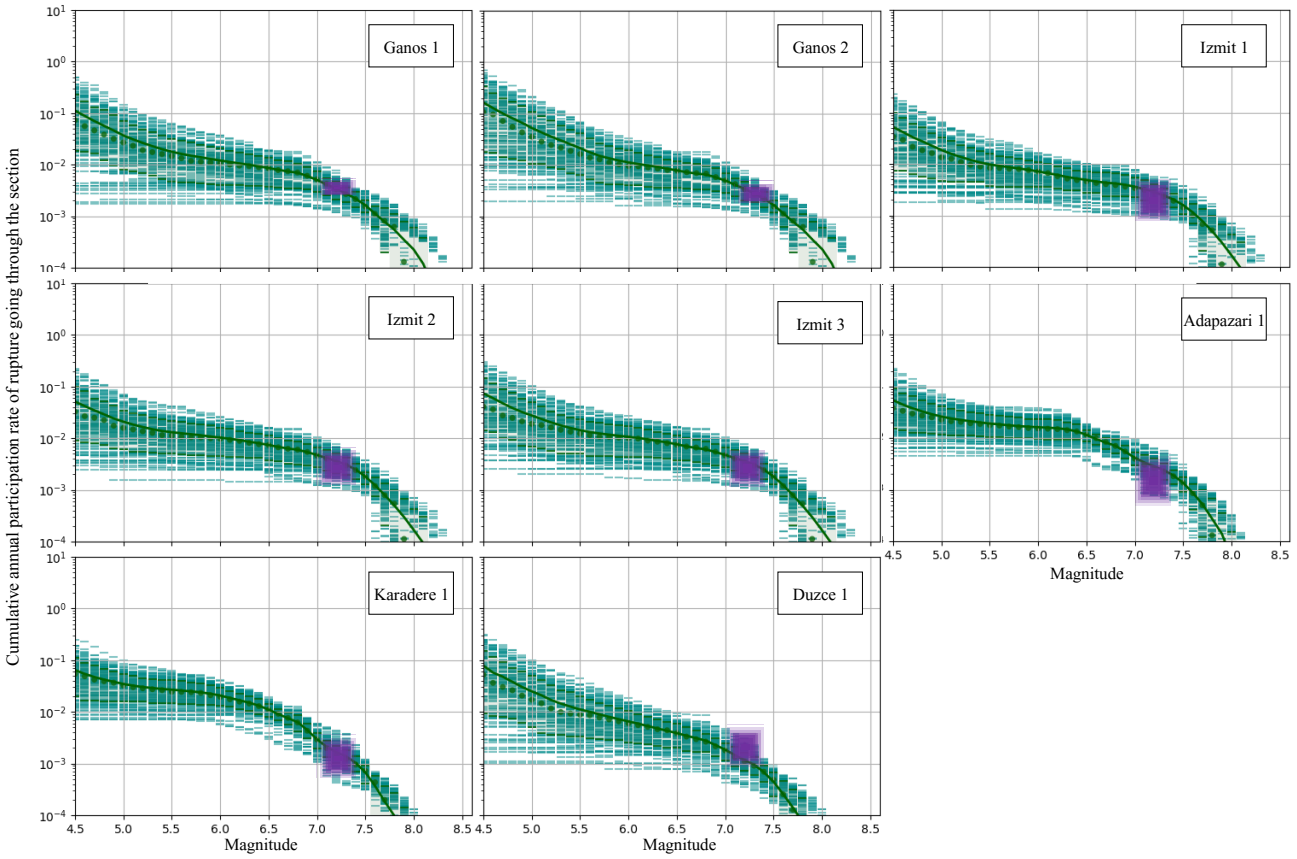

Figure 2.8: *Comparison of the modelled rupture rates with the rates calculated from the paleoearthquake record at each paleoearthquake site (Figure 2.2). The green curves are the modelled rate of ruptures rupturing the given section. The dashed lines are each individual model of the logic tree. The dotted line are the 16th and 84th percentiles of the distribution and the continuous line is the mean value of the distribution. The paleoearthquake rates and references are exposed in Table 2.3.*

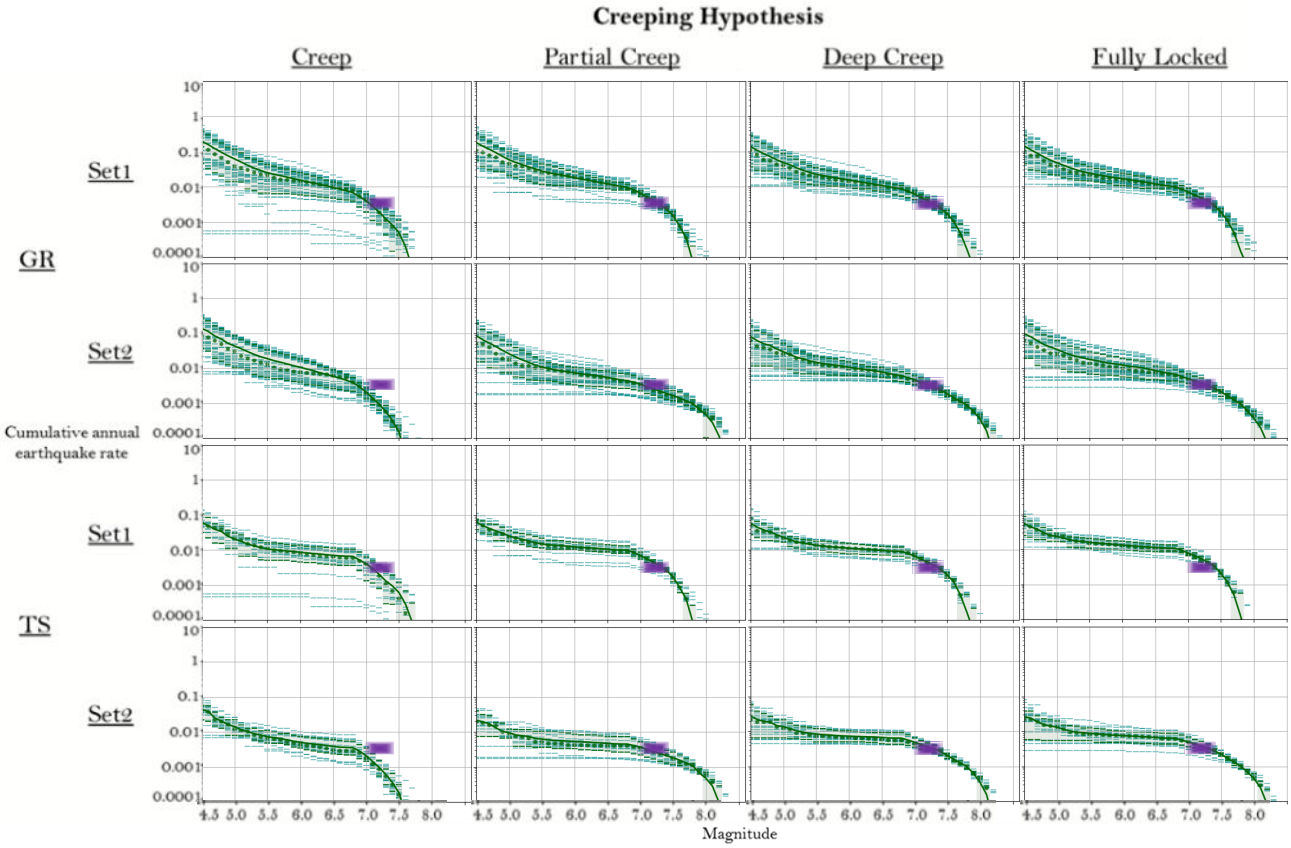

Figure 2.9: *Comparison of the modelled rupture rates with the rates calculated from the paleoearthquake record at Ganos 1 site (Figure 2.2) for different branches of the logic tree. The green curves are the modelled rate of ruptures rupturing the Ganos segment. The dashed lines are each individual model of the logic tree. The dotted line are the 16th and 84th percentiles of the distribution and the continuous line is the mean value of the distribution. The paleoearthquake rates are derived from Rockwell et al 2009 and are represented by the purple box.*

is indicative of the inability in SHERIFS to satisfy the hypotheses of MFD, slip-rate budget and the set of ruptures. The introduction of a fault completely creeping in the western part of the Marmara Sea limits the number of large earthquakes along the Ganos fault. In the SHERIFS framework, the ability of the Ganos faults to release seismic moment jointly with the faults of the Marmara Sea is required in order for it to spend its full budget.

2.3.2 The physics-based approach : RSQSim

Core principle and main hypotheses

The comparison between the modelled earthquake rates and the data allows to reduce part of the uncertainties explored in the logic tree but some uncertainties still remain, notably the uncertainties concerning the MFD shape and the maximum rupture size.

In the hope of reducing these uncertainties, we implemented our fault system in the physics-based simulator RSQSim (Richards-Dinger and Dieterich [2012]) and analyzed the resulting synthetic earthquake catalogs. RSQSim is a boundary element model that applies the rate and state equation (Dieterich [1978]) where each element of a fault system can be in three different states: loading, nucleating the earthquake rupture or sliding. Elements interactions are considered to be quasi-static therefore neglecting the dynamic influence of seismic waves that would be generated by sliding elements. Where the elements are sliding, they are doing so at a constant slip-rate. These simplifications of the earthquake physics and the rate and state law allow RSQSim to be computationally efficient and to generate long earthquake histories for a large fault network.

RSQSim takes as input the same information as SHERIFS concerning the fault parameters but doesn't require a target MFD shape nor a set of possible FtF ruptures. The MFD of the fault system and the ruptures will be deduced from the synthetic earthquake catalog which result from the combination of the faults loading rates and the rate and state friction law as implemented in RSQSim.

In this study, we will use the same friction parameters a and b for all the elements of the fault system (0.01 and 0.015 respectively). These a and b value are based on empirical measurements from laboratory earthquake experiments and have been widely used for modelling earthquake cycles with the rate and state equation (Marone [1998]). The stress loading rate of the faults in the system is defined using back-slip loading. Back-slip loading calculates the stress rate for each element of the fault that corresponds to the long-term slip-rate of the fault.

We are using triangular elements of 1 km size. For this reason, we will only discuss earthquakes larger than magnitude 6 that rupture a number of elements large enough (around 100) to be representative of an earthquake rupture.

Earthquake rates in RSQSim

Using the fault geometry and slip-rate with the fully locked hypothesis (table 2.1, figure 2.2) as input for RSQSim, we simulate a 10 000 years long earthquake catalog from which we remove the first 1000 years of seismicity to let the fault system run through several earthquake cycles before starting the statistical analysis.

Richards-Dinger and Dieterich [2012] compared RSQSim with fully dynamic earthquake simulations and have shown that both modelling approaches lead to similar earthquake ruptures in terms of slip history and final slip on the fault. We observe that the scaling of earthquakes in RSQSim is in good agreement with the scaling laws (Tullis et al. [2012]) in terms of geometrical scaling and of stress drop. Interestingly, RSQSim reproduces quite well the rupture extent and magnitude of the 1999 of Izmit and Duzce earthquakes. Based on these comparisons we conclude that even if the physics of earthquake is simplified to allow for cost efficient calculations, the simulated catalog is representative of

the natural system at least to the first order. In order to not over-interpret the results we limit our use of the RSQSim simulated catalog to the discussion of two first order questions : (1) does the long term MFD of the fault system follows a GR law?, (2) do we observe magnitude 8.0 earthquakes in this fault system?

Considering the MFD of the synthetic catalog, we can reach two conclusions: RSQSim cannot reproduce a GR MFD with the given fault system and the maximum magnitude that can be generated is closer to magnitude 7.7 than to a magnitude 8.0. These two conclusions are not affected by uncertainties on a and b values (figure 2.10).

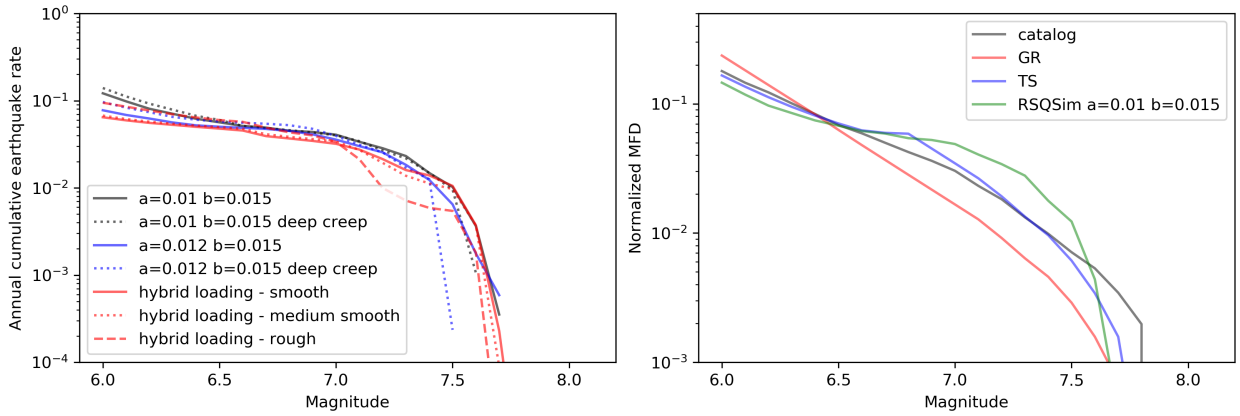

Figure 2.10: *Left : MFD of the synthetic catalogs generated from the RSQSim simulations exploring a range of input parameters. Right : Normalized MFD of the earthquake catalog, the MFD of the faults modelled using SHERIFS with a GR MFD or a TS MFD (the background seismicity has been removed), and the RSQSim synthetic catalog (a=0.01 b=0.015).*

We also explored the impact of having a region of the fault creeping in the Western Marmara Sea. In a way similar to the "deep creep" model used in SHERIFS, in RSQSim, elements that are below 5km depth were attributed a b value much lower than the a value. The resulting MFD (figure 2.10) leads to the same first order conclusions as for the fully locked hypothesis. This is to be expected. Indeed, the region where creep is suspected is of a negligible size compared to the entire fault system and does not drastically affect the shape at the system level MFD. The creeping condition of the fault in the western Marmara sea most likely affects the participation rates of neighboring faults, however, here we only consider the first order results of the calculation.

The results are compared with the hybrid loading method used in Shaw et al. [2018] with the same input parameters as in the Californian application. The hybrid loading method allows to remove the edge effects on the stress field of the fault that occur with the standard back-slip loading (Shaw et al. [2018]). Different smoothness parameters are tested. The synthetic catalogs do not follow GR MFD, strengthening our previous observation.

2.4 Discussion on the weight of the epistemic uncertainties in the logic tree

The flexibility of SHERIFS for modelling the earthquakes rates in the NAFS allowed to explore a wide range of uncertainties in a logic tree framework. The SHERIFS approach doesn't contain any physical constraints; it can therefore accommodate the different input hypotheses that are being discussed by the scientific community but also can lead to a large range of uncertainties on the earthquake rates. Since SHERIFS only uses the fault data as input, it is possible to compare the modelled rates with rates calculated from the earthquake catalog and paleoearthquakes which can be considered independent data from the SHERIFS input.

Branches weights in a logic tree are usually set by judging the scientific value of each hypothesis explored in the logic tree but not on the capacity of the modelled earthquake rates in each branch to reproduce the data. The weight of an individual branch of the logic tree is simply the product of the weights of each hypothesis used for the branch. In this discussion, we propose an innovative approach to set the weight of each branch of the logic tree accounting for both the input hypotheses used and the ability to reproduce the independent data.

We set up a quantitative scoring system in order to set the weight of each individual branch of the logic tree. For each branch, four scores are calculated (figure 2.11). One score judges the ability of the combination of input parameters and hypotheses to spend the slip-rate budget of the faults as earthquake rate and not generate a large amount of NMS (S1). Two scores compare the model to the data, judging the capacity of the modelled earthquake rates to reproduce the earthquake rates deduced from the catalog (S2), and to reproduce the earthquake rates deduced from the paleoseismic records (S3). The last score rates a-priory the input hypotheses based on the analysis of the RSQSim synthetic catalog (S4). The final weight of each branch of the logic tree is given by the combination of the four scores (figure 2.11).

2.4.1 Uncertainty on the creep

The creep conditions assumed on the Terkidag section play a predominant role in the modeling of earthquake rates on the adjacent Ganos section. The models considering the Terkidag fault as completely creeping have difficulties reproducing the paleoearthquake rates estimated on the Ganos fault (figure 2.9) and lead to a high level of NMS on this fault. In the fully creeping model, the Ganos section is only able to convert 54% of its slip-rate in seismic moment rate and 46% is considered NMS slip. In the partly creeping model, the Ganos section spends 71% of its slip-rate as moment rate. In the deep-creep model and the fully locked model, the Ganos section spends 90% of its slip-rate as seismic moment rate. We calculate the NMS value for each model and each fault section of the central zone. Since

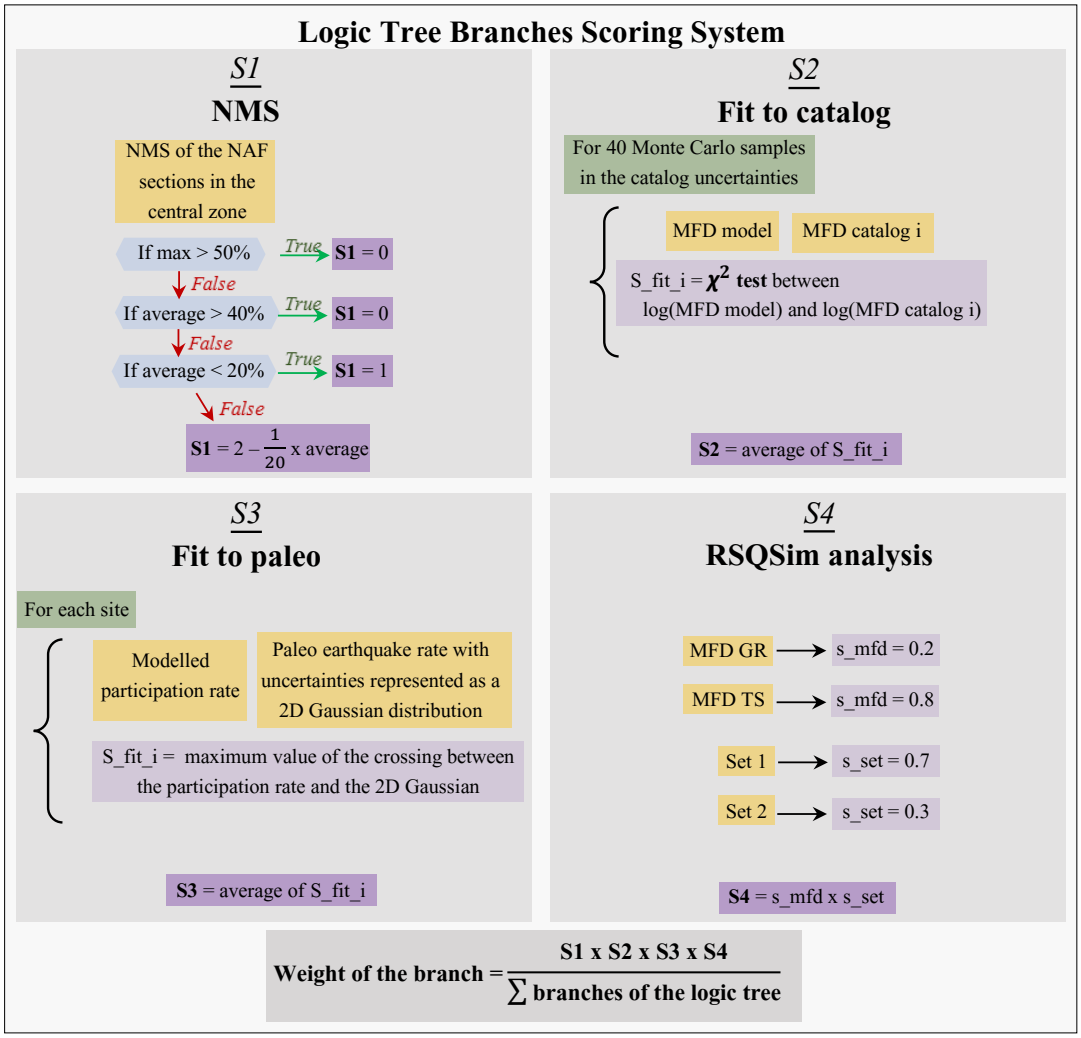

Figure 2.11: *Scoring system allowing the individual weighting of each branch of the logic tree according to both its capacity to spend the fault slip rate budget into earthquake rate, its capacity to reproduce the earthquake rates observed in the data and the agreement between the input hypotheses and the results of the discussion on the physics-based synthetic catalog.*

a large NMS value is likely linked to incompatibilities between input hypotheses, given the SHERIFS framework, a low score is attributed to models with high NMS slip value. For a given model, if the mean NMS value for the fault sections of the central zone is greater than 40% or if the NMS value of one of the faults of the central region is greater than 50%, the score is zero. The score is 1 if the average NMS of the sections of the central region is less than 20%. Between the average values of 20% and 40%, the score linearly decreases from 1 to 0 as the NMS value increases. As a result, the average score of the fully creeping models is 0.27, the average score of the partly creeping models is 0.6 and the average scores of both the deep creep and fully locked models are 0.96.

Figure 2.9 exposes the underestimation of the rate of earthquakes on the Ganos section compared to data when considering the Creep branch of the logic tree. The comparison with the paleoseismicity is done at all sites where

paleoseismicity is available for each model of the logic tree (see above). At each site, the rate of paleo-earthquake is extracted from the literature (table 2.3) and represented as a 2D Gaussian distribution in order to capture the uncertainty both in terms of magnitude and annual rate of occurrence. At each site, the modelled rates are compared to this distribution. The closer the participation rate for a given section is from the maximum of the paleoearthquake rate distribution the larger the score. The score of the branch is the average score for each paleo-site. The scores for the Creep, Partial Creep, Deep Creep and Fully Locked models are 0.65, 0.73, 0.82 and 0.81 respectively.

2.4.2 Uncertainty on the shape of the MFD target

The models imposing a GR MFD as a target shape for the MFD of the fault system show a good agreement with the rate of large earthquakes calculated from both the catalog and the paleoseismicity but overestimate the rate of small to intermediate earthquakes.

However, it can be argued that the deviation of the MFD shape from the GR shape is an artifact of the observation period, which is too short for accurately capturing the full earthquake cycle. Such arguments have been brought forward in California in support for a GR MFD target shape even when the apparent shape of the earthquake catalog close to the faults might differ from the GR shape (Page and Felzer [2015]). Stirling and Gerstenberger [2018] have shown that for four fault systems in New Zealand, while the data doesn't show a good agreement with the GR MFD hypothesis, ETAS simulations suggest that the observation time in the data is too short to accurately calculate the long term rate of a clustered earthquake catalog.

However, the 10 000 years long synthetic catalog generated by RSQSim, complete and representative of the seismic cycle by definition, also diverges from the GR MFD shape. In the RSQSim catalog, there are almost no earthquakes of magnitude between 6.5 and 7.0. According to the scaling laws (Thingbaijam et al. [2017]), 6.5 is the magnitude of an earthquake rupturing the whole fault width. We speculate that in RSQSim, an earthquake rupturing the entire width of the fault, is more likely to rupture neighbouring faults unless stopped by geometrical complexities or the presence of faults that are still in the early part of their cycle. The average length of the straight sections of the faults is around 100 km (figure 2.1), which corresponds to an earthquake of magnitude around 7.2. Ruptures are able to overcome bends and asperities if the neighboring fault is close to rupture leading to larger magnitude earthquakes. Since they require several sections to be close to rupture simultaneously, these ruptures are less frequent. We obtain therefore a bi-modal distribution with the earthquakes up to magnitude 6.3 following a GR, rupturing different portion of the fault plane, very few earthquakes between magnitude 6.3 and 7.0, and a population of earthquakes larger than magnitude 7.0 with the number of earthquake decreasing with the magnitude.

Based on the comparison between the modelled rates with the rates of the earthquake catalog and the results

of RSQSim, and in consideration of the scientific debate around the MFD, we suggest a stronger score (s_mfd) on the "Tuned shape" MFD branch. The branch exploring the "GR MFD" should remain part of the logic tree, however. Consequently, we set the score of the branch "Tuned shape MFD" to 0.8 and the score of the branch "GR MFD" to 0.2.

2.4.3 Uncertainty on the largest possible rupture

Based on the comparison between the modelled rates from SHERIFS and the rate calculated from the earthquake catalog and the paleoseismicity studies, it is not possible to weight differently the two branches of the logic tree exploring the uncertainty on the rupture scenarios. While the fit between the modelled rates and the catalog rates using the Set 2 of ruptures, allowing larger ruptures, leads to a slightly better fit than the models using the Set 1 of ruptures, both fits can be considered as satisfying (figure 2.7). Whether ruptures up to magnitude 8.0 are considered or not, the rates in the models are matching the rates from the data when using the TS MFD as a target.

In the several 10 000 years long catalogs simulated by RSQSim, we do not observe earthquake larger than 7.7. In the SHERIFS models, earthquake with magnitude larger than or equal to 8.0 have an annual frequency of 5 10-4. The likelihood of observing at least one earthquake in a 10 kyr long catalog is 99.3 %. While the bends and other geometrical complexities of the fault trace are crossed by ruptures in the model, no rupture crosses enough complexities to generate such large earthquakes. We can assume that the fault system is very chaotic and having a large number of neighbouring fault sections close to rupture is very unlikely. However, it is important to recall that RSQSim does not fully model the dynamic effects of the earthquake rupture. It is possible that this simplification in the modelling leads to underestimating the likelihood of a rupture propagating through fault complexities. In consideration of these arguments, we set the s_set score of the branch "Set 1" to 0.7 and the s_set score of the branch "Set 2" to 0.3.

2.4.4 The weighted logic tree

In the discussion, we established four types of scores for the logic tree branches: the weights established from the analysis of the RSQSim synthetic catalogs and those established from the comparison of the earthquake rates modelled using SHERIFS with the rates calculated from the earthquake catalog and the paleoearthquake record. For each individual branch of the logic tree, these scores are convolved into a final weight unique to the branch (figure 2.11).

The final weight of each hypothesis is calculated by summing all the branches using a given hypothesis (figure 2.12). Due to the high NMS value of the models using the Creeping hypothesis and their underestimation of the paleoearthquake rates, this hypothesis has the lightest weight in the logic tree.

The weights of the MFD hypothesis branches and the Set of Rupture scenario branches are strongly influenced

by the weights imposed after the analysis of the RSQSim synthetic catalog. While the fit to the catalog is better with the TS MFD hypothesis, the fit to the paleoearthquake rates and the ratio of NMS is similar for both branches. Therefore the comparison with the data affects the weight ever-so slightly. The same can be said about the set of rupture scenario branches.

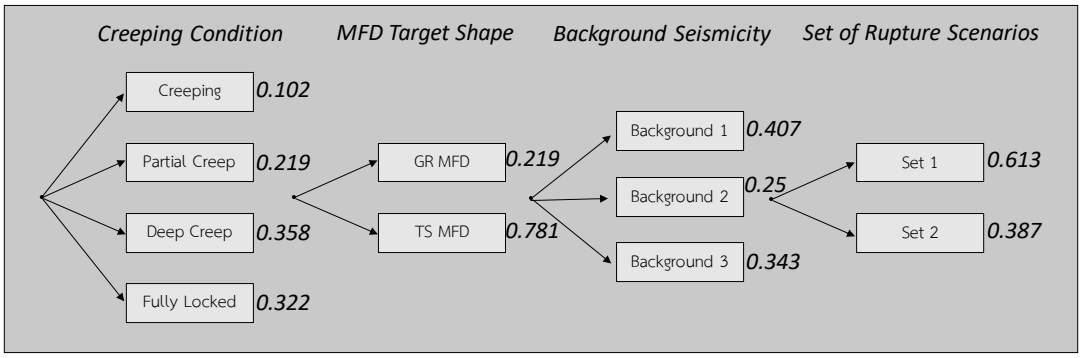

Figure 2.12: *Weighted logic tree established in this study. For each branch, the scaling law parameters and the slip-rate uncertainties are explored through 10 random samples. The weight of each branch is indicated by the bold number.*

The weights of the background hypothesis branches are only affected by the comparison with the data. Overall, the background 1 hypothesis reproduces the earthquake rates better than the other two background hypotheses. The scores for the fit to the catalog, the fit to the paleoearthquake rates, and the NMS ratio are respectively 0.54, 0.77 and 0.77 for the background 1; 0.49, 0.76 and 0.70 for the background 2 ;and 0.48, 0.72 and 0.61 for the background 3. In figure 2.7, we can observe that the modelled MFD tends to be slightly higher than the catalog MFD for the majority of models. Therefore, the background 3, adding more seismicity to the background than the two other hypotheses (table 2.4) necessarily leads to earthquake rates that are in general slightly higher than the observed rates. It is worth noting that we are discussing small differences in trends that are correctly captured by the weights in the logic tree with the three hypotheses having similar weights with a slightly stronger weight for background 1.

The impact of the weights on the annual earthquake rates are presented in figure 2.13. The weighting leads to lower earthquake rates for the small and intermediate magnitudes and larger rates for the magnitudes greater than 7.0. This is due to the reduced weight of the branches using the GR hypothesis.

2.5 Conclusion and perspectives

In this study, we calculated earthquake rates in the Marmara region relying on the fault slip-rate and geometry as primary information. We combined two innovating approaches, the SHERIFS approach that relies on statistical rules and the RSQSim approach that relies on physical rules.

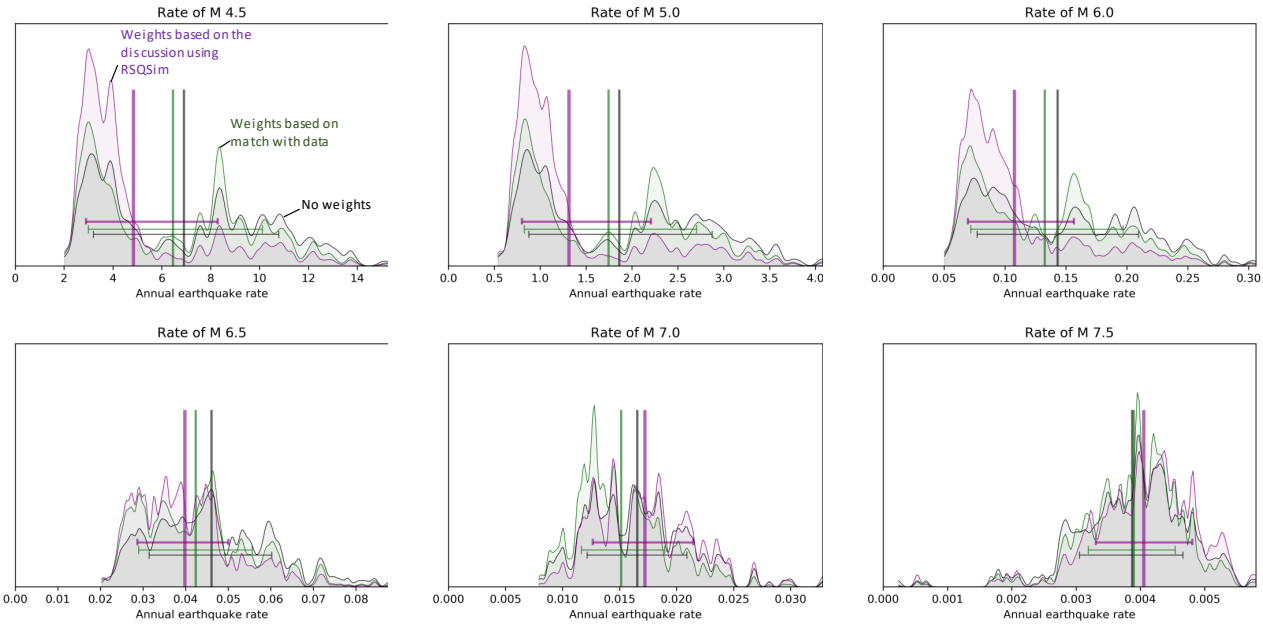

Figure 2.13: *Distribution of the cummulative earthquake rates in the logic tree for the central zone for different magnitudes. In black : the density distribution when only the scoring according to the NMS is applied. In green : the density distribution weighted according to the fit with the data (catalog and paleo). In purple : the final distribution of earthquake rates, weighted according to the match with the data and the discussion using RSQSim. The vertical bar shows the mean annual earthquake rate of each distribution. The horizontal bar extends between the 16th percentile and the 84th percentile of each distribution.*

With SHERIFS, we explored an extensive logic tree of uncertainties concerning the locking condition of the NAF fault in the Marmara region, the shape of the MFD, the ratio of seismicity between the background and the faults, the largest possible rupture as well as uncertainties on the slip-rates and the maximum magnitude predicted by the scaling law. Rather than basing the weights of the branches of the logic only on the expert judgment of each hypothesis, we also take advantage of model performance by comparing results with data (earthquake catalog and paleoearthquake) and weigh the branch accordingly. In addition, the analysis of the synthetic catalog simulated by RSQSim showed a MFD diverging from a GR and didn't contain any earthquakes of magnitude larger than 7.7. This allowed us to attribute a stronger weight to the branches of the logic tree showing similar features leading to a final weighted distribution of modelled annual rates that properly represents the state of knowledge of the NAFS in the vicinity of Istanbul (figure 2.13).

In the next chapter of this study, we carry out a PSHA using the weighted logic tree and calculate the risk of collapse of a building in Istanbul. We expose the impact of each uncertainty of the logic tree on the probability of collapse and we deaggregate the risk results in order to identify the sources controlling the seismic risk.

Chapter 3

Calculating the Probability of Collapse of a Building in Istanbul

Résumé du chapitre

En utilisant les taux de sismicité modélisés dans le chapitre 2, nous calculons la probabilité d'effondrement d'un bâtiment à Istanbul sur une période de 50 ans. Pour chaque branche de l'arbre logique, nous calculons grâce au calcul de l'aléa probabiliste, la courbe d'aléa au site, décrivant la probabilité d'être dépassée pour un ensemble d'accélérations du sol. En convoluant cette courbe d'aléa avec la courbe de fragilité du bâtiment, qui décrit la probabilité d'effondrement du bâtiment en fonction de l'accélération du sol, nous obtenons la mesure du risque qui est la probabilité d'effondrement.

Nous calculons cette probabilité d'effondrement pour deux types de bâtiments représentant la majorité des bâtiments à Istanbul: un immeuble de moyenne élévation (3 à 7 étages), en béton armé, l'un construit en suivant le code de construction de 1975 (TEC75) et l'un construit en suivant le code de construction de 1998 (TEC98).

La probabilité d'effondrement du bâtiment TEC75 en 50 ans est de 2.8 à 8.1% et celle du bâtiment TEC98 est de 0.4 à 1.3%. L'évolution du code de construction entre 1975 et 1998 réduit donc le risk d'effondrement par un facteur 6. La pondération de l'arbre logique en utilisant la comparaison aux données et au catalogue synthétique de RSQSim a permis de réduire l'incertitude par un facteur 1.6 pour le bâtiment TEC75 et par un facteur 1.8 pour le bâtiment TEC98.

Grâce à l'utilisation de l'arbre logique, nous calculons l'impact de chaque incertitude explorée sur l'estimation du risque. L'incertitude sur la forme de la MFD entraine le plus d'incertitude sur la probabilité d'effondrement, suivie par le choix du set de rupture, de la GMPE et de l'hypothèse de sismicité de fond. A cause de la distance d'environ 70

km entre la zone de creep potentiel et le site, l'incertitude sur l'hypothèse de creep n'entraine que peu d'incertitude sur l'estimation du risque à Istanbul.

Nous présentons une méthode de désagrégation du risque sismique permettant d'identifier les sources sismotectoniques contrôlant la probabilité de collapse. Pour le cas du bâtiment à Istabul, deux sources participent au risk d'effondrement: les séismes de magnitudes supérieures à 7 sur la section de la faille Nord Anatolienne la plus proche et les séismes de magnitudes inférieures à 6.5 dans la zone de sismicité de fond à une distance inférieure à 10 km du site. Les séismes de magnitude 4.5 à 5.5 participent à la probabilité d'effondrement du bâtiment TEC75 mais ne génèrent pas de mouvements du sol suffisant pour entrainer l'effondrement du bâtiment TEC98.

L'approche développée dans ce chapitre nous permet d'identifier les pistes de recherches qui permettront d'améliorer la connaissance du risque sismique à Istanbul.

3.1 Introduction

The city of Istanbul being located in the vicinity of the North Anatolian Fault (NAF) (figure 3.1), the seismic hazard estimate has been a growing concern over the past decades (Erdik et al. [1999]). Since the 1999 Izmit earthquake that occurred on the NAF, East of the Marmara Sea, about 50 km east of Istanbul, several studies have assessed the seismic hazard in Istanbul (Erdik et al. [2004], Woessner et al. [2015], Giardini et al. [2018], Sesetyan et al. [2018]). These studies provided results in the form of hazard maps of the Peak Ground Acceleration (PGA) and of various frequencies of Pseudo Spectral Acceleration (PSA) that have a 10% or 2% probability of being exceeded in 50 years.

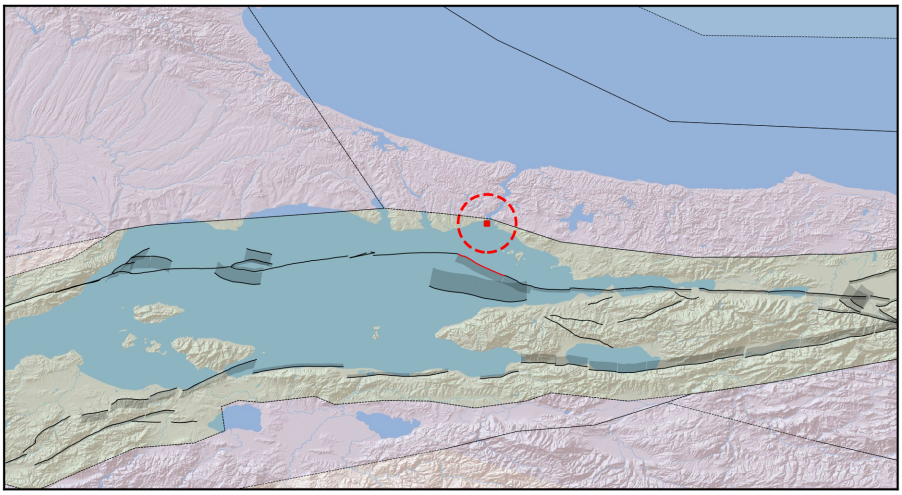

Figure 3.1: *Map of the Marmara region. The faults are indicated in black (see chapter 2). The site of calculation is the red dot. The red circle has a 10 km radius around the site of calculation. The background zone of the fault system is in green. The SHARE host model(Woessner et al. [2015]), that is to say the area zones for the regions surrounding the fault system are in shaded blue.*

In this study, rather than presenting hazard metrics, we present the result of the analysis in terms of a risk assessment metric, the probability of collapse in a period of 50 years for a building in Istanbul. Probabilistic seismic risk studies have been used in the past but they have been carried on by structural engineers, thus focusing more on the modelling of the building response to the ground motion rather than the modelling of the seismic sources. In this study, we will focus our attention on the impact that different hypotheses on the modelling of the seismogenic sources have on the risk assessment.

Using a risk targeted approach, Iervolino et al. [2018] show that "generally, design procedures are such that seismic structural reliability tends to decrease with increasing seismic hazard of the building site, despite the homogeneous return period of exceedance of the design seismic ground-motion". Despite using a building code constructed using a hazard that represents the PGA at 10% over 50 years over the whole Italian territory, the risk of damage of buildings is much higher in area where the hazard is higher. Using a hazard map presenting a uniform

probability of exceeding a given ground motion (e.g. 10% over 50 years) can mislead the audience of the map (decision maker and the public) and make them think that the risk level is also uniform over the territory. Iervolino et al. [2018] showed that, in Italy, the risk level is not uniform.

In order to increase the chances of seismic hazard results reaching a wider audience, the company Temblor (through it's website `temblor.net`) presents risk results that integrate both the hazard results and the building response. This allows to frame the question in terms that can be more easily understood by non-PSHA practitioners.

Hong and Goda [2006] have shown that when looking at the risk for a specific building, the seismogenic sources controlling the hazard at a site can differ from the sources controlling the risk.

Most PSHA studies present the hazard as a 10% chance of exceedance in 50 years which is the classical value used for the establishment of codes for residential buildings. This value of 10% is set as an acceptable level of risk to be taken for general use buildings. For critical structures such as chemical of nuclear plans this value is 2%, 0.5% or even lower. However, since this value on the hazard level doesn't directly reflects the risk level (Iervolino et al. [2018]), presenting the hazard without consideration of the risk can be misleading for the decision making process.

In more general terms, the hazard map poorly answers the question posed by the general public and decision makers alike: what is the probability of damage for a building during a given period of time?

In this study, the seismic hazard (probability of exceeding different levels of ground motion) is calculated using the earthquake rates modelled in the chapter 2 exploring a wide logic tree of epistemic uncertainties (figure 3.2, table 3.1). We also explore uncertainties on the Ground Motion Prediction Equation (GMPE) through a logic tree set up by Sesetyan et al. [2018].

Using the probability of exceeding different levels of ground motion (e.g. the hazard curve) and the probability of a building to collapse given a ground motion level (e.g. the fragility curve), we calculate the probability of collapse for a building in Istanbul over a period of 50 years. We perform the calculation using two different fragility curves both corresponding to a 5 floors reinforced concrete building but for two different building codes, the TEC75 building code and the TEC98 building code (Ozmen et al. [2010]).

We track the impact on the probability of collapse of each uncertainty explored in the logic tree and show the leading sources of uncertainty on the probability of collapse.

Using a novel approach for risk deaggregation, we expose the earthquake sources that control the probability of collapse of a building in Istanbul for different models and input hypotheses.

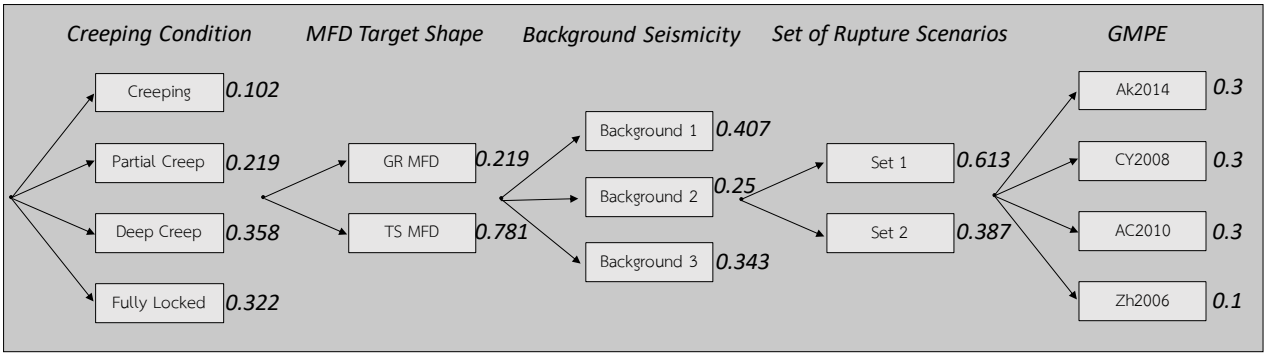

Figure 3.2: *Weighted logic tree established in chapter 2. For each branch, the scaling law parameters and the slip-rate uncertainties are explored through 10 random samples. The weight of each branch is indicated by the bold number. GR and TS MFD: Gutenberg-Richter and Tuned Shape Magnitude frequency distribution, Creeping condition in west of the Marmara, Sea Ak2014: Akkar et al. [2014], CY2008: Chiou and Youngs [2014], AC2010: Akkar and Çağnan [2010], Zh2006: Zhao et al. [2006]*

3.2 Seismic hazard in Istanbul

In this study, we assess the seismic hazard at a site in Istanbul (figure 3.1) using the Cornell McGuire approach (Cornell [1968], McGuire [1976]) based on the following equation 1.

$$P(A > a^*) = \int_M \int_D \int_\epsilon \lambda_{M,D} \times P(A > a^* | M, D, \epsilon) \tag{3.1}$$

With $\lambda_{M,D}$ the annual rate of earthquakes of magnitude M at the distance D; and $P(A > a^*|M, D)$ the probability of the ground motion A to exceed the target level a^* given the magnitude M at the distance D. $P(A > a^*|M, D, \epsilon)$ is obtained using a Ground Motion Prediction Equation (GMPE) with the ϵ value . ϵ is defined as the number of standard deviations by which an observed logarithmic spectral acceleration deviates from the mean logarithmic spectral acceleration of a GMPE.

Based on the earthquake rates modeled using SHERIFS (chapter 2), we conduct the Probabilistic Seismic Hazard Assessment (PSHA) using the code Openquake (Pagani et al. [2014]). Each branch of the Seismic Source Characterization (SSC) logic tree set up in the chapter two is ran with four Ground Motion Prediction Equation (GMPE). This GMPE logic tree was constructed in a former PSHA of Turkey (Sesetyan et al. [2018]) and we chose to use it with the same GMPE weights (figure 3.2).

The logic tree presented in figure 3.2 explores a large number of uncertainties on the input hypotheses of the fault modelling (table 3.1). This translates into a large uncertainty in the hazard result (figure 3.3). In this study, we will not expose the impact of each uncertainty on the hazard metric but rather we will present how each uncertainty

Logic Tree node	Hypothesis name	Short description
Locking Model	Creeping	NAF creeping in the Western Marmara Sea
	Partial Creep	NAF 50% creeping in the Western Marmara Sea
	Deep Creep	NAF creeping at depth in the Western Marmara Sea
	Fully Locked	NAF is locked
Magnitude Frequency Distribution	GR	The modelled seismicity follows a Gutenberg Richter distribution
	TS	The modelled seismicity follows a distribution similar to the one observed in the earthquake catalog
Background Seismicity		Ratio of the seismicity on the faults for each magnitude range
		4.0 | 4.5 | 5.0 | 5.5 | 6.0 | 6.5 | 7.0 | 7.5 | 8.0+
	BG_1	0.8 | 0.8 | 0.8 | 0.8 | 0.8 | 0.9 | 1.0 | 1.0 | 1.0
	BG_2	0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.8 | 0.9 | 1.0 | 1.0
	BG_3	0.4 | 0.4 | 0.4 | 0.45 | 0.5 | 0.6 | 0.8 | 1.0 | 1.0
Scenario Set	Set_1	The largest rupture possible in the system are in the size of the largest historical earthquakes (M7.7)
	Set_2	The largest rupture possible in the system (M8.0) exceeds the historical earthquakes

Table 3.1: Description of the hypotheses explored in the logic tree.

impacts the risk, expressed as the probability of collapse of a building.

3.3 Methododology for the calculation of seismic risk

3.3.1 Probability of damage

In this study, we set up an approach to explore the impact of the different epistemic uncertainties considered in the logic tree through the lens of the probability of collapse. The hazard curve resulting from each branch of the logic tree is convolved with a fragility curve describing the building response to the seismic solicitation. The hazard curve shows the probability of exceedance of a given ground motion (e.g. PGA) during a given time period (e.g. 50 years). From this hazard curve, we create an incremental hazard curve giving the probability of a ground motion to be the strongest

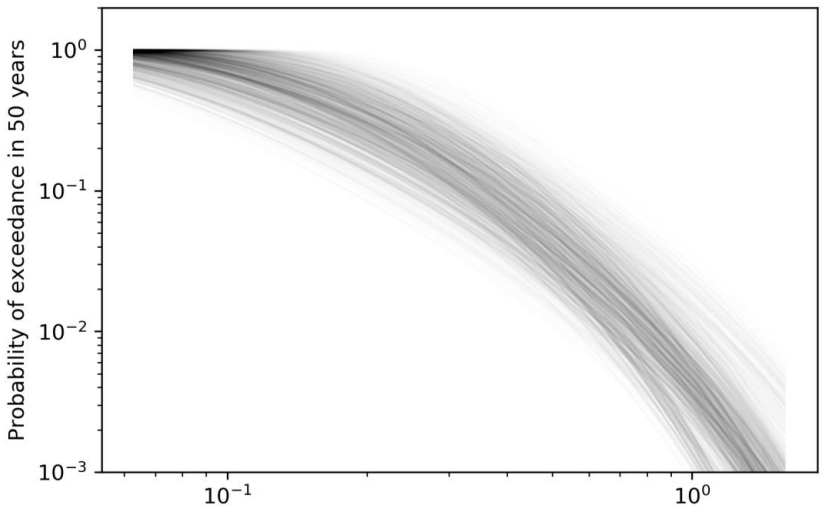

Figure 3.3: *Hazard curves calculated for each branch of the logic tree presented in figure 3.2. The opacity of the curve represents the weight of the branch in the logic tree.*

ground motion observed during the given time period. This probability is then multiplied by the probability of the ground motion to damage the building according to the predictions of a given fragility curve. The integration on the ground motion level gives us a probability of damage of the building during the time period (equation 2).

$$P(damage) = \sum_{A} P(A = a_n) \times P(damage|a_n) \tag{3.2}$$

With A the ground motion level, P(A) the probability of the ground motion being the largest ground motion observed in the time period according to the hazard curve, P(damage|A) the probability of damage with a ground motion of A according to the fragility curve.

3.3.2 Risk deaggregations

Convolving hazard and fragility curves provides a single risk indicator, such as the probability of damage, give us a better idea of how different input hypotheses impact risk, it also makes it more difficult to understand what are the earthquake sources contributing the most to this risk. Similarly to the deaggregations developed by McGuire (McGuire [1995]) for hazard deaggregation, we identify the participation of each source to risk by developing a methodology to deaggregate the probability of damage. First, deaggregations of the hazard results are carried out for a range of ground motions, ranging from very small ground motion with a negligible change of damaging a building to very large ground motions with a negligible probability of exceedance. Each deaggregation gives the conditional probability of exceedance of the ground motion for each range of distance, magnitude, and epsilon value (eq 4). The epsilon value ϵ is the number of standard deviation from the mean ground motion predicted by the GMPE. $\epsilon = 0$ indicates the mean

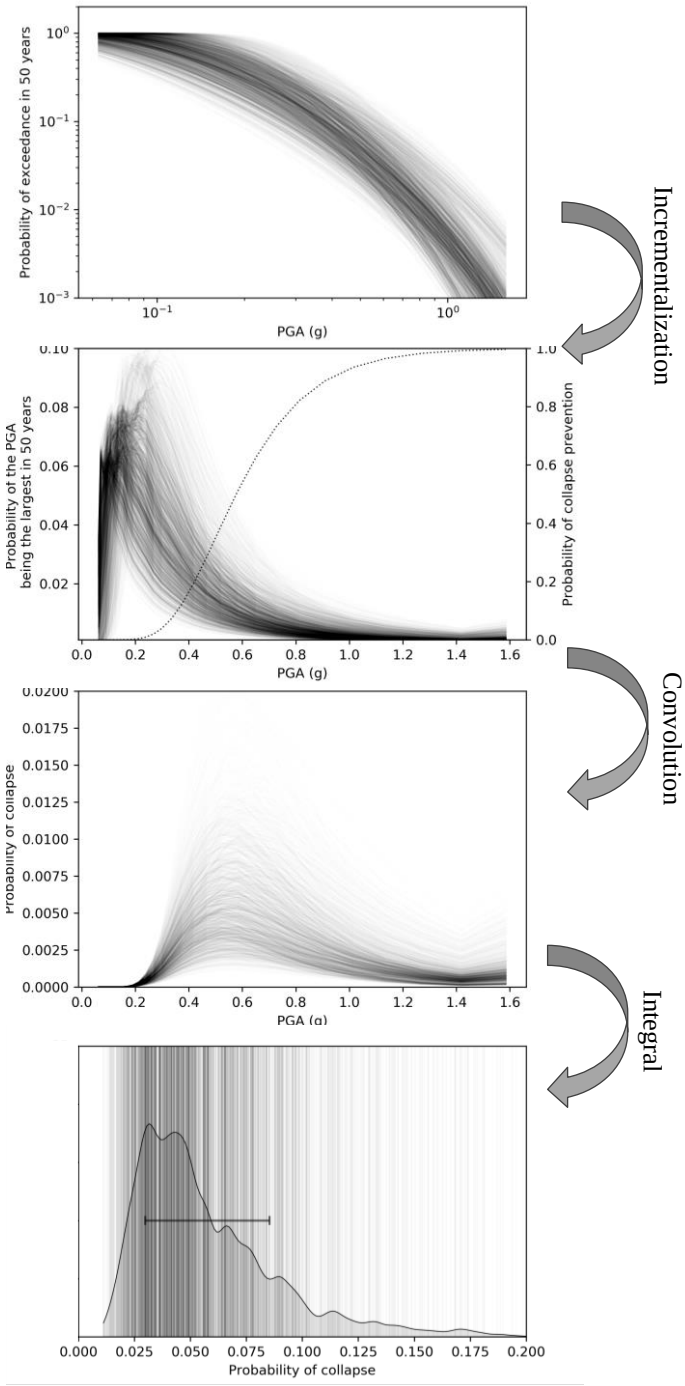

Figure 3.4: *Calculation of the probability of collapse. The opacity of each curve reflects the weight of the branch in the logic tree. From top to bottom: (a) Hazard curves obtained using the logic tree presented in figure 3.2. (b) Incremental hazard curves giving the probability of every PGA value to be the largest observed in 50 years. The dotted line is a building fragility curve linking the probability of the building to collapse or requiring collapse for a given PGA. (c) Convolution of the incremental hazard curves and the fragility function. For each model of the logic tree, the probability of collapse the integral of this curve. (d) Distribution of the probability of collapse in 50 years. Vertical lines indicate the probability of collapse for each individual branch.*

value participates to the hazard, ϵ = -1 means that the values of acceleration one standard deviation smaller than the mean participate to the hazard, and a $\epsilon = 1$ means that acceleration one standard deviation larger than the mean participate to the hazard.

$$P(A > a_n) = \sum_D \sum_M \sum_\epsilon P(A > a_n | M, D, \epsilon) \tag{3.3}$$

In order to obtain the conditional probability of the ground motion to be the largest observed in the time period of interest, each conditional probability is reduced by the conditional probability of exceeding a stronger ground motion (eq 4)

$$P(A = a_n) = \sum_D \sum_M \sum_\epsilon P(A > a_n | M, D, \epsilon) - P(A > a_{n+1} | M, D, \epsilon) = \sum_D \sum_M \sum_\epsilon P(A = a_n | M, D, \epsilon) \tag{3.4}$$

Where, $P(A > a_n | M, D, \epsilon)$ is the conditional probability of exceeding the ground motion a_n given the magnitude M, the distance site to source D and ϵ value; and $P(A > a_{n+1} | M, D, \epsilon)$ is the conditional probability of exceeding a ground motion such as $a_{n+1} > a_n$.

In order to obtain a precise result, the hazard deaggregation ($P(A > a_n | M, D, \epsilon)$) needs to be calculated for a large number of ground motion a_n. However, because deaggregation can be computationally expensive, the number of ground motion level calculated needs to be optimized depending on the computational capacities.

$$P(damage | M, D, \epsilon) = \sum_n P(A = a_n | M, D, \epsilon) \times P(damage | a_n) \tag{3.5}$$

where $P(damage | M, D, \epsilon)$ is the probability of a building being damaged over the time period of interest for a given set of magnitude, distance and epsilon value.

Deaggregating the probability of damage provides a better understanding of the sources contributing to the seismic risk for a building and allows a more efficient targeting of future studies aiming to reduce risk uncertainties.

This approach also offers the possibility to use magnitude, distance and/or source dependant fragility curves where the the probability of damage $P_{M,D,\epsilon}(damage | a_n)$ can be calculated using the following equation if such fragility curves are available.

$$P(damage) = \sum_M \sum_D \sum_n P(A = a_n | M, D) \times P_{M,D}(damage | a_n) \tag{3.6}$$

3.4 Results

3.4.1 Probability of collapse for a building in Istanbul

Several fragility curves describing the building response to ground motions are available for Turkish buildings. In this study, we use the fragility curves for mid-rise reinforced concrete (RC) buildings developed by Ozmen et al. [2010]. RC buildings is the most common construction type in the Marmara region with 75% of the total building stock (Bal et al. [2008]). Fragility curves provide the probability of the building being in a state of "collapse prevention" where structural elements of the building have been damaged and the reuse of the building is not safe. In order to simplify the language, we will address the "probability of collapse prevention" as the "probability of collapse". Ozmen et al. [2010] provide fragility curves for two types of RC buildings, following the Turkish building code established in 1975 (TEC75) and following the building code established in 1998 (TEC98). The probability of collapse according to the PGA is shown in figure 3.5. These building models are representative of the majority of building stock in Turkey (Ozmen et al. [2010]). Ozmen et al. [2010] also provide fragility curves giving the probability of the building to be no longer suitable for immediate occupancy and for life safety.

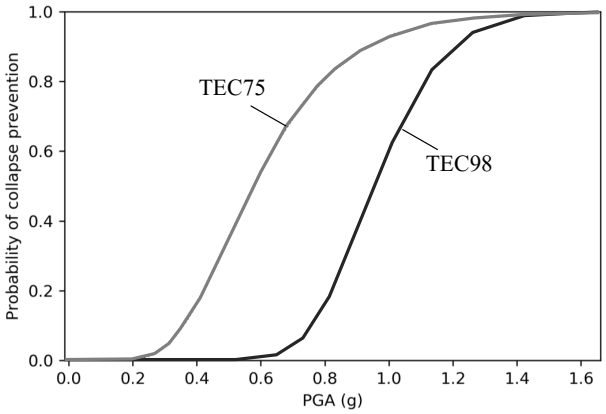

Figure 3.5: *Fragility curves giving the probability of the building requiring collapse prevention for the 1975 Turkish building code (TEC75) and the 1998 Turkish building code (TEC98) (Ozmen et al. [2010]).*

Using the hazard curves giving the probability of exceedance of PGA over 50 years presented in figure 3.4, we calculate the probability of collapse.

Comparing the probability of collapse of a building in 50 years for the two building codes, indicates a reduction of a factor 6 for the TEC98 building with respect to the TEC75 building (figure 3.6). For a TEC75 building, the information gained from adding the comparison to the data in the weighting of the logic tree reduces the uncertainty on the probability of collapse from between 3.4% and 11.8% to between 3.2% and 10.3%. Adding the weighting suggested by the physics-based modelling, the uncertainty on the probability of collapse is brought down to between 2.8% and

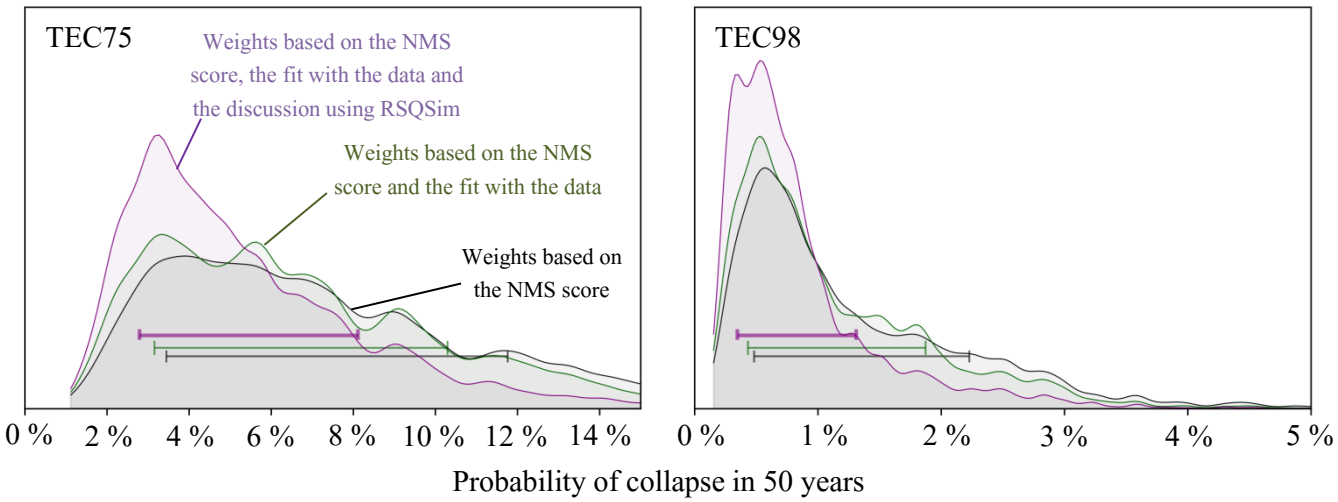

Figure 3.6: *Probability of collapse in 50 years for a mid-rise RC building in Istanbul built following the 1975 building code (left) or the 1998 building code (right). In black, probability of collapse using the logic tree weighted using the NMS score of the branch. In green, probability of collapse weighted according to the fit between the modelled earthquake annual rates and the annual rates calculated from the data. In purple, the probability of collapse weighted according to the fit between the data and the results of the physics-based modeling. (see chapter 2 for more details)*

8.1%. In total, the uncertainty is reduced by a factor 1.6.

Similarly, for a TEC98 building, the information gained from adding the comparison to the data in the weighting of the logic tree reduces the uncertainty on the probability of collapse from between 0.48% and 2.23% to between 0.44% and 1.88%. Adding the weighting suggested by the physics-based modelling, the uncertainty on the probability of collapse is brought down to between 0.35% and 1.31%. In total, the uncertainty is reduced by a factor 1.8.

3.4.2 Impact of uncertainties on the probability of collapse

In order to understand the impact of each uncertainty explored in the logic tree on the probability of collapse, we illustrate the results according to each branch of the logic tree (figures 3.7 to 3.11).

The probability of collapse is not dependant on the creep hypothesis used in the model (figure 3.7). The distance between Istanbul and the creeping section of the fault prohibits a strong influence of this distant source on the seismic hazard.

The choice of GMPE does affect the probability of collapse with the Akkar and Cagnan 2010 GMPE leading to significantly lower probabilities compared to the high values of Zhao et al. 2006 GMPE (figure 3.8).

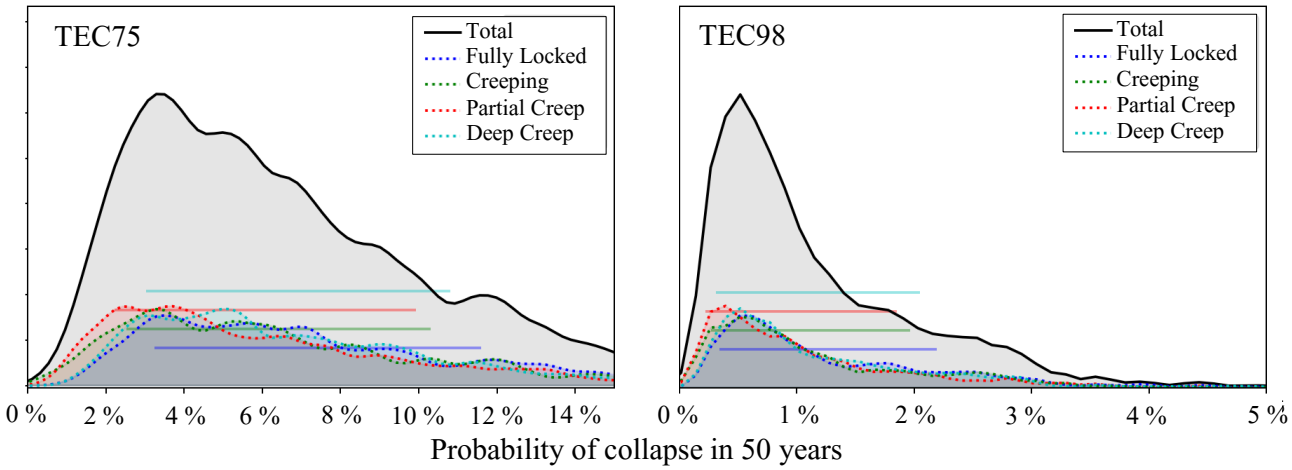

Figure 3.7: *Probability of collapse over 50 years colored according to the creep hypothesis.*

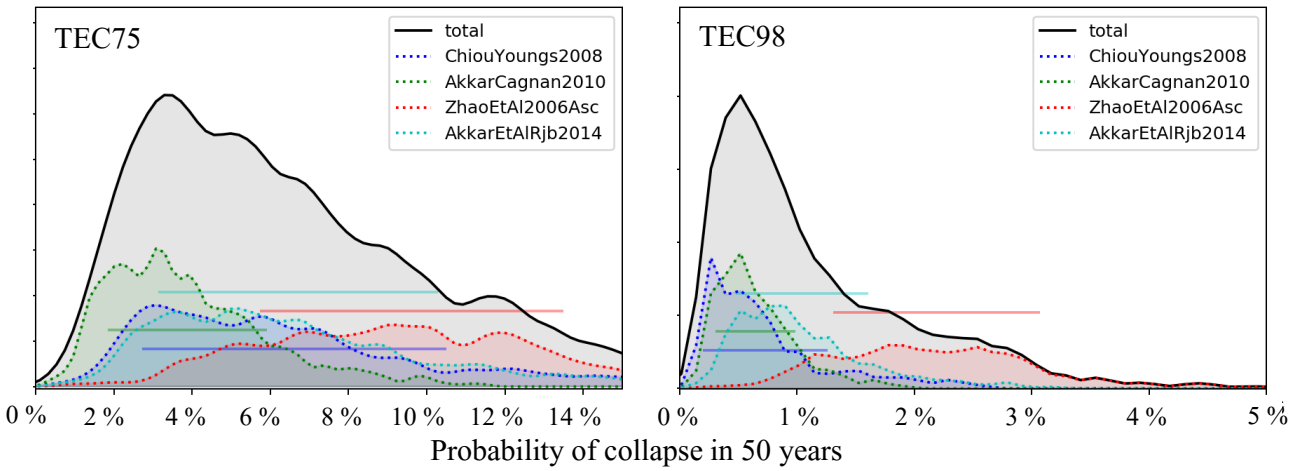

Figure 3.8: *Probability of collapse over 50 years colored according to the choice of GMPE.*

Using the GR hypothesis as a target shape leads to a significantly larger probability of collapse over 50 years than using the TS hypothesis (figure 3.9). The models using the GR hypothesis predicts a larger rate of earthquakes of magnitude 4.5 to 6.5 than the TS hypothesis, leading to a probability of collapse around two times larger.

Allowing larger ruptures to occur in the Set 2 branch consummates a large portion of the moment rate budget of the faults which leads to lower annual rate of rupture for earthquakes of all magnitudes. This change in the balance of the moment rate budget leads to probabilities of collapse smaller for Set 2 and for Set 1 by a factor of around 1.5 (figure 3.10).

In chapter 2, the results of the physics-based model RSQSim suggested giving less weight to both the GR hypothesis and the Set 2 hypothesis. The impact of this weighting on the probability of collapse can be observed in figure 3.6. However, in this particular case, the GR hypothesis leads to a higher risk and the Set1 to a lower risk. The

combined influence of these branches makes weighting less influential on the mean value of collapse but the overall uncertainty is reduced.

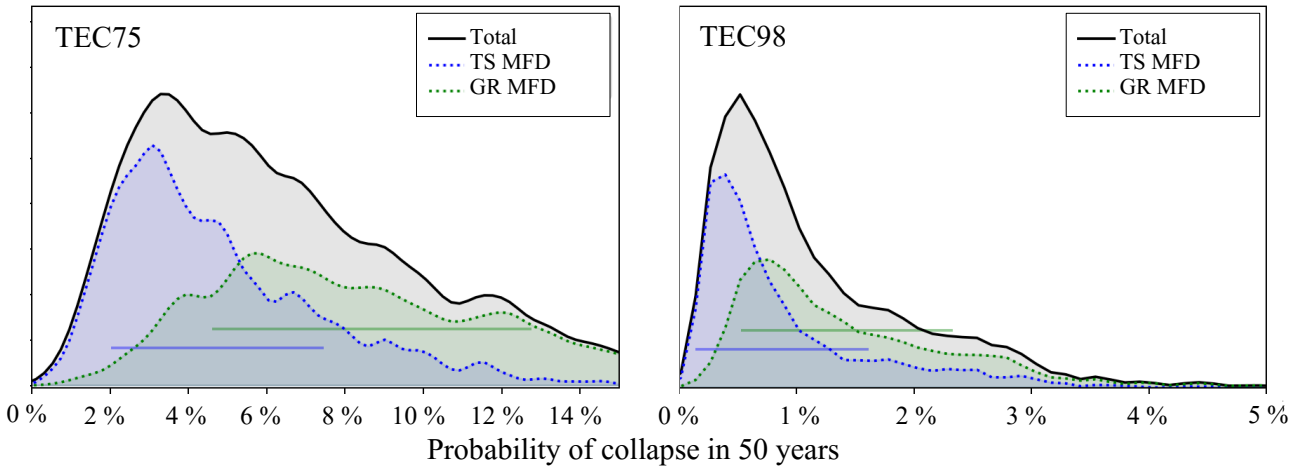

Figure 3.9: *Probability of collapse over 50 years colored according to the MFD hypothesis.*

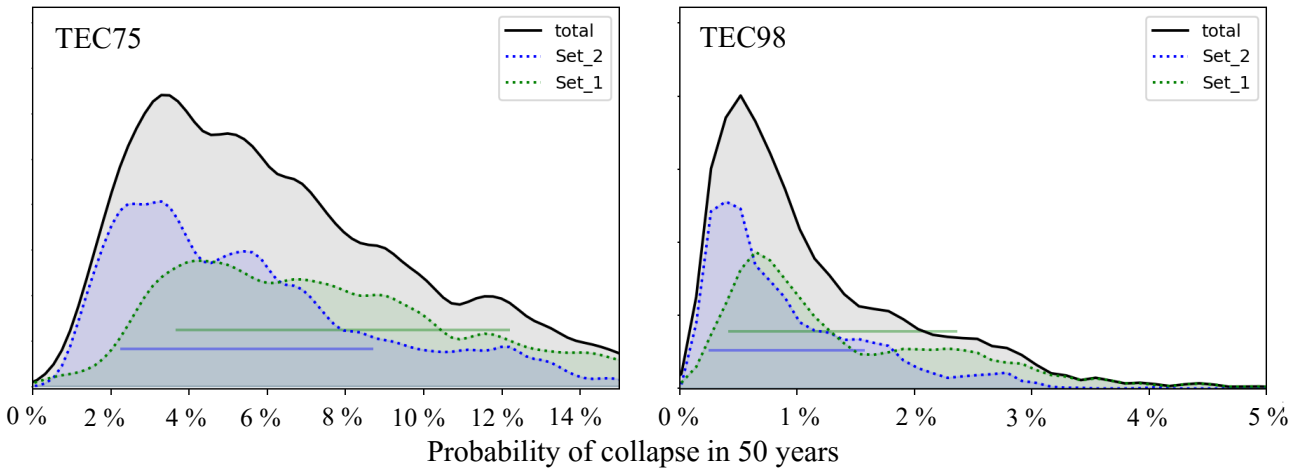

Figure 3.10: *Probability of collapse over 50 years colored according to set of scenarios.*

We evaluate (figure 3.12) the contribution of the uncertainties of each node of the logic tree by calculating the absolute difference between the distribution of the branches presented in figures 3.7 to 3.11. The most influential branches of the logic tree show the largest differences between the different distributions.For both building codes, the largest source of uncertainty is the choice of target MFD followed by the choice of the set of rupture scenario. The uncertainty with the less impact on the risk is the choice of locking model. For the TEC75 building, the background hypothesis node of the logic tree is generating a similar uncertainty on the risk to the choice of GMPE node. However, for the TEC98 building, while both nodes generate more uncertainty, the choice of GMPE is a larger source of uncertainty.

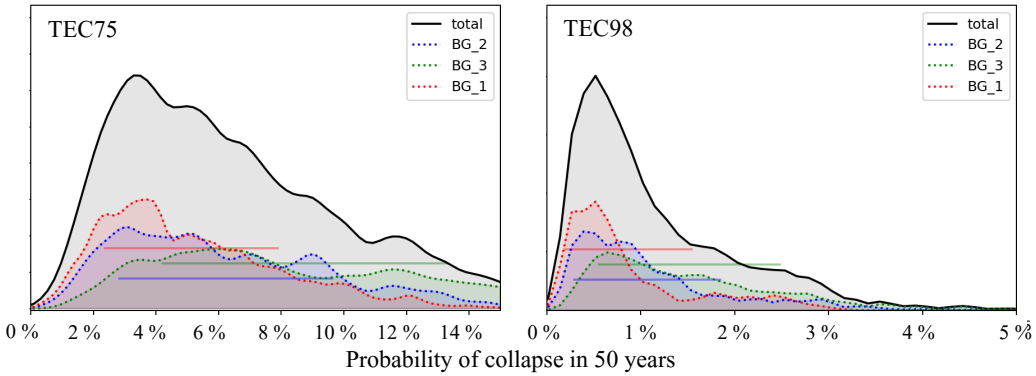

Figure 3.11: *Probability of collapse over 50 years colored according to the background hypothesis.*

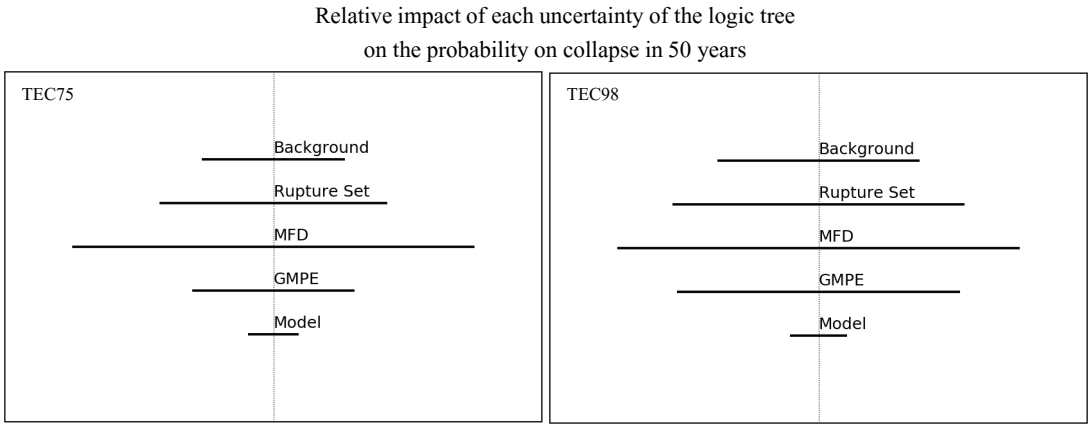

Figure 3.12: *Relative impact of each uncertainty on the probability of collapse for TEC75 (left) and TEC98 (right). The length of the line is relative to its impact on the uncertainty. The lengths are normalized. The uncertainties are from top to bottom : the background hypothesis, the rupture set, the MFD, the GMPE and the locking model of the NAF in the western Marmara Sea.*

3.4.3 Seismogenic sources controlling the risk

The deaggregation of the risk is ran for the 3 branches that have the strongest weight in the logic tree to analyze the impact on the risk. For comparison purposes, deaggregation is also computed using the faults and background model developed within the SHARE project. For each model, the two fragility curves TEC75 and TEC98 are used (figure 3.13).

For all models, the deaggregation of the seismic risk shows a clear bi-modal participation with a contribution of the distant large M7 and greater earthquakes occurring on the NAF and the participation of close small to moderate earthquakes occurring in the background. The choice of background hypothesis affects the participation of small to moderate earthquakes (figure 3.13 a and b). The choice of target MFD is the most impacting with a drastic increase of the participation of small earthquakes to the risk when the GR MFD is used compared to the TS MFD (figure 3.13 a

and c). While the probability of an M5 earthquake damaging a building is relatively small, the rate of M5 earthquakes in the background, e.g. close to the site, is larger in the GR hypothesis than in the TS one.

The SHARE model leads to a greater probability of collapse than the models from the logic tree proposed in this study (figure 3.13) due to higher estimates of both background seismicity rates and of M6.5 to 7.0 earthquakes on the faults compared to those proposed in this study (figure 3.14).

The contribution of each magnitude-distance bin to the probability of collapse depends strongly on the fragility curve considered for the building. Between figure 3.13 b and f, we can observe the decrease of the participation of M4.5 to 5.5 earthquakes when considering a TEC75 or a TEC98 building. As shown by the epsilon value close to 2 (figure 3.13 b), M4.5 to 5.5 earthquakes have a non negligible probability of generating PGA that are damaging for TEC75 building (around 0.4 g, cf figure 3.5). For a TEC98 building, a damaging PGA are around 0.8g; a PGA value unlikely to be generated by the M4.5 to 5.5 earthquakes (epsilon value are superior to 2.5 in figure 3.13 f).

In SHARE, 100% of earthquakes are in the background for magnitude lower or equal to 6.3 and 100% of the earthquakes are on the faults for magnitudes greater or equal to 6.4. This artificial separation is clearly visible in the deaggregation (figure 3.13). In the SHERIFS models, especially using the BG3 hypothesis, the rate of 4.5 $<$M$<$ 5.5 earthquakes on the faults is non-negligible. However, since the fault is almost 20 km away from the site of calculation, these moderate magnitudes don't participate to the risk of collapse in this case. The participation of the earthquake source to the risk of collapse is directly linked to the spatial distribution of the rate of earthquakes vis-a-vis the calculation site.

From the hazard source model, we extracted the magnitude frequency distribution (MFD) of the section of the NAF closest to Istanbul and the background MFD from a circular zone of 10 km radius around Istanbul (figure 3.1, figure 3.14). The 10 km radius was chosen to be representative of the distance up to which earthquakes are participating to the risk in the background region of Istanbul (figure 3.13). The larger rate of earthquakes assumed in the SHARE model (figure 3.14), both for the background seismicity and the faults, explains the stronger probability of collapse over 50 years. In SHERIFS, seismicity is shared between background and faults over the entire magnitudes range with the exception of the largest one (for the largest magnitude bin the ratio of earthquakes in the background is 0, for M$>$ than 7.0 in BG1 and M$>$7.5 in BG2 and BG3). For low magnitude earthquakes (4.5 $<$M$<$ 5.5) this ratio is 0.2 in BG1, 0.3 in BG2 and 0.6 in BG3 (teble 3.1).

For all models, the contribution of earthquakes for magnitude between 4.5 and 6.5 is linked to PGA in the upper part of the GMPE distribution (mean epsilon $>$ 2). Probability of collapse of the TEC98 building is linked to higher epsilon of the GMPE than for the TEC75 building. In order to lead to the collapse the TEC98 building, the PGAs are linked to the upper part of the GMPE distribution.

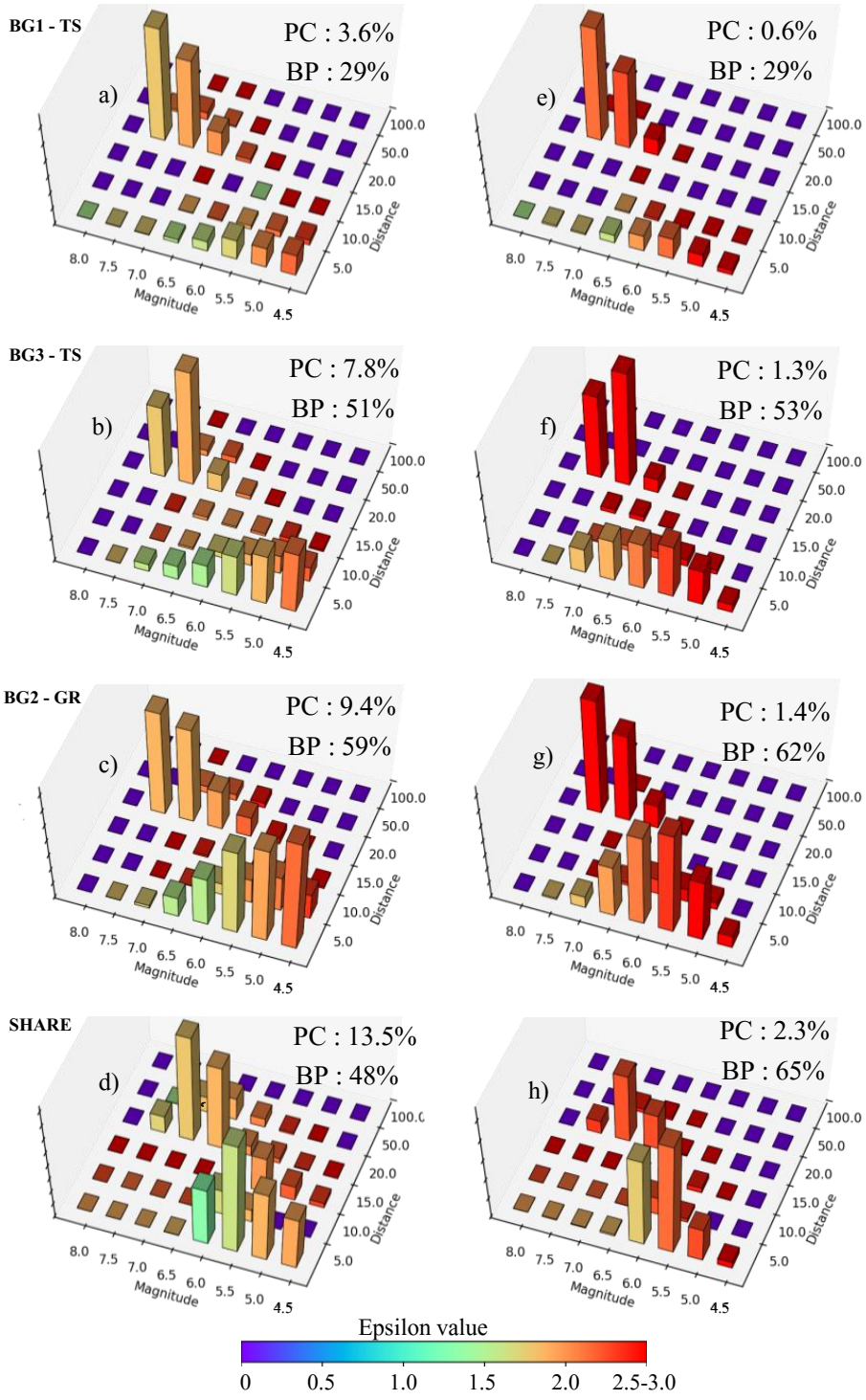

Figure 3.13: *Deaggregation of the probability of collapse over 50 years using the building code TEC75 (a,b,c,d) and TEC98 (e,f,g,h)and the GMPE Ak14. Three different models from the logic tree are used: all with partial creep model and the Set 1 of rupture scenario, the BG1 and TS MFD (a,e), BG3 and TS MFD (b,f) and BG2 and GR MFD (c,g). Figure d and h use the fault and background model from the SHARE project. The color represent the mean of the epsilon value of the GMPE for each bin. PC : Probability of Collapse. BP : Background Participation. The figures are normalized vertically since the vertical scale is the conditional probabilities.*

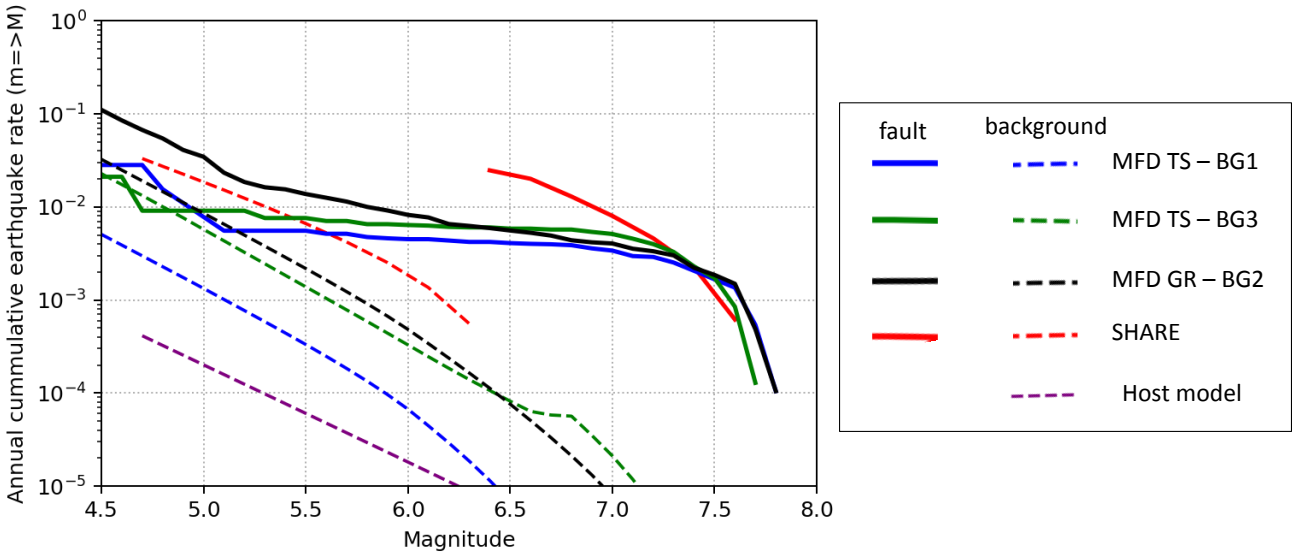

Figure 3.14: *For the four models used in figure 3.13 : participation rate of the fault section closest to the site (solid line) and for the background seismicity (dashed line) for a 10 km radius (figure 3.1) around the site. Three models from the logic tree are used: the BG1 and TS MFD (blue), BG3 and TS MFD (green) and BG2 and GR MFD (black). The fault and background MFD SHARE are in red. All models include the background zone north of Istanbul that crosses the 10km radius (figure 3.1) shown in purple.*

3.5 Discussion

3.5.1 Probability of collapse

The methodology used in the study allows to describe the differences in seismic risk estimates when considering two buildings following different construction codes in Istanbul (TEC75 and TEC98 type buildings). The TEC75 buildings have a probability of collapse in 50 years of 2.8 to 8.1% whereas the TEC98 buildings have a probability of collapse in 50 years of 0.4 to 1.3% (figure 3.6). The evolution of the building code between TEC75 and TEC98 leads to a 6 fold reduction of the risk of collapse over 50 years.

We used the data and the discussion based on the RSQSim modelling to weigh the logic tree (see chapter 2). Because of the weighting, the uncertainty on the risk is reduced by a factor of 1.6 for the TEC75 building and by a factor of 2.0 for the TEC98 building.

3.5.2 Linking input hypothesis and results

In this study we explore a larger range of epistemic uncertainties that is usually explored in the risk assessment studies and we observe the impact of these uncertainties by computing the probability of collapse. By using a streamlined methodology, we are able to expose the impact of each uncertainty on the final probability of collapse and therefore give an orientation for future research with a strong potential for improving the risk assessment.

The Magnitude Frequency Distribution

The largest source of uncertainty is on the choice of the target MFD (figure 3.12), especially for the TEC75 building. Future work on the earthquake catalog, both on the historical and the instrumental part, might be able to shed new light on the long term MFD in the western NAFS. However, since the number of large magnitude earthquakes will remain relatively low, the statistical analysis will still suffer from uncertainties in the large magnitude range.

After analyzing the MFD of several fault systems in New Zealand, Stirling and Gerstenberger [2018] advise to always explore this uncertainty in the logic tree for hazard studies. Since our results show the important impact of this uncertainty on the uncertainty concerning the risk, we also support the exploration of this uncertainty by default in future studies in order to have a proper estimation of the uncertainty on the risk.

However, reducing this uncertainty is possible. The long synthetic catalogs generated using "physics-based" model RSQSim show a clear deviation from the GR MFD. Taking into account this observation in the weighting of the logic tree significantly reduces the uncertainty on the probability of collapse (figure 3.6, figure 3.9). More studies with alternative "physics-based" approaches might be able to reduce even more the uncertainty.

Finally, a worldwide comparison of the MFD similar fault systems (e.g. San Andreas Fault System, Dead Sea Fault System, Hope Fault System in New Zealand...) could bring some new information on the validity of a MFD different from the GR MFD for a strike-slip fault system.

The largest possible rupture

The probability of collapse in 50 years is reduced in Set 2 compared to Set 1 (figure 3.10). Including larger earthquakes in a fixed moment rate budget system leads to reducing the overall rate of earthquakes. Since the ground motion generated by magnitude 7.5 earthquakes have the potential to already lead to the collapse of buildings (figure 3.13), the reduction of their annual rates has a greater impact on the probability of collapse than considerations of potential larger magnitude earthquakes, hence leading to an overall reduction of the probability of collapse for Set 2.

In chapter 2, the comparison between the model and the data (earthquake catalog and paleoseismicity) was not sufficient to discard the possibility of a magnitude 8 earthquake on the NAF close to Istanbul. It is possible that future paleoseismic studies reveal the occurrence of magnitude 8 earthquakes in the Marmara region. In order to reduce this uncertainty, we modelled the NAF with RSQSim. The absence of magnitude 8 earthquakes in the RSQSim synthetic catalogs allowed to set a stronger weight to the Set 1 branch and therefore reducing the uncertainty on the probability of collapse due to the rupture set branch. However, RSQSim doesn't capture the dynamic effects of the earthquake rupture which could allow for larger earthquakes than the ones suggested by RSQSim. Further physics-based modelling of the NAF have the potential to reduce even more the uncertainty caused by this branch of the logic tree .

The background seismicity

The uncertainty on the way the background seismicity is modelled is not usually explored in seismic risk analysis. Using the background modelling approach of SHERIFS, we explored three hypotheses of seismicity ratio between the faults and the background (table 3.1). This uncertainty has a large impact on the probability of collapse (figures 3.11 and 3.12); the larger the proportion of earthquakes in the background, the larger the probability of collapse.

Reducing this uncertainty will remain a challenge since the number of damaging earthquakes in a given area is too small for a conclusive statistical analysis. Geodetic strain maps and geologic investigations of deformation away from the major faults could provide pieces of information useful for setting the seismicity ratio between the faults and the background.

The creep condition model

In this study, the hypothesis of creep on the fault in the western Marmara Sea has the smallest impact on the uncertainty concerning the risk of collapse. The potentially creeping section is more than 50 km away from the risk calculation site in Istanbul. Because of this large distance, the risk is much more sensitive to the modelling parameters affecting the fault section directly in front of Istanbul.

3.6 Conclusion and Perspectives

Rather than exposing hazard metrics such as the 10% probability of exceedance of a PGA over 50 years, we chose to present risk metric such as the probability of collapse of a building in 50 years. Using a risk assessment approach we can draw the following conclusions :

- The probability of a 5 story RC building to require collapse prevention in Istanbul over a period of 50 years is 2.8

to 8.1 % for a TEC75 building and 0.4 to 1.3% for a TEC98 building.

- The evolution of the building code between TEC75 and TEC98 decreases the probability of collapse by a factor of 6 according to our study.
- By using the comparison with the data and the physics-based catalog from RSQSim, the uncertainty on the probability of collapse was reduced by a factor of 1.6 for the TEC75 building and by a factor 1.8 for the TEC98 building.
- The leading source of uncertainties is the choice of MFD followed by the choice of rupture set, the GMPE and the background hypothesis. The locking condition of the NAF doesn't affect the probability of collapse since the potentially creeping section is 70km away from Istanbul.
- Both magnitude 7+ earthquakes occurring on the NAF and moderate earthquakes of magnitude between 4.5 and 6.5 occurring within 10 km from Istanbul participate to the probability of collapse.

Identifying the sources controlling the hazard and the risk allows to target future studies that aim to reduce this uncertainty. For example, in Istanbul, the use of RSQSim has reduced uncertainties by allowing more weight on some branches. Alternative models could help to know if the GR MFD branch should be removed from the logic tree completely, thus further reducing the uncertainty.

Risk oriented studies have the potential to improve communication between two isolated communities that are the geosciences community and the engineering community by developing tools with which both hazard modeler and engineers can discuss. In this study, we only use PGA fragility curves. More fragility curves could be explored and multi dimensional (PGA, PSA, duration,..) fragility curves could also be used in this type of studies.

The approach presented in this study could also be used to draw maps of probability of collapse for different types of buildings. Such type of outputs could help public authorities to set up risk mitigation plans.

Chapter 4

Conclusions and Perspectives

This chapter presents the main conclusions and perspectives of this thesis. It is structured in three sections presenting each the conclusions and perspectives of an aspect of the thesis:

- Modelling earthquake rates in fault systems
- Physics-based modelling
- The risk oriented approach

A final section gives the general conclusions of the thesis.

4.1 Modelling earthquake rates in fault systems

4.1.1 The system level approach

In this study, we present a system level approach for earthquake rate modelling in fault systems, taking as input the slip-rate and geometry of the faults, an hypothesis on the MFD of the fault system, and an hypothesis on the ruptures that are possible in the system. Through an iterative procedure, the fault slip-rate budget is converted into annual rate of rupture. In order to accommodate the different input hypotheses, part of the slip-rate budget of a fault cannot be converted into annual rate of ruptures and is thus considered as aseismic. A proportion of the seismicity can be attributed in the background zone surrounding the fault system, representing the possibility of an earthquake happening on non-modelled faults. It is possible to explore a wide range of epistemic uncertainties in a logic tree. By comparing the modelled earthquake rates with the local data (earthquake catalog and paleoseismicity), a weighting of

the logic tree branches is proposed. Through this approach, the uncertainties affecting the modelling of the earthquake rates on faults are properly propagated in the seismic risk assessment.

4.1.2 Magnitude Frequency Distribution

In chapter 2, after analysis of the earthquake rates in the catalog, we observe a shape of the MFD that diverges from the GR distribution. We explored two hypotheses in a logic tree, the MFD following the GR distribution and the MFD following a shape tuned to that of the catalog MFD (TS MFD). Between the epistemic uncertainties explored in the logic tree in chapter 2, we found that the MFD is the leading cause of uncertainty on the probability of collapse in 50 years of a building in Istanbul. Following the conclusions of Stirling and Gerstenberger [2018], we suggest that this uncertainty should be explored for seismic hazard assessment.

The synthetic earthquake catalog generated by the physics based model RSQSim shows a clear divergence from the GR distribution towards a more characteristic distribution of the seismicity. Using this analysis, we were able to set a stronger weight to the TS MFD branch in the logic tree and therefore reduce the uncertainty on the assessment of the probability of collapse.

At the system level, the earthquake rates follow the MFD set as input. However, at the fault level, the participation rate (or rate of rupture going through the fault) can be very different from the regional MFD, sometimes tending towards a characteristic earthquake MFD even when a GR MFD is imposed at the system level. Similar observations can be made for the participation rates in the UCERF3model (Field et al. [2014]). As the intensity of the ground motion decreases with the square of the distance between the source and the site of calculation, the local seismogenic sources often control the hazard. Hence, the participation rate MFD can have a very strong impact on the hazard estimate. In Istanbul, the models constructed using SHERIFS lead to very different probability of collapse estimates than the models based on SHARE (see chapter 3). SHARE uses the classical fault level approach to model the faults and imposes a GR MFD on each fault, the risk disaggregation shows the impact of this hypothesis on the risk estimate. Allowing the local MFD to be different than the system level MFD is one of the features of the system level approach that has the largest impact on the risk estimate.

4.1.3 Complex fault rupture

Allowing ruptures to propagate between fault sections, moves the discussion from segmentation of the fault to possible ruptures in the system. This allows an easy exploration of the epistemic uncertainty on the maximum rupture that is possible in the fault system. Comparing the modelled earthquake rates with the rates from the earthquake catalog and the paleoseismicity studies can eliminate some rupture hypotheses and therefore reduce the epistemic uncertainty. This

allowed to test some rupture hypotheses laid out by paleoseismologists and compare how they perform against data.

4.1.4 The background seismicity

In SHERIFS, a novel methodology for modelling background seismicity is introduced. The proportion of seismicity on the faults and in the background is user-defined for each magnitude as an input. It acknowledges the intrinsic lack of knowledge of a given region and several hypothesis are easily explored. Due to a lack of modelling tools, this uncertainty is usually not explored in seismic hazard studies. In the Istanbul example, we showed that the background seismicity , depending on how it is considered in the fault model, can be an important source of uncertainty for the estimate of the probability of collapse. This uncertainty should be systematically explored for a complete picture of the epistemic uncertainty affecting a risk estimate.

4.1.5 Perspectives on the use of SHERIFS

While SHERIFS is operational and has been applied to a variety of fault systems, it has never been applied to a subduction zone. In theory, the methodology should be applicable in the same way; however, unforeseen challenges may arise. The application of this methodology to a subduction would allow to explore different approaches and developments to model the coupling of the interface.

Many aspects of SHERIFS could be improved, especially in term of computational efficiency. Faster computation enables the user to explore a wider range of uncertainties on a laptop computer. Features like automatic consistency checks could be implemented. For example if a rupture set is not compatible at all with the other hypotheses in the system, SHERIFS could notify the user and prevent him from interpreting a flawed set of input hypotheses.

Several additional refinements could be added to SHERIFS, one of which being the modelling of complex slip-distribution along the ruptures. At present SHERIFS spends the same amount of slip-rate budget for all the faults involved in a given rupture, the possibility of using other distributions can be implemented.

So far, SHERIFS calculates only time-independent rate of earthquakes in the fault system which is the first step towards building a time-dependent model. In time-dependent models, depending on the elapsed time since the last earthquake, the probability of an earthquake in the near future is modified by a probability function. Several functions are possible and UCERF3-TD (time dependent, Field et al. [2015]) presents how this approach can be applied to system level approaches. The way to model time dependency is still under discussion in the scientific community and, to my knowledge, no national building code relies on such models. It would be interesting to transform SHERIFS in a tool allowing users to join this discussion and to explore the impact of different hypotheses on the risk assessment.

Finally, SHERIFS is a community tools used by geologists and hazard modelers. By sharing the risk calculation codes, we could bring engineers in this community. Working as a community to address the problems of seismic risk is the most efficient way to bring the best available science into the decisions on risk mitigation measures. We can follow the example of the SCEC community (Southern California Earthquake Center) that has enabled advances in the field of seismic hazard assessment (UCERF - Field et al. [2014], CyberSHAKE - Graves et al. [2011], ...).

4.2 Physics-based models and alternative approaches of earthquake modelling

4.2.1 Possible ruptures

In accordance with dynamical modelling results of Aochi and Ulrich [2015], we define earthquake ruptures that can run through the fault bend of the NAF in the vicinity of Istanbul.

On the NAF, no earthquakes larger than 7.7 have been observed in the historical time. However, since the length of the historical catalog is limited to several hundred years, it is possible that larger earthquakes with long recurrence times may occur along the western NAF. Therefore we explore this hypothesis in a logic tree branch that allows earthquake ruptures corresponding to a magnitude 8 earthquakes.

The 10 000 years long synthetic earthquake catalogs generated using the earthquake simulator RSQSim didn't contain any earthquake of magnitude greater than 7.7. This observation suggests a stronger weight for the branch of the logic tree that doesn't include the larger rupture. Since the set of possible ruptures has an impact on the probability of collapse, the weighting of the logic tree allows to reduce the uncertainty on the risk.

4.2.2 Earthquake rates

In the Marmara region, the MFD calculated from the earthquake catalog shows a divergence from the GR MFD. However, it can be argued that this observation is due to an incompleteness of the earthquake catalog or the uncertainty of the historical magnitudes. Nevertheless, we set up logic tree branches exploring this hypotheses: a GR MFD and a MFD tuned to the shape of the catalog (Tuned Shape TS).

The MFD of the 10 000 years long catalog simulated using RSQSim follows a MFD significantly different from a GR MFD. Therefore, we set a stronger weight to the branch using the TS MFD. The choice of the MFD being the leading source of uncertainty on the risk, the setting of the weight led to an important reduction of the uncertainty affecting the assessment of the probability of collapse in 50 years of a building in Istanbul.

However, it is worth noting that RSQSim simplifies the dynamics of the earthquake rupture and it is not excluded that other physics-based models could predict larger ruptures or a different MFD shape. At the time of writing this thesis, there are no simulators able to generate long earthquake histories of an extensive 3D fault network while integrating the full earthquake dynamics.

4.2.3 Perspectives on the physics based modelling

In this study, we have demonstrated how physics based modelling can be used to improve seismic risk estimates by using RSQSim as a discussion tool. In order not to over-interpret the results of RSQSim, given the modelling limitation, we limited the discussion to two first order questions concerning the NAF: are magnitude 8 earthquakes possible and can the MFD differ from a GR distribution?

As physics-based models continue to improve and evolve, the answers to these questions can be more accurate in the future and more detailed questions can be asked. For example, the local participation rate of a fault section is extremely important for assessing the seismic risk of neighbouring sites. Participation rates from statistical methods such as SHERIFS could be compared to those produced from physics based models.

In this study, we didn't focus on the ground motion prediction aspect of the seismic hazard assessment and used GMPE available in the literature. GMPE are empirically constructed and their accuracy to represent the ground motion can only be as good as the data used to derive them. As very few ground motion have so far been recorded near faults, GMPEs predictive power is reduced for sites located less than 10 km away from the seismogenic source, typically the sites exposed to the highest risk levels. Dynamic rupture models have the potential to provide the synthetic data needed to complement the ground motion database on which GMPEs are constructed.

Finally, one can imagine that fully physics-based hazard models will come to replace the traditional statically and empirically based hazard models of the Cornell approach, such as the ones constructed in this study. By linking earthquake rupture simulation and wave propagation for a long synthetic catalog, it is possible to calculate the probability of exceeding a given damaging ground motion (Shaw et al. [2018]), even extract some non-Poissonian behaviours of the system. Such model could have the potential to generate an earthquake hazard forecast as efficiently as meteorological model can forecast the weather. Efforts for the integration of physics-based geosciences in hazard assessment are carried on within SCEC (Southern Californian Earthquake Center) (Shaw et al. [2018]) and international projects such as ChEESE (`https://cheese-coe.eu`).

While the physics of the earthquake rupture and the wave propagation can be properly written into equations and, in theory, integrated in a 3D calculation software; one major roadblock stands in the way of a fully physics based hazard assessment: the large epistemic uncertainties affecting geophysical systems. Physics based models require an

important level of detail for the geometry and properties both of the fault and of the medium where the wave is propagating. This is required in order to be able to simulate the large bandwidth of frequency content of a ground motion that can damage a building. Meteorological models solve similar issues by comparing the models to the data and gradually improving. However, the weather is a constantly occurring phenomenon for which most parameters can be assessed by in-situ sampling or through light-waves observation. Earthquakes and solid earth parameters are for the most part only accessible when earthquakes occur and rarely with direct in-situ measurements of the parameters. Therefore, since most of the parameters will likely never be known with a sufficient detail for all points of a model, exploring the epistemic uncertainty is necessary but is computationally very impractical with these detailed models.

Nonetheless, pursuing the development of fully physics based hazard model will allow improving our general knowledge of the earthquake phenomena and, equally importantly, will generate interactions between different sectors of the geosciences community by creating a common goal as well as shared methodologies and modelling tools that can be used to inform traditional models. In my opinion, these beneficial outcomes have the potential to drastically improve the seismic hazard assessment even if the practical applicability of a fully physics based hazard model seems impossible in the foreseeable future.

In the annex, I discuss the use of Machine Learning methods for the seismic risk assessment.

4.3 The risk oriented approach

In this study, we have deliberately chosen to present the result in term of risk (e.g. probability of collapse) and not in the more classical terms of hazard (e.g. expected acceleration associated to a given probability of being exceeded). The reasons for this choice are twofold.

First, the usual hazard metrics such as the UHS (Uniform Hazard Spectrum) or the hazard map are presented for a given probability of exceedance, for example 10% over 50 years. While 50 years is a comprehensive universal number representing the average time of occupancy of a building by an individual, the 10% is a political choice for an acceptable level of risk taken which can change drastically depending on the application of the study. In this study, we don't have a given application in mind but simply want to expose and understand the seismic risk that affects a building in Istanbul, therefore the probability of collapse is a more adapted and comprehensive metric.

Second, using a risk oriented approach is an opportunity to open the discussion of this study, about the seismogenic sources controlling the risk and its uncertainty, to a wider audience than the hazard modelers community, bringing together geosciences and engineering. In order to reach this aim we developed tools easily used by both geoscientists and hazard modelers to identify the sources contributing to the seismic risk and the origins of the

uncertainties. At the same time, the same tools allow engineers to appreciate the reduction of risk when more precise geological/seismological hypotheses are available for the faults.

4.3.1 Identifying the sources of the uncertainties

Concerning the probability of collapse for a mid-rise RC building in Istanbul, we identified the MFD of the fault system as the leading source of uncertainty. We were able to reduce this uncertainty using the discussion with RSQSim. The other uncertainties explored in the logic tree (the rupture scenario set, the background seismicity, the creep model of the fault, and the GMPE) all have an impact. The choice of creep model hypothesis has a minimal impact on the risk estimate due to the 70km distance between the potentially creeping section of the fault and the site of calculation.

By weighting the logic tree using the comparison between the modelled earthquake rates and the data as well as the discussion based on the RSQSim synthetic catalog, we were able to enhance the accuracy of the risk assessment by a factor 1.6 for the older 1975 building and by a factor of 2 for the 1998 building. According to our calculations, the evolution of the building code between TEC75 and TEC98 leads to a 6 fold reduction of the risk of collapse over 50 years.

4.3.2 Risk disaggregation

We developed a methodology for risk disaggregation that exposes the contributing seismogenic sources of the probability of collapse for a building during a period of time, here 50 years.

In Istanbul, the probability of collapse is controlled by two sources, the magnitude 7+ earthquakes occurring on the nearest section of the NAF 17km away and the magnitude 4.5 to 6.5 earthquakes occurring in the background seismicity occurring within 10 km from the site. The relative participation to the risk of the background compared to the fault is variable depending on the background hypothesis used, ranging from 30 to 60% of the participation.

The TEC75 building is more sensitive to the background hypothesis than the TEC98 building. The disaggregation also shows that the ground motion that can lead to the collapse of the TEC98 building are more extreme compared to the mean ground motion of the GMPE than the ground motion that lead to the collapse of the TEC75 building.

4.3.3 Perspectives on the risk calculation

The methodology set up in this study is implemented in a code that takes the outputs of SHERIFS and OpenQuake to calculate the risk of building damage and the disaggregation. The work needed to make this code user friendly and available will be carried in the future. The experience of sharing SHERIFS showed important improvements in the discussion of modelling earthquake rates on faults. Sharing the risk calculation has a potential to enhance communication between the seismic hazard community and the engineering community.

So far, the methodology only considered as input the fragility curve provided by the engineer. With more exchanges with the engineers, it could be possible to imagine other types of inputs such as multi-dimensional fragility curves that can take several ground motion parameters as input instead of one.

The methodology can also be used to generate other types of outputs such as a map of the probability of collapse for a given type of building or a map of the uncertainties, a useful tool to orient risk mitigation measures. This could identify areas where the risk is potentially high but very uncertain and help setting up future studies in order to reduce this uncertainty.

Bibliography

Akkar, S. and Çağnan, Z. (2010). A local ground-motion predictive model for turkey, and its comparison with other regional and global ground-motion models. *Bulletin of the Seismological Society of America*, 100(6):2978–2995.

Akkar, S., Sandıkkaya, M., and Bommer, J. (2014). Empirical ground-motion models for point-and extended-source crustal earthquake scenarios in europe and the middle east. *Bulletin of earthquake engineering*, 12(1):359–387.

Albini, P., Rovida, A., Scotti, O., and Lyon-Caen, H. (2017). Large eighteenth–nineteenth century earthquakes in western gulf of corinth with reappraised size and location. *Bulletin of the Seismological Society of America*, 107(4):1663–1687.

Ambraseys, N. (2002). The seismic activity of the marmara sea region over the last 2000 years. *Bulletin of the Seismological Society of America*, 92(1):1–18.

Aochi, H. and Ulrich, T. (2015). A probable earthquake scenario near istanbul determined from dynamic simulations. *Bulletin of the Seismological Society of America*, 105(3):1468–1475.

Armijo, R., Meyer, B., Hubert, A., and Barka, A. (1999). Westward propagation of the north anatolian fault into the northern aegean: Timing and kinematics. *Geology*, 27(3):267–270.

Armijo, R., Meyer, B., King, G., Rigo, A., and Papanastassiou, D. (1996). Quaternary evolution of the corinth rift and its implications for the late cenozoic evolution of the aegean. *Geophysical Journal International*, 126(1):11–53.

Avallone, A., Briole, P., Agatza-Balodimou, A. M., Billiris, H., Charade, O., Mitsakaki, C., Nercessian, A., Papazissi, K., Paradissis, D., and Veis, G. (2004). Analysis of eleven years of deformation measured by gps in the corinth rift laboratory area. *Comptes Rendus Geoscience*, 336(4-5):301–311.

Baker, C., Hatzfeld, D. L., Lyon-Caen, H., Papadimitriou, E., and Rigo, A. (1997). Earthquake mechanisms of the adriatic sea and western greece: implications for the oceanic subduction-continental collision transition. *Geophysical Journal International*, 131(3):559–594.

Bal, İ. E., Crowley, H., Pinho, R., and Gülay, F. G. (2008). Detailed assessment of structural characteristics of turkish rc building stock for loss assessment models. *Soil Dynamics and Earthquake Engineering*, 28(10-11):914–932.

Bankston III, C. L., Barnshaw, J., Bevc, C., Capowich, G. E., Clarke, L., Das, S. K., Donato, K. M., Dynes, R. R., Eargle, L. A., Elliott, J. R., et al. (2010). *The sociology of Katrina: Perspectives on a modern catastrophe*. Rowman & Littlefield Publishers.

Barka, A. and Kadinsky-Cade, K. (1988). Strike-slip fault geometry in turkey and its influence on earthquake activity. *Tectonics*, 7(3):663–684.

Bernard, P., Briole, P., Meyer, B., Lyon-Caen, H., Gomez, J.-M., Tiberi, C., Berge, C., Cattin, R., Hatzfeld, D., Lachet, C., et al. (1997). The m s= 6.2, june 15, 1995 aigion earthquake (greece): evidence for low angle normal faulting in the corinth rift. *Journal of Seismology*, 1(2):131–150.

Bernard, P., Lyon-Caen, H., Briole, P., Deschamps, A., Boudin, F., Makropoulos, K., Papadimitriou, P., Lemeille, F., Patau, G., Billiris, H., et al. (2006). Seismicity, deformation and seismic hazard in the western rift of corinth: New insights from the corinth rift laboratory (crl). *Tectonophysics*, 426(1-2):7–30.

Bohnhoff, M., Martínez-Garzón, P., Bulut, F., Stierle, E., and Ben-Zion, Y. (2016). Maximum earthquake magnitudes along different sections of the north anatolian fault zone. *Tectonophysics*, 674:147–165.

Bohnhoff, M., Wollin, C., Domigall, D., Küperkoch, L., Martínez-Garzón, P., Kwiatek, G., Dresen, G., and Malin, P. E. (2017). Repeating marmara sea earthquakes: indication for fault creep. *Geophysical Journal International*, 210(1):332–339.

Boiselet, A. (2014). *Cycle sismique et aléa sismique d'un réseau de failles actives: le cas du rift de Corinthe (Grèce)*. PhD thesis, Ecole Normale Superieure de Paris-ENS Paris.

Bommer, J., Spence, R., Erdik, M., Tabuchi, S., Aydinoglu, N., Booth, E., del Re, D., and Peterken, O. (2002). Development of an earthquake loss model for turkish catastrophe insurance. *Journal of Seismology*, 6(3):431–446.

Brandenberg, S. J., Wang, P., Nweke, C. C., Hudson, K., and al. (2019). Preliminary report on engineering and geological effects of the july 2019 ridgecrest earthquake sequence. *USGS*.

Chartier, T., Scotti, O., Clément, C., Jomard, H., and Baize, S. (2017a). Transposing an active fault database into a fault-based seismic hazard assessment for nuclear facilities–part 2: Impact of fault parameter uncertainties on a site-specific psha exercise in the upper rhine graben, eastern france. *Natural Hazards and Earth System Sciences*, 17(9):1585–1593.

Chartier, T., Scotti, O., and Lyon-Caen, H. (2019). Sherifs: Open-source code for computing earthquake rates in fault systems and constructing hazard models. *Seismological Research Letters*.

Chartier, T., Scotti, O., Lyon-Caen, H., and Boiselet, A. (2017b). Methodology for earthquake rupture rate estimates of fault networks: example for the western corinth rift, greece. *Natural Hazards and Earth System Sciences*, 17(10):1857–1869.

Cheloni, D., Giuliani, R., D'Anastasio, E., Atzori, S., Walters, R., Bonci, L., D'Agostino, N., Mattone, M., Calcaterra, S., Gambino, P., et al. (2014). Coseismic and post-seismic slip of the 2009 l'aquila (central italy) mw 6.3 earthquake and implications for seismic potential along the campotosto fault from joint inversion of high-precision levelling, insar and gps data. *Tectonophysics*, 622:168–185.

Chiou, B. S.-J. and Youngs, R. R. (2014). Update of the chiou and youngs nga model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, 30(3):1117–1153.

Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the seismological society of America*, 58(5):1583–1606.

Danciu, L., Şeşetyan, K., Demircioglu, M., Gülen, L., Zare, M., Basili, R., Elias, A., Adamia, S., Tsereteli, N., Yalçın, H., et al. (2018). The 2014 earthquake model of the middle east: seismogenic sources. *Bulletin of Earthquake Engineering*, 16(8):3465–3496.

Demircioğlu, M. B., Şeşetyan, K., Duman, T. Y., Can, T., Tekin, S., and Ergintav, S. (2018). A probabilistic seismic hazard assessment for the turkish territory: part ii—fault source and background seismicity model. *Bulletin of Earthquake Engineering*, 16(8):3399–3438.

Diao, F., Walter, T. R., Solaro, G., Wang, R., Bonano, M., Manzo, M., Ergintav, S., Zheng, Y., Xiong, X., and Lanari, R. (2016). Fault locking near istanbul: Indication of earthquake potential from insar and gps observations. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, 205(1):490–498.

Dieterich, J. H. (1978). Time-dependent friction and the mechanics of stick-slip. In *Rock Friction and Earthquake Prediction*, pages 790–806. Springer.

Dikbaş, A. and Akyüz, H. S. (2011). Kaf zonu üzerinde izmit-sapanca gölü segmentinin fay morfolojisi ve paleosismolojisi. *İTÜDERGİSİ/d*, 9(3).

Dikbaş, A., Akyüz, H. S., Meghraoui, M., Ferry, M., Altunel, E., Zabcı, C., Langridge, R., and Yalçıner, C. Ç. (2018). Paleoseismic history and slip rate along the sapanca-akyazı segment of the 1999 izmit earthquake rupture (mw= 7.4) of the north anatolian fault (turkey). *Tectonophysics*, 738:92–111.

Duman, T. Y., Çan, T., Emre, Ö., Kadirioğlu, F. T., Baştürk, N. B., Kılıç, T., Arslan, S., Özalp, S., Kartal, R. F., Kalafat, D., et al. (2018). Seismotectonic database of turkey. *Bulletin of Earthquake Engineering*, 16(8):3277–3316.

Durand, V., Hok, S., Boiselet, A., Bernard, P., and Scotti, O. (2016). Dynamic rupture simulations on a fault network in the corinth rift. *Geophysical Journal International*, 208(3):1611–1622.

Duverger, C., Godano, M., Bernard, P., Lyon-Caen, H., and Lambotte, S. (2015). The 2003–2004 seismic swarm in the western corinth rift: Evidence for a multiscale pore pressure diffusion process along a permeable fault system. *Geophysical Research Letters*, 42(18):7374–7382.

Emre, Ö., Duman, T. Y., Özalp, S., Şaroğlu, F., Olgun, Ş., Elmacı, H., and Can, T. (2018). Active fault database of turkey. *Bulletin of Earthquake Engineering*, 16(8):3229–3275.

Erdik, M., Biro, Y. A., Onur, T., Sesetyan, K., and Birgoren, G. (1999). Assessment of earthquake hazard in turkey and neighboring. *Annals of Geophysics*, 42(6).

Erdik, M., Demircioglu, M., Sesetyan, K., Durukal, E., and Siyahi, B. (2004). Earthquake hazard in marmara region, turkey. *Soil Dynamics and Earthquake Engineering*, 24(8):605–631.

Ergintav, S., Reilinger, R., Çakmak, R., Floyd, M., Cakir, Z., Doğan, U., King, R., McClusky, S., and Özener, H. (2014). Istanbul's earthquake hot spots: Geodetic constraints on strain accumulation along faults in the marmara seismic gap. *Geophysical Research Letters*, 41(16):5783–5788.

Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., et al. (2014). Uniform california earthquake rupture forecast, version 3 (ucerf3)—the time-independent model. *Bulletin of the Seismological Society of America*, 104(3):1122–1180.

Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., et al. (2015). Long-term time-dependent probabilities for the third uniform california earthquake rupture forecast (ucerf3). *Bulletin of the Seismological Society of America*, 105(2A):511–543.

Field, E. H., Jordan, T. H., and Cornell, C. A. (2003). Opensha: A developing community-modeling environment for seismic hazard analysis. *Seismological Research Letters*, 74(4):406–419.

Flerit, F., Armijo, R., King, G., and Meyer, B. (2004). The mechanical interaction between the propagating north anatolian fault and the back-arc extension in the aegean. *Earth and Planetary Science Letters*, 224(3-4):347–362.

Ford, M., Rohais, S., Williams, E. A., Bourlange, S., Jousselin, D., Backert, N., and Malartre, F. (2013). Tectono-sedimentary evolution of the western corinth rift (central greece). *Basin Research*, 25(1):3–25.

Geist, E. L. and Parsons, T. (2018). Determining on-fault earthquake magnitude distributions from integer programming. *Computers & geosciences*, 111:244–259.

Giannopoulos, D., Rivet, D., Sokos, E., Deschamps, A., Mordret, A., Lyon-Caen, H., Bernard, P., Paraskevopoulos, P., and Tselentis, G.-A. (2017). Ambient noise tomography of the western corinth rift, greece. *Geophysical Journal International*, 211(1):284–299.

Giardini, D., Danciu, L., Erdik, M., Şeşetyan, K., Tümsa, M. B. D., Akkar, S., Gülen, L., and Zare, M. (2018). Seismic hazard map of the middle east. *Bulletin of earthquake engineering*, 16(8):3567–3570.

Graves, R., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., Milner, K., et al. (2011). Cybershake: A physics-based seismic hazard model for southern california. *Pure and Applied Geophysics*, 168(3-4):367–381.

Grünthal, G., Wahlström, R., and Stromeyer, D. (2013). The share european earthquake catalogue (sheec) for the time period 1900–2006 and its comparison to the european-mediterranean earthquake catalogue (emec). *Journal of seismology*, 17(4):1339–1344.

Gülerce, Z., Buğra Soyman, K., Güner, B., and Kaymakci, N. (2017). Planar seismic source characterization models developed for probabilistic seismic hazard assessment of istanbul. *Natural Hazards and Earth System Sciences*, 17(12):2365–2381.

Gülerce, Z. and Ocak, S. (2013). Probabilistic seismic hazard assessment of eastern marmara region. *Bulletin of Earthquake Engineering*, 11(5):1259–1277.

Gutenberg, B. and Richter, C. (1954). Seismicity of the earth, 310 pp.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in california. *Bulletin of the Seismological society of America*, 34(4):185–188.

Gómez-Novell, O., Chartier, T. H., García-Mayordomo, J., Ortuño, M., Masana, E., Insua-Arévalo, J., López, R., and Scotti, O. (2018). Preliminary characterization of the magnitude-frenquency distributions of multi-fault rupture hypotheses at the eastern betic shear zone (se spain) by using the sherifs approach. *Iberfault, 3ª Reunión Ibérica sobre Fallas Activas y Paleosismología, Alicante (España), 11-13 June 2018.*

Hanks, T. C. and Kanamori, H. (1979). A moment magnitude scale. *Journal of Geophysical Research: Solid Earth*, 84(B5):2348–2350.

Hecker, S., Abrahamson, N., and Wooddell, K. (2013). Variability of displacement at a point: Implications for earthquake-size distribution and rupture hazard on faults. *Bulletin of the Seismological Society of America*, 103(2A):651–674.

Hergert, T. and Heidbach, O. (2010). Slip-rate variability and distributed deformation in the marmara sea fault system. *Nature Geoscience*, 3(2):132.

Hong, H. and Goda, K. (2006). A comparison of seismic-hazard and risk deaggregation. *Bulletin of the Seismological Society of America*, 96(6):2021–2039.

Hornblow, S., Quigley, M., Nicol, A., Van Dissen, R., and Wang, N. (2014). Paleoseismology of the 2010 mw 7.1 darfield (canterbury) earthquake source, greendale fault, new zealand. *Tectonophysics*, 637:178–190.

Hsu, L. and Bürgmann, R. (2006). Surface creep along the longitudinal valley fault, taiwan from insar measurements. *Geophysical research letters*, 33(6).

Iervolino, I., Spillatura, A., and Bazzurro, P. (2018). Seismic reliability of code-conforming italian buildings. *Journal of Earthquake Engineering*, 22(sup2):5–27.

Jackson, J., Gagnepain, J., Houseman, G., King, G., Papadimitriou, P., Soufleris, C., and Virieux, J. (1982). Seismicity, normal faulting, and the geomorphological development of the gulf of corinth (greece): the corinth earthquakes of february and march 1981. *Earth and Planetary Science Letters*, 57(2):377–397.

Jacottin, M. (2019). Apport du gps pour contraindre la sismicité du massif armoricain. *M2, Université de Montpellier/IRSN*.

Jayaram, N. and Baker, J. W. (2009). Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering & Structural Dynamics*, 38(15):1687–1708.

Kadirioğlu, F. T., Kartal, R. F., Kılıç, T., Kalafat, D., Duman, T. Y., Azak, T. E., Özalp, S., and Emre, Ö. (2018). An improved earthquake catalogue (m >= 4.0) for turkey and its near vicinity (1900–2012). *Bulletin of Earthquake Engineering*, 16(8):3317–3338.

Kim, Y., Kim, M., and Kim, W. (2013). Effect of the fukushima nuclear disaster on global public acceptance of nuclear energy. *Energy Policy*, 61:822–828.

Klein, E., Duputel, Z., Masson, F., Yavasoglu, H., and Agram, P. (2017). Aseismic slip and seismogenic coupling in the marmara sea: What can we learn from onland geodesy? *Geophysical Research Letters*, 44(7):3100–3108.

Klinger, Y., Okubo, K., Vallage, A., Champenois, J., Delorme, A., Rougier, E., Lei, Z., Knight, E. E., Munjiza, A., Satriano, C., et al. (2018). Earthquake damage patterns resolve complex rupture processes. *Geophysical Research Letters*, 45(19):10–279.

Klinger, Y., Sieh, K., Altunel, E., Akoglu, A., Barka, A., Dawson, T., Gonzalez, T., Meltzner, A., and Rockwell, T. (2003). Paleoseismic evidence of characteristic slip on the western segment of the north anatolian fault, turkey. *Bulletin of the Seismological Society of America*, 93(6):2317–2332.

Knuepfer, P. (1989). Implications of the characteristics of end-points of historical surface fault ruptures for the nature of fault segmentation. In *Fault Segmentation and Controls of Rupture Initiation and Termination*, volume 89, pages 193–228. US Geol. Surv. Open-File Rept.

Lambotte, S., Lyon-Caen, H., Bernard, P., Deschamps, A., Patau, G., Nercessian, A., Pacchiani, F., Bourouis, S., Drilleau, M., and Adamova, P. (2014). Reassessment of the rifting process in the western corinth rift from relocated seismicity. *Geophysical Journal International*, 197(3):1822–1844.

Lange, D., Kopp, H., Royer, J.-Y., Henry, P., Cakir, Z., Petersen, F., Sakic, P., Ballu, V., Bialas, J., Özeren, M. S., et al. (2019). Interseismic strain build-up on the submarine north anatolian fault offshore istanbul. *Nature communications*, 10(1):3006.

Le Pichon, X., Chamot-Rooke, N., Rangin, C., and Sengör, A. (2003). The north anatolian fault in the sea of marmara. *Journal of Geophysical Research: Solid Earth*, 108(B4).

Le Pichon, X., Şengör, A., Demirbağ, E., Rangin, C., Imren, C., Armijo, R., Görür, N., Çağatay, N., De Lepinay, B. M., Meyer, B., et al. (2001). The active main marmara fault. *Earth and Planetary Science Letters*, 192(4):595–616.

Lefevre, M., Klinger, Y., Al-Qaryouti, M., Le Béon, M., and Moumani, K. (2018). Slip deficit and temporal clustering along the dead sea fault from paleoseismological investigations. *Scientific reports*, 8(1):4511.

Leonard, M. (2010). Earthquake fault scaling: Self-consistent relating of rupture length, width, average displacement, and moment release. *Bulletin of the Seismological Society of America*, 100(5A):1971–1988.

Lienkaemper, J. J., DeLong, S. B., Domrose, C. J., and Rosa, C. M. (2016). Afterslip behavior following the 2014 m 6.0 south napa earthquake with implications for afterslip forecasting on other seismogenic faults. *Seismological Research Letters*, 87(3):609–619.

Lyon-Caen, H., Papadimitriou, P., Deschamps, A., Bernard, P., Makropoulos, K., Pacchiani, F., and Patau, G. (2004). First results of the crln seismic network in the western corinth rift: evidence for old-fault reactivation. *Comptes Rendus Geoscience*, 336(4-5):343–351.

Makropoulos, K., Kaviris, G., and Kouskouna, V. (2012). An updated and extended earthquake catalogue for greece and adjacent areas since 1900. *Natural Hazards and Earth System Sciences*, 12(5):1425–1430.

Marone, C. (1998). Laboratory-derived friction laws and their application to seismic faulting. *Annual Review of Earth and Planetary Sciences*, 26(1):643–696.

Marti, E. (2018). Estimation des taux de séismes le long de la faille du levant (israël/jordanie) à partir des données publiées. *M2, IPGP/IRSN*.

McGuire, R. K. (1976). Fortran computer program for seismic risk analysis. Technical report, US Geological Survey,.

McGuire, R. K. (1995). Probabilistic seismic hazard analysis and design earthquakes: closing the loop. *Bulletin of the Seismological Society of America*, 85(5):1275–1284.

Meghraoui, M., Aksoy, M. E., Akyüz, H. S., Ferry, M., Dikbaş, A., and Altunel, E. (2012). Paleoseismology of the north anatolian fault at güzelköy (ganos segment, turkey): Size and recurrence time of earthquake ruptures west of the sea of marmara. *Geochemistry, Geophysics, Geosystems*, 13(4).

Mignan, A., Danciu, L., and Giardini, D. (2015). Reassessment of the maximum fault rupture length of strike-slip earthquakes and inference on m max in the anatolian peninsula, turkey. *Seismological Research Letters*, 86(3):890–900.

Murru, M., Akinci, A., Falcone, G., Pucci, S., Console, R., and Parsons, T. (2016). M >= 7 earthquake rupture forecast and time-dependent probability for the sea of marmara region, turkey. *Journal of Geophysical Research: Solid Earth*, 121(4):2679–2707.

Nason, R. D. (1973). Fault creep and earthquakes on the san andreas fault. *Stanford Univ. Publs. geol. Sci*, 13:275–85.

Oglesby, D. D., Mai, P. M., Atakan, K., and Pucci, S. (2008). Dynamic models of earthquakes on the north anatolian fault zone under the sea of marmara: Effect of hypocenter location. *Geophysical Research Letters*, 35(18).

Ordaz, M., Aguilar, A., García-Elías, A., and Córdova, A. (2017). Analysis of validation results of crisis considering the peer project (preliminary version).

Ozmen, H., Inel, M., Meral, E., and Bucakli, M. (2010). Vulnerability of low and mid-rise reinforced concrete buildings in turkey. In *Proceedings of the 14th European conference on earthquake engineering, Ohrid, Macedonia.*

Pace, B., Visini, F., and Peruzza, L. (2016). Fish: Matlab tools to turn fault data into seismic-hazard models. *Seismological Research Letters*, 87(2A):374–386.

Pagani, M., Monelli, D., Weatherill, G., Danciu, L., Crowley, H., Silva, V., Henshaw, P., Butler, L., Nastasi, M., Panzeri, L., et al. (2014). Openquake engine: An open hazard (and risk) software for the global earthquake model. *Seismological Research Letters*, 85(3):692–702.

Page, M. and Felzer, K. (2015). Southern san andreas fault seismicity is consistent with the gutenberg–richter magnitude–frequency distribution. *Bulletin of the Seismological Society of America*, 105(4):2070–2080.

Pantosti, D., De Martini, P. M., Koukouvelas, I., Stamatopoulos, L., Palyvos, N., Pucci, S., Lemeille, F., and Pavlides, S. (2004). Palaeoseismological investigations of the aigion fault (gulf of corinth, greece). *Comptes Rendus Geoscience*, 336(4-5):335–342.

Pantosti, D., Pucci, S., Palyvos, N., De Martini, P., D'Addezio, G., Collins, P., and Zabci, C. (2008). Paleoearthquakes of the düzce fault (north anatolian fault zone): Insights for large surface faulting earthquake recurrence. *Journal of Geophysical Research: Solid Earth*, 113(B1).

Papazachos, B. and Papazachou, C. (2003). The earthquakes of greece, ziti publ. *Co., Thessaloniki, Greece*, page 286.

Parsons, T., Geist, E. L., Console, R., and Carluccio, R. (2018). Characteristic earthquake magnitude frequency distributions on faults calculated from consensus data in california. *Journal of Geophysical Research: Solid Earth*, 123(12):10–761.

Reilinger, R., Ergintav, S., Bürgmann, R., McClusky, S., Lenk, O., Barka, A., Gurkan, O., Hearn, L., Feigl, K., Cakmak, R., et al. (2000). Coseismic and postseismic fault slip for the 17 august 1999, m= 7.5, izmit, turkey earthquake. *Science*, 289(5484):1519–1524.

Reilinger, R., McClusky, S., Vernant, P., Lawrence, S., Ergintav, S., Cakmak, R., Ozener, H., Kadirov, F., Guliev, I., Stepanyan, R., et al. (2006). Gps constraints on continental deformation in the africa-arabia-eurasia continental collision zone and implications for the dynamics of plate interactions. *Journal of Geophysical Research: Solid Earth*, 111(B5).

Richards-Dinger, K. and Dieterich, J. H. (2012). Rsqsim earthquake simulator. *Seismological Research Letters*, 83(6):983–990.

Rizzo, P., Shaw, D., and Jarecki, S. (1975). Development of real/synthetic time histories to match smooth design spectra. *Nuclear Engineering and Design*, 32(1):148–155.

Rockwell, T., Ragona, D., Seitz, G., Langridge, R., Aksoy, M. E., Ucarkus, G., Ferry, M., Meltzner, A. J., Klinger, Y., Meghraoui, M., et al. (2009). Palaeoseismology of the north anatolian fault near the marmara sea: implications for fault segmentation and seismic hazard. *Geological Society, London, Special Publications*, 316(1):31–54.

Schmittbuhl, J., Karabulut, H., Lengliné, O., and Bouchon, M. (2016a). Long-lasting seismic repeaters in the central basin of the main marmara fault. *Geophysical Research Letters*, 43(18):9527–9534.

Schmittbuhl, J., Karabulut, H., Lengliné, O., and Bouchon, M. (2016b). Seismicity distribution and locking depth along the main marmara fault, turkey. *Geochemistry, Geophysics, Geosystems*, 17(3):954–965.

Schorlemmer, D., Wiemer, S., and Wyss, M. (2005). Variations in earthquake-size distribution across different stress regimes. *Nature*, 437(7058):539.

Schwartz, D. P. (2018). Past and future fault rupture lengths in seismic source characterization—the long and short of it. *Bulletin of the Seismological Society of America*, 108(5A):2493–2520.

Schwartz, D. P. and Coppersmith, K. J. (1984). Fault behavior and characteristic earthquakes: Examples from the wasatch and san andreas fault zones. *Journal of Geophysical Research: Solid Earth*, 89(B7):5681–5698.

Şengör, A. C. and Zabcı, C. (2019). The north anatolian fault and the north anatolian shear zone. In *Landscapes and Landforms of Turkey*, pages 481–494. Springer.

Sesetyan, K., Demircioglu, M. B., Duman, T. Y., Can, T., Tekin, S., Azak, T. E., and Fercan, Ö. Z. (2018). A probabilistic seismic hazard assessment for the turkish territory—part i: the area source model. *Bulletin of Earthquake Engineering*, 16(8):3367–3397.

Shaw, B. E., Milner, K. R., Field, E. H., Richards-Dinger, K., Gilchrist, J. J., Dieterich, J. H., and Jordan, T. H. (2018). A physics-based earthquake simulator replicates seismic hazard statistics across california. *Science advances*, 4(8):eaau0688.

Sokos, E., Zahradník, J., Kiratzi, A., Janskỳ, J., Gallovič, F., Novotny, O., Kosteleckỳ, J., Serpetsidaki, A., and Tselentis, G.-A. (2012). The january 2010 efpalio earthquake sequence in the western corinth gulf (greece). *Tectonophysics*, 530:299–309.

Stepp, J. (1972). Analysis of completeness of the earthquake sample in the puget sound area and its effect on statistical estimates of earthquake hazard. In *Proc. of the 1st Int. Conf. on Microzonazion, Seattle*, volume 2, pages 897–910.

Stirling, M. and Gerstenberger, M. (2018). Applicability of the gutenberg–richter relation for major active faults in new zealandapplicability of the gutenberg–richter relation for major active faults in new zealand. *Bulletin of the Seismological Society of America*, 108(2):718–728.

Stucchi, M., Rovida, A., Capera, A. G., Alexandre, P., Camelbeeck, T., Demircioglu, M., Gasperini, P., Kouskouna, V., Musson, R., Radulian, M., et al. (2013). The share european earthquake catalogue (sheec) 1000–1899. *Journal of Seismology*, 17(2):523–544.

Thingbaijam, K. K. S., Martin Mai, P., and Goda, K. (2017). New empirical earthquake source-scaling laws. *Bulletin of the Seismological Society of America*, 107(5):2225–2246.

Toda, S., Stein, R. S., Richards-Dinger, K., and Bozkurt, S. B. (2005). Forecasting the evolution of seismicity in southern california: Animations built on earthquake stress transfer. *Journal of Geophysical Research: Solid Earth*, 110(B5).

Toro, G. R., Abrahamson, N. A., and Schneider, J. F. (1997). Model of strong ground motions from earthquakes in central and eastern north america: best estimates and uncertainties. *Seismological Research Letters*, 68(1):41–57.

Tullis, T. E., Richards-Dinger, K., Barall, M., Dieterich, J. H., Field, E. H., Heien, E. M., Kellogg, L. H., Pollitz, F. F., Rundle, J. B., Sachs, M. K., et al. (2012). A comparison among observations and earthquake simulator results for the allcal2 california fault model. *Seismological Research Letters*, 83(6):994–1006.

Visini, F., Valentini, A., Chartier, T., Scotti, O., and Pace, B. (2019). Computational tools for relaxing the fault segmentation in probabilistic seismic hazard modelling in complex fault systems. *Pure and Applied Geophysics*, pages 1–23.

Wang, Y.-J., Lee, Y.-T., Chan, C.-H., and Ma, K.-F. (2016). An investigation of the reliability of the taiwan earthquake model psha2015. *Seismological Research Letters*, 87(6):1287–1298.

Wells, D. L. and 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(4):974–1002.

Wesnousky, S. G. (1986). Earthquakes, quaternary faults, and seismic hazard in california. *Journal of Geophysical Research: Solid Earth*, 91(B12):12587–12631.

Wesnousky, S. G. (1988). Seismological and structural evolution of strike-slip faults. *Nature*, 335(6188):340.

Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grünthal, G., Valensise, G., Arvidsson, R., Basili, R., Demircioglu, M. B., et al. (2015). The 2013 european seismic hazard model: key components and results. *Bulletin of Earthquake Engineering*, 13(12):3553–3596.

Yamamoto, R., Kido, M., Ohta, Y., Takahashi, N., Yamamoto, Y., Pinar, A., Kalafat, D., Özener, H., and Kaneda, Y. (2019). Seafloor geodesy revealed partial creep of the north anatolian fault submerged in the sea of marmara. *Geophysical Research Letters*, 46(3):1268–1275.

Yazdani, A., Nicknam, A., Eftekhari, S. N., and Dadras, E. Y. (2016). Sensitivity of near-fault psha results to input variables based on information theory. *Bulletin of the Seismological Society of America*, 106(4):1858–1866.

Youngs, R. R. and 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(4):939–964.

Zhao, J. X., Zhang, J., Asano, A., Ohno, Y., Oouchi, T., Takahashi, T., Ogawa, H., Irikura, K., Thio, H. K., Somerville, P. G., et al. (2006). Attenuation relations of strong ground motion in japan using site classification based on predominant period. *Bulletin of the Seismological Society of America*, 96(3):898–913.

Annex 1 : Annexes of chapter 1

A1.1 Supplementary material of Chartier et al. [2017]

Figure 1, presents an example of the earthquake rates calculation step by step for a fault system composed of three faults.

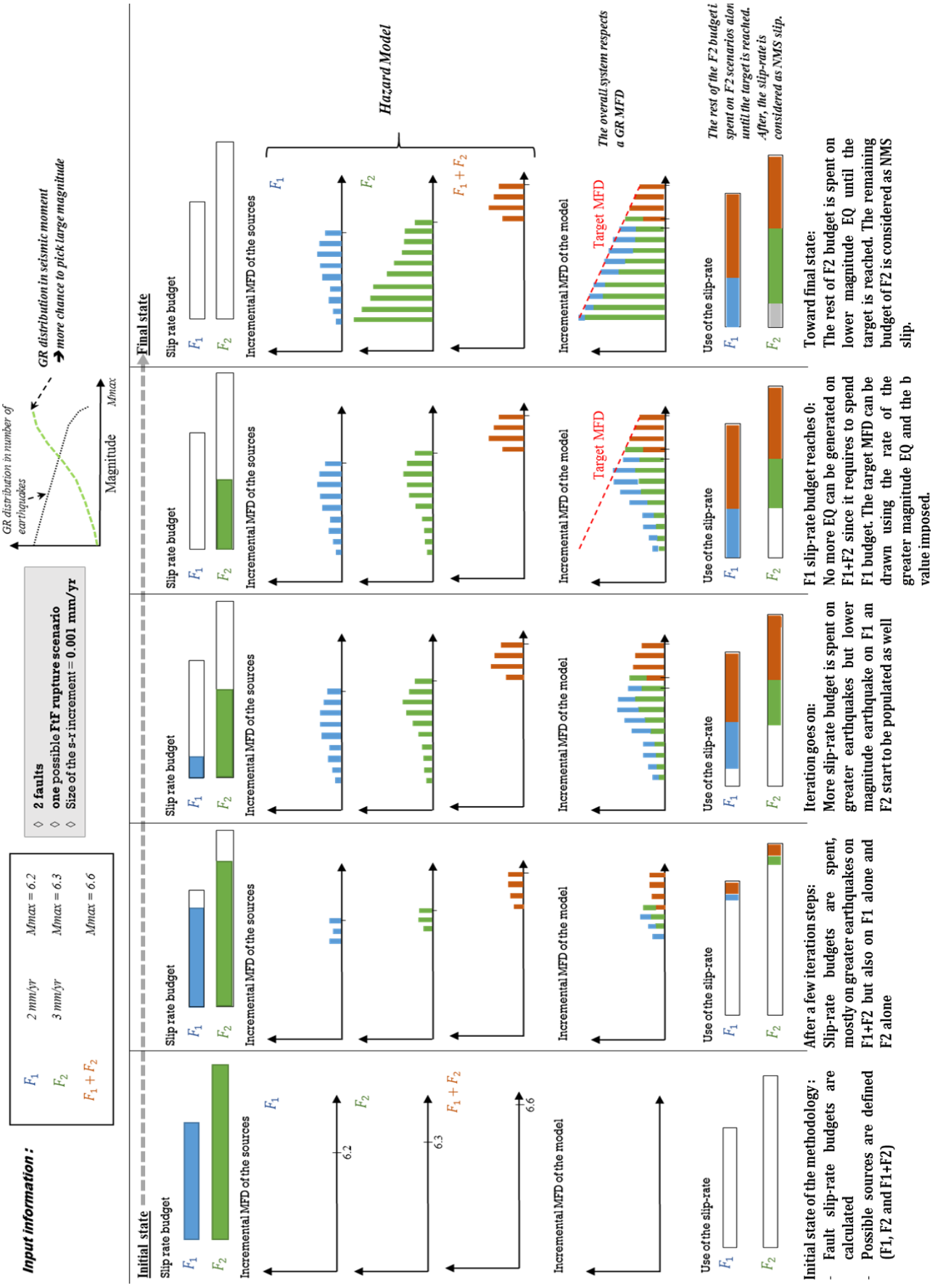

Figure 1: *Illustration of the methodology step by step.*

A1.2 Supplementary material of Chartier et al. [2019]

A1.2.1 The Modified Young and Coppersmith (1985)

For some fault systems, depending on the spatial scale of the study, the Gutenberg and Richter (1954; hereafter, GR) magnitude–frequency distribution (MFD) might not be appropriate. It has been shown that a combined MFD with a log decrease with magnitude and a characteristic earthquake might be more appropriate to explain the rate observed at different time scales (Young and Coppersmith, 1985, hereafter, YC; Zielke [2018]). The YC MFD aims to better represent this bimodal distribution. The bimodal distribution proposed by YC concentrates an important proportion of the moment rate on the higher magnitude range of the distribution. When working with a fixed moment-rate budget as SHERIFS does, it implies then that the overall absolute value of the earthquake rate will be very dependent on the maximum magnitude in the fault system. Because SHERIFS allows exploring the maximum magnitude within the range admitted by the scaling law, a change of 0.1 in Mmax between two models with the same overall moment rate can result in an important difference in the rate of lower magnitude earthquakes (figure 2).

To minimize this effect, we propose the following modified YC MFD expression

$$p = \begin{cases} \frac{\beta e^{-\beta(M_{max}-1.5-M_{min})}}{1-e^{-\beta(M_{max}-M_{min})}} \times \frac{\phi((M-(M_{max}-0.25))\times 5)}{\phi(0)} \times s, & if M_{max}-0.5 < M < M_{max} \\ (\frac{\beta e^{-\beta(M-M_{min})}}{1-e^{-\beta(M_{max}-M_{min})}}) \times e^{\frac{1}{b(M_{max}-M_f)}}, & if M_{min} < M < M_{max}-0.5 \\ 0, & if M > M_{max} \end{cases}$$

in which b is the slope of the logarithmic part of the distribution, $\beta = ln(10) \times b$, M_{max} is the maximum magnitude in the fault system, M_{min} is the minimal magnitude of earthquakes considered for the model, M_f is the optimal maximum magnitude (see the definition in the next paragraph), ϕ is a normal probability distribution of mean = 0 and $\sigma = 1$ and s is the scaling factor for the rate of characteristic earthquakes (default value = 1.0).

This expression allows reducing the variation of the rate of smaller earthquakes when the maximum magnitude varies but the moment rate of the MFD remains constant. In this formulation, the relative shape between the exponential part and the characteristic part of the distribution becomes dependent on the difference between the actual Mmax and Mf, the optimal Mmax. The user needs to define a magnitude Mf within the SHERIFS code such that when Mmax = Mf the distribution is in the shape of an YC distribution in which the characteristic part is a Gaussian distribution of width 0.5, centered on Mmax 0.25. If Mmax < Mf, the rate of earthquakes in the characteristic part is increased in proportion to the exponential part of the distribution. If Mf < Mmax, this rate is decreased. Figure S1 illustrates how this allows reducing the variability of the rate of small-magnitude earthquakes while exploring the maximum magnitude.

The parameter s allows increasing or decreasing the ratio between the logarithmic part of the distribution and the characteristic part. If s = 1, the rate at the center of the normal distribution defining the characteristic earthquake is equal to the rate in the logarithmic distribution for a magnitude equal to the maximum magnitude minus 1.5; if s > 1, this rate is higher, if s < 1 this rate is lower.

We choose to define the characteristic part of the distribution using a normal distribution. Wells and

Coppersmith (1985) use a boxcar distribution instead. Because of the lack of data in most fault systems for characterizing the rate of large earthquake precisely, this choice is open to discussion. SHERIFS being an open-source code, we encourage the user to try other distributions if needed.

In the SHERIFS methodology, modeled earthquake rates are compared with the rates calculated from the earthquake catalog, mostly populated with low-magnitude earthquakes. Because the actual size and shape of the characteristic part of the MFD cannot be constrained by the too few large earthquakes in the catalog; with the proposed MFD, it is possible; given a fixed moment-rate budget to explore uncertainties on the maximum magnitude in the system while keeping the rate of low-magnitude earthquakes stable.

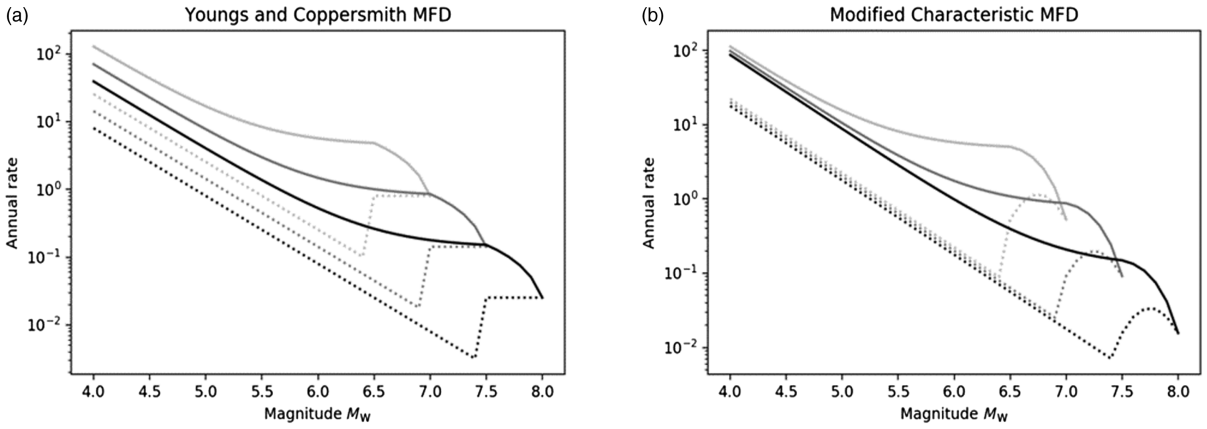

Figure 2: *Illustration of (a) the YC MFD and (b) the modified characteristic MFD proposed in this article. The exponential part of the distribution has a b-value of 1, the Mf value of the modified characteristic distribution is fixed at 7.5 and the scaling factor s = 1. Three maximum magnitudes are explored: 7.0 (light gray curve), 7.5 (dark gray curve), and 8.0 (black curve). All of the six MFDs have the same moment rate of 1020 N·m. Solid curves are the cumulative distribution (M m) and dotted curves are the incremental distributions.*

A1.2.2 Host Models

SHERIFS hazard model for the fault system can be easily incorporated into a larger "host model." The host model can be any model for the region in the form of an OpenQuake input file. This allows the user to run calculations integrating other types of sources such as area zones for the regions surrounding the fault system and to use the resulting model for hazard and risk studies for sites outside of the background zone.

SHERIFS cuts an imprint in the host model using the user-defined background shape and then pastes the background and the modeled fault sources in the cut. The geometries and earthquake rates of the host model cropped sources are then automatically redefined. The seismic activity of the modified host zones is reduced to take into account the reduction of area where the background zone defined around the faults overlaps with the zone. The seismic activity of the host area sources is not recalculated using the earthquake catalog.

Bibliography

Chartier, T., Scotti, O., and Lyon-Caen, H. (2019). Sherifs: Open-source code for computing earthquake rates in fault systems and constructing hazard models. *Seismological Research Letters*.

Chartier, T., Scotti, O., Lyon-Caen, H., and Boiselet, A. (2017). Methodology for earthquake rupture rate estimates of fault networks: example for the western corinth rift, greece. *Natural Hazards and Earth System Sciences*, 17(10):1857–1869.

Zielke, O. (2018). Earthquake recurrence and the resolution potential of tectono-geomorphic recordsearthquake recurrence and the resolution potential of tectono-geomorphic records. *Bulletin of the Seismological Society of America*, 108(3A):1399–1413.

Annex 2 : Discussion on Machine learning application in seismic risk assessment

Machine learning allows fast calculation for complex systems after the model has been trained on a dataset (Bergen et al. [2019]). The data can be actual observations or output of models. Machine learning algorithms can extract trends that would be difficult to identify using statistical analysis alone. The resulting algorithm can have a tremendous predictive power for other areas of application and be very computationally efficient.

However, the resulting algorithm can only be as good as the data it's trained on. While there are many possible applications in the field of seismology (Kong et al. [2018]), for seismic hazard and risk assessment technical limitations are likely to reduce the applicability of these methods.

Hazard is often mostly controlled by local sources since the intensity of the ground motion decreases with the square of the distance of the source to the site. Since the number of records of large earthquakes in a given region is very sparse, ergodic assumptions would be required in order to gather an amount of data large enough to train and test an algorithm.

Machine learning doesn't aim to explain a phenomenon but rather aims to produce an operational tool that can be used even without a full understanding of the mechanisms at work. Algorithms can be tested against independent data in order to assess their accuracy. For seismic hazard and risk assessment, dealing with rare and extreme events, the amount of data for testing is likely to be too small.

Furthermore, users of hazard and risk estimates are decision makers who don't necessarily have a scientific background. The robustness of the products are under intense scrutiny since they can have important social and economical impacts. If the model is based on state of the art modelling and hypotheses about earthquakes, the decision makers can understand the validity and limitations of the results and elaborate strategies accordingly. The statistical test of the validity of a PSHA results remains an extremely difficult task due to the low probability of moderate or large events that affect the hazard. However, each component of the model can be tested against data (earthquake rate, ground motion prediction equation).

In the case of a model relying on machine learning, proving the statistical validity against independent data will be difficult. Machine learning generates "black box" algorithms in which the different components cannot be independently tested against data. It will be difficult for decision makers to rely on these models that offer less visibility.

The start-up company One Concern developed a risk mitigation tool based on machine learning that has been used by different public entities including the city of San Francisco. After realizing the lack of consistency between different outputs and the lack of transparency of the model, the city of San Francisco has decided to terminate the contract with One Concern. According to the head of the city's emergency management department : "We can't be cutting-cutting edge without knowing for sure that we can validate data" (`https://www.nytimes.com/2019/08/09/us/emergency-response-disaster-technology.html`).

Nonetheless, while a direct application of the machine learning algorithms to seismic risk assessment seems limited due to the lack of transparency of the algorithms and the fundamental lack of training and testing data, the

many possible applications in the fields of seismology, dynamic modelling and geodesy (Bergen et al. [2019]) will provide improved inputs for the classical PSHA approaches. Recent studies using machine learning showed very promising results that will likely lead to improvements in the accuracy of seismic hazard assessment.

For example, Rouet-Leduc et al. [2018] have demonstrated the capability of machine learning for inferring stress evolution surrounding a laboratory fault from acoustic emissions only once the algorithm has been trained. Rouet-Leduc et al. [2017] have trained an algorithm to predict the time to future rupture in the laboratory. These type of studies will bring valuable information for the calibration of physics based models and for the setting of new probability functions that can be used in time-dependant probabilistic seismic hazard assessment.

In the same way that physics-based models do not have to be perfect to be able to bring valuable input to the risk assessment and to reduce the uncertainty, as we have shown in this study, the machine learning algorithms will bring new insight on problems that classical models have trouble investigating adequately.

Bibliography

Bergen, K. J., Johnson, P. A., Maarten, V., and Beroza, G. C. (2019). Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363(6433):eaau0323.

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., and Gerstoft, P. (2018). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, 90(1):3–14.

Rouet-Leduc, B., Hulbert, C., Bolton, D. C., Ren, C. X., Riviere, J., Marone, C., Guyer, R. A., and Johnson, P. A. (2018). Estimating fault friction from seismic signals in the laboratory. *Geophysical Research Letters*, 45(3):1321–1329.

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., and Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, 44(18):9276–9282.

Annex 3 : List of contributions

A3.1 Articles

- Methodology for earthquake rupture rate estimates of fault networks: example for the western Corinth rift, Greece. Thomas Chartier, Oona Scotti, Hélène Lyon-Caen, Aurélien Boiselet, Natural Hazards and Earth System Sciences 2017
- SHERIFS: Open-Source Code for Computing Earthquake Rates in Fault Systems and Constructing Hazard Models. Thomas Chartier, Oona Scotti, Hélène Lyon-Caen. SRL. 2019
- Computational Tools for Relaxing the Fault Segmentation in Probabilistic Seismic Hazard Modelling in Complex Fault Systems. F Visini, A Valentini, T Chartier, O Scotti, B Pace - Pure and Applied Geophysics, 2019
- Modelling earthquake multi-fault ruptures across complex fault systems based on geological data: the Eastern Betics Shear Zone. Octavi Gomez-Novell, Thomas Chartier, Julian Garcia-Mayordomo, María Ortuño, Eulalia Masana, Juan Miguel Insua- Arévalo, Oona Scotti. Engineering Geology. submitted.
- Transposing an active fault database into a seismic hazard fault model for nuclear facilities–Part 1: Building a database of potentially active faults (BDFA) for metropolitan France . Jomard, Hervé and Cushing Edward Marc and Palumbo Luigi and Baize Stéphane and David Claire and Chartier Thomas. Natural Hazards and Earth System Sciences. 2017
- Transposing an active fault database into a fault-based seismic hazard assessment for nuclear facilities – Part 2: Impact of fault parameter uncertainties on a site-specific PSHA exercise in the Upper Rhine Graben, eastern France, Chartier Thomas and Scotti Oona and Clément Christophe and Jomard Hervé and Baize Stéphane, Natural Hazards and Earth System Sciences. 2017

A3.2 Posters and Presentations

- 2017
  - Journées des thèses IRSN
  - Congrès des doctorants ED
- 2018
  - EGU
    - Seismic Hazard and Earthquake Rate In Fault Systems–SHERIFS: a python tool for earthquake geologists, fault-hazard modelers and decision-makers. Thomas Chartier, Scotti Oona, Lyon-Caen Hélène
    - Earthquake rupture hypotheses at the Eastern Betic Shear Zone (SE Spain): what does the SHERIFS approach tell about their magnitude-frequency distributions? Octavi Gómez-Novell, Julián García-Mayordomo, María Ortuño, Eulàlia Masana, Robert López, Thomas Chartier, Oona Scotti

- SCEC Annual Meeting : Constraining epistemic uncertainties on hazard models in the Marmara region using SHERIFS (Seismic Hazard and Earthquake Rates in Fault Systems). Thomas Chartier, Oona Scotti, Hélène Lyon-Caen
- Fault2SHA 2018 KAUST
  - Seismic Hazard and Earthquake Rates in Fault Systems: The SHERIFS methodology applied in the Marmara region (Turkey). Thomas Chartier, Oona Scotti, Keith Richards-Dinger, James Dieterich, and Hélène Lyon-Caen.
  - Need for inclusion of physics-based fault models into seismic hazard assessment: The example of the Dead Sea Fault. Thomas Chartier, Etienne Marti, Oona Scotti, Yann Klinger
- Journées des thèses IRSN
- Congrès des doctorants ED

- 2019
  - Fault2SHA Barcelona : REDUCING UNCERTAINTIES IN SEISMIC RISK ESTIMATES IN ISTANBUL BY COMBINING STATISTICS- AND PHYSICS-BASED APPROACHES by: Thomas Chartier, Oona Scotti, Hélène Lyon-Caen, Keith Richards-Dinger, Jim Dieterich, Aurélien Boiselet.
  - EGU : Physics-based earthquake simulators as a tool in Fault-PSHA: example for the Marmara region, Turkey. Thomas Chartier, Oona Scotti, Keith Richards-Dinger, James Dieterich, and Hélène Lyon-Caen.
  - Journées des thèses IRSN
  - Congrès des doctorants ED

A3.3 SHERIFS GitHub and User Manual

SHERIFS code is available at `https://github.com/tomchartier/SHERIFS` with a user manual.

Figure 1: *Screen shot of the SHERIFS user manual. Available at* `https://github.com/tomchartier/SHERIFS`.

RÉSUMÉ

Les taux de sismicité sont un élément clé du calcul probabiliste de l'aléa et du risque sismique. Une méthodologie innovante est développée pour modéliser les taux de ruptures complexes dans un système de failles (SHERIFS). La flexibilité de la méthodologie permet de la partager et de l'appliquer sur des systèmes de failles différents et de s'en servir comme outil de discussion des hypothèses concernant la sismicité sur les failles. Les taux de sismicité dans la partie Ouest du Rift de Corinthe en Grèce et dans la région de Marmara en Turquie sont modélisés en explorant l'incertitude épistémique sur le scenario de rupture maximale, la distribution en magnitude et fréquence ou encore les conditions de glissement sur la faille. A Marmara, les hypothèses sont pondérées dans un arbre logique en comparant les taux modélisés aux taux calculés à partir des données et des résultats d'un modèle physique basé sur les équations "rate and state". Pour chaque hypothèse de l'arbre logique, le risque sismique à Istanbul, en termes de probabilité d'effondrement d'un immeuble, est calculé pour deux bâtiments de même type construits respectivement suivant le code de construction de 1975 et celui de 1998. Le risque sismique est six fois plus important pour le bâtiment le plus ancien. Parmi les incertitudes explorées, la source d'incertitude plus importante est liée à la distribution en magnitude et fréquence. L'utilisation des données et du modèle physique permet de réduire l'incertitude sur le risque par un facteur 1.6. Une nouvelle méthodologie de désagrégation du risque permet de montrer que le risque à Istanbul est contrôlé en partie par les séismes de magnitude supérieure à 7 sur la Faille Nord Anatolienne et en partie par les séismes de magnitude modérée (4.5 à 6) dans la sismicité de fond, sur des failles non connues, et à une distance du bâtiment inférieure à 10 km.

MOTS CLÉS

Taux de Sismicité, Faille, Système de Failles, Aléa Probabiliste, Risque Sismique, Istanbul, Faille Nord Anatolienne, Désagrégation

ABSTRACT

Earthquake rates are a key component of probabilistic seismic hazard and risk assessment. A novel methodology to model earthquake rates in complex ruptures in a fault system (SHERIFS) is developed. The flexibility of the methodology allows sharing it and applying it to various fault systems as well as using it as a tool to discuss the hypotheses concerning the seismicity on faults. The earthquake rates in the Western Corinth Rift in Greece and the Marmara Region in Turkey are modeled while exploring the associated epistemic uncertainty such as on the maximum rupture scenario, the magnitude frequency distribution or the locking condition of the faults. In Marmara, the hypotheses are weighted in a logic tree by comparing them to the rate calculated from the data (earthquake catalog and paleoseismicity) and the results of a "physics-based" model applying the rate and state equations. For each hypothesis in the logic tree, the seismic risk in Istanbul, represented by the probability of collapse of a building, is calculated for two buildings of the same type but one constructed following the 1975 building code and on following the 1998 building code. The seismic risk is six times larger for the older building. Within the explored uncertainties, the one concerning the magnitude frequency distribution is the largest source of uncertainty on the risk of collapse. By using the data and the physics-based model to weight the logic tree, the uncertainty on the risk is reduced by a factor of 1.6. Through a novel risk deaggregation methodology, we identify two seismogenic sources controlling the risk in Istanbul: the large earthquakes (magnitude larger than 7) on the neighboring North Anatolian Fault and the moderate magnitude earthquakes (magnitude between 4.5 and 6) occurring in the background zone, on unknown faults, and at a distance shorter than 10 km from the building.

KEYWORDS

Earthquake Rates, Fault, Fault System, Probabilistic Hazard, Seismic Risk, Istanbul, North Anatolian Fault, Deaggregation