Author:
Mancini, Simone

Title:
Improving the Performance of Physics-based and Statistical Forecasts of Natural and Induced Seismicity

Improving the Performance of Physics-based and Statistical Forecasts of Natural and Induced Seismicity

By

Simone Mancini

School of Earth Sciences
University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

September 2020

Word count: 42,820

Abstract

The demand for reliable and replicable short-term probabilistic earthquake forecasts is becoming increasingly compelling, as we continue to witness seismic sequences with occasionally multiple disturbing or damaging earthquakes.
Purely statistical models of earthquake clustering adequately capture the patterns of triggered seismicity and currently represent the standard approach for different operational earthquake forecasting systems.
On the other hand, developing and testing physics-based forecast models let us validate the most popular physical hypotheses for earthquake triggering and clustering. These models couple complex stress interactions between faults with laboratory-derived frictional laws providing a framework for earthquake forecasting in the context of continuum mechanics. However, while featuring the unique characteristic of integrating many products of observational seismology, they are extremely data-intensive especially in near real-time settings where their applicability is still contentious.
Over the last decade, the scientific advancements in seismology have provided higher resolution seismic catalogues as well as improved fault characterisations; this presents us with great opportunities to (1) evaluate their usefulness in improving the short-term performance of models of both forecast categories, (2) explore which specific modelling choices driven by real-time data quality and availability boost our forecasting skills and by how much, and (3) assess what are the data products required for such model improvements to be operationally delivered.
To answer the above points, this thesis presents three forecasting experiments offering a novel experimental strategy, where the absolute and relative performance of statistical and physics-based models is formally quantified under different forecasting modes and modelling choices, in both cases of tectonic and induced seismicity.
Looking ahead to the future improvements in near real-time input data quality promised by the most recent progresses in artificial intelligence techniques, the results of these experiments suggest what are the pathways that should be undertaken for future model developments.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: Simone Mancini ..... DATE: 17/09/2020 .....

Acknowledgements

First and foremost, I would like to thank my supervisors Margarita Segou (British Geological Survey) and Max Werner (University of Bristol) for their support, patience, and availability. Our discussions represented key learning moments for me, and their constructive comments were of fundamental importance to develop my own critical thinking.

I am indebted to Camilla Cattania (Stanford University, now at MIT) for the many Skype calls we had (sometimes at quite odd hours given the time difference) to discuss about my work and to guide me through her code. Her advice was extremely important to accomplish the results of this thesis.

I would like to thank all the participants to the NERC-NSF supported project "*The Central Apennines Earthquake Sequence Under a New Microscope*" (NE/R0000794/1) for providing feedback on my work and for facilitating fruitful conversations about its potential future directions.
I am also very grateful to the organisers of the session "Better Earthquake Forecasts" at the 2019 SSA Annual Meeting in Seattle (WA) and to the conveners of the CSEP workshop at the 2019 SCEC Annual Meeting in Palm Springs (CA) for inviting me to give oral talks and present my research at renowned international conferences.

A very special thank goes to my family for their love and the full support they always give me in many different ways.

Finally, thanks Erika for understanding and motivating me, and for the many happy moments over the last four years, especially during the stressful months of the Covid-19 lockdown when we took good care of each other.

This Ph.D. project was funded by a GW4+ Doctoral Training Partnership studentship from the Natural Environment Research Council (NE/L002434/1) and by the British Geological Survey University Funding Initiative (BUFI) studentship (S350).

Contents

List of Figures

List of Tables

Acronyms and Symbols

General acronyms

ANSS	Advanced National Seismic System
CRS	Coulomb Rate-and-State
CSEP	Collaboratory for the Study of Earthquake Predictability
DISS	Database of Individual Seismogenic Sources
EGS	Enhanced Geothermal System
ETAS	Epidemic-Type Aftershock Sequence model
HF	Hydraulic Fracturing
SCEDC	Southern California Earthquake Data Center
SCSN	Southern California Seismic Network
UCERF3	Uniform California Earthquake Rupture forecast (third version)
USGS	United States Geological Survey

CRS parameters and acronyms

CFF	Coulomb Failure Function
τ	Shear stress
μ'	Coefficient of effective friction
σ_n	Normal stress
B_k	Skempton's coefficient
R	Seismicity rate
r_o	Background seismicity rate
γ	State variable
$\dot{\tau}$	Secular shear stressing rate

ΔS	Stress step
Aσ	Fault constitutive parameter
t_a	Aftershocks duration time
OOP	Optimally Oriented Planes
GRP	Geological Receiver Planes
USD	Uniform slip distribution
FFM	Finite-fault slip model
I	Isotropic stress field
SUP	Spatially uniform receiver planes
SVP	Spatially variable receiver planes
SSI	Smoothed stress inversion
Opt RS	Optimised rate-and-state parameters
HoBR	Spatially homogeneous background seismicity rate
HeBR	Spatially heterogeneous background seismicity rate

ETAS parameters

λ	Conditional intensity function
H_t	Seismic history
μ	Background rate
K_0	Productivity parameter
α	Productivity parameter
c	Omori-Utsu c-value
p	Omori-Utsu p-value
d	Spatial kernel parameter
q	Spatial kernel parameter
γ	Aftershock zone scaling factor
n	Branching ratio
c_f	Constant of proportionality between injection rate and seismicity rate

Performance evaluation metrics

LL	Log-likelihood
λ	Expected seismicity rate
ω	Observed seismicity rate
Λ	Forecasted seismicity catalogue
Ω	Observed seismicity catalogue
jLL	Joint log-likelihood
δ_1, δ_2	N-test quantiles
R_N	N-test rejection ratio
IG	Information gain per earthquake

Others

M_0	Seismic moment
M_L	Local magnitude
M_w	Moment magnitude
M_c	Magnitude of completeness
FM	Focal mechanisms
I_R	Injection rate
MLE	Maximum Likelihood Estimation

Publications

During the course of this Ph.D. I have published the following papers:

- Mancini, S., Segou, M., Werner, M. J., and Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016–2017 Central Italy Earthquake Cascade. *J. of Geophys. Res. Solid Earth*, 124, 8626– 8643. doi:10.1029/2019JB017874.

  Contributions: I gathered the datasets, run the models, written the manuscript and produced all figures and tables. M.S. and M.J.W. advised on the experimental design and supervised model development. C.C. provided the code and helped defining the methods. All co-authors provided feedback on the results and on the manuscript drafts.

- Mancini, S., Segou, M., Werner, M. J., and Parsons, T. (2020). The Predictive Skills of Elastic Coulomb Rate-and-State Aftershock Forecasts during the 2019 Ridgecrest, California, Earthquake Sequence. *Bull. Seismol. Soc. Am.*, 110(4), 1736–1751. doi:10.1785/0120200028.

  Contributions: I gathered the datasets, set up the experimental design, run the models, written the manuscript and produced all figures and tables. M.S. and M.J.W. have supervised model development. T.P. has consulted on large scale stress calculations. All co-authors have provided feedback on the results and on the manuscript drafts.

The following article has been submitted and is currently in review:

- Mancini, S., Werner, M. J., Segou, M., and Baptie, B. J. (2020). Probabilistic Forecasting of Hydraulic Fracturing Induced Seismicity Using an Injection-Rate-Driven ETAS model. *Submitted to Seismol. Res. Lett.*.

Contributions: I set up the experimental design, run the models, written the manuscript and produced all figures and tables. M.S. and M.J.W. have supervised model development. M.J.W. provided part of the code. B.B., M.J.W. and M.S. helped gathering the datasets from the Oil and Gas Authority. All co-authors have provided feedback on the results and on the manuscript drafts.

Chapter 1

Introduction

Earthquakes continue to produce losses of human lives and enormous damages around the world. Predicting the time, location and magnitude of destructive earthquakes has been considered the ultimate goal of modern seismology since its birth in the early 20^{th} century.
For many years, scientists have been searching for observable and measurable precursors that would enable us to precisely anticipate the occurrence of seismic events. Different hypotheses on earthquake preparatory phenomena were formulated in the 1970s and 1980s with occasional successes and several failures.
Nowadays, that early enthusiasm has mostly vanished as seismicity precursors proved unreliable, and the so-called 'deterministic' prediction of earthquakes is considered as a pipe dream by the seismological community.

The current research in the field focuses instead on probabilistic approaches that fall into the category of mathematical models known as 'earthquake forecasts'. Such modelling effort is made possible in the first place by the study of the large scale characteristics of Earth's seismicity: our knowledge of plate tectonics, paleoseismological investigations of earthquake recurrence times, together with studies of historical seismicity and fault mapping, provide us with an understanding of which areas are more likely to experience strong ground shaking in the long term (*i.e.* several decades or centuries). These elements contribute to time-independent earthquake forecasts, which usually take the form of seismic hazard maps express-

ing the convolution of long term earthquake probabilities with site conditions. This kind of research product is generally important for uses in risk management including building code developments but also covers wider applications in the field of insurance and re-insurance financial products.

However, a critical element comes from the observations that earthquakes tend to cluster in space and time in what we commonly call 'earthquake sequences'. Such empirical evidence, which is not considered in time-independent models, suggests that the occurrence of an earthquake alters the short-term spatiotemporal probabilities of further seismic events, usually called 'aftershocks'. Hence, short-term models (*e.g.* few minutes to several months) are required to capture the space-time fluctuations of the seismic hazard, as aftershock cascades can generate various moderate to large magnitude events over weeks, months or years that expand the damage zones extensively causing even more severe disruption to livelihoods.

1.1 Short-term Earthquake Forecasting

A turning point in the field of short-term earthquake forecasting is represented by the destructive 2009 L'Aquila (Central Italy) earthquake, where an accelerating pattern of M3+ seismicity was felt by the population a few days before a Mw = 6.3 mainshock that caused more than 300 fatalities. In that case, seven members of the Italian National Commission for the Forecast and Prevention of Major Risks were accused of inaccurate risk communication and eventually indicted on charges of manslaughter (for more details, see *Stucchi et al.,* 2016).
While efforts to develop time-dependent earthquake forecast models were surely under way well before the 2009 Italian case, these tragic events dramatically epitomised (1) the need to quickly move towards a community effort to establish testable aftershock forecasts to inform short-term decision-making protocols, and (2) the challenge to enhance societal awareness and preparedness to pending disasters by delivering hazard information to the general public in an effective and timely manner.
Therefore, following the L'Aquila disaster, the Italian government convened an International Commission on Earthquake Forecasting for Civil Protection (ICEF),

where a group of experts was asked to describe the state-of-the-art of short-term earthquake forecasting and provide suggestions for its future developments. In the resulting report, the framework of Operational Earthquake Forecasting (OEF) was conceptualised for the first time, and defined as: "*the continual updating of authoritative information about the future occurrence of potentially damaging earthquakes, and the officially sanctioned dissemination of this information to enhance earthquake preparedness in threatened communities*" (*Jordan et al.*, 2011).

Instead of directly focusing on the definition of OEF protocols (that actually involve critical research in other fields such as risk communication), this thesis concerns the scientific advance that supports OEF improvements in the form of testable aftershock forecast models.

These models can be divided into three categories:

1. **Statistical models**, that are based on purely empirical relationships providing a description of the probabilistic evolution of triggered seismicity. These include the modified Omori-Utsu (OU) law for the aftershock time decay, the Gutenberg-Richter (GR) law for the earthquake magnitude distribution, and other relationships that scale the aftershock area and the number of triggered events (*i.e.* the aftershock productivity) with the magnitude of the mainshock.

2. **Physics-based models** that seek to forecast the space-time distribution of future seismicity based on the physical mechanisms that are believed to drive earthquake triggering. These models are commonly named "stress-based" forecasts as they couple the co-seismic and/or post-seismic stress interactions between faults with constitutive laws that project earthquake rates over different time horizons.

3. **Hybrid models** implement a combination of the two previously mentioned categories. For example, successful hybrid models sample earthquake magnitudes and occurrence times from the GR and OU laws and then redistribute events in space using the fault-to-fault stress interaction patterns (*e.g. Steacy et al.*, 2014; *Cattania et al.*, 2018).

Because of their recognised ease of implementation and robustness in describing the short-term spatiotemporal patterns of triggered seismicity, statistical forecasts are the most widespread models currently adopted in OEF systems. In particular, the Epidemic-Type Aftershock Sequence (ETAS) forecasts (*Ogata*, 1988; 1998) have shown considerable skills in capturing the clustering characteristics of triggered seismicity and are an integral part of many OEF strategies worldwide, where they are either used alone or in combination with other models (*e.g. Gerstenberger & Rhoades*, 2010; *Marzocchi et al.*, 2014; *Field et al.*, 2017; *Omi et al.*, 2019).
ETAS models are also commonly recognised as the most robust benchmark against which modellers evaluate any performance improvement of competing forecasting techniques. On the other hand, these models offer limited insight into the physics of earthquake nucleation and short-term fault interaction in terms of continuum mechanics.
Here, an indirect benefit of OEF emerges: while representing a practical aspect of science to understand how well proposed models forecast seismicity, it also serves as a community get-together to establish the veracity of the underlying scientific hypotheses, such as those for the physical mechanisms governing earthquake triggering and clustering.

1.2 The Challenge of Stress-based Forecasts

The idea of causal relationships between the occurrence of earthquakes dates back to the second half of the twentieth century, when it was first postulated that earthquakes change the equilibrium of stresses on neighbouring faults (*Richter*, 1958). Possible correlations between areas of increased shear stresses and aftershock locations were proposed by *Das and Scholz* (1981). However, it was not until the seminal work of *Harris and Simpson* (1992) that the so-called static stress transfer hypothesis was formally formulated. According to this hypothesis, which now represents a commonly accepted physical interpretation for earthquake triggering, a discrete dislocation in the Earth's crust (*i.e.* an earthquake) statically (that is, quasi-permanently) perturbs the state of stress in the surrounding crustal volume. It follows that shear and normal stresses on fault surfaces are modified, inhibiting

some ruptures and making others more likely.
The way static stress changes are implemented, that is, the vehicle to quantify the actual amount of statically transferred stress is the Coulomb stress theory (see Chapter 2): seismicity is promoted on neighbouring faults experiencing positive static stress changes, while it is suppressed in those areas of negative stress change commonly indicated as 'stress shadows'.
Despite promising applications in several regions of the world (*e.g. Stein et al.*, 1992; *King et al.*, 1994), the stress transfer hypothesis has been challenged by many authors pointing to the fact that a non-negligible portion of triggered seismicity (in extreme cases up to 30-40%) occurs in stress shadows (*e.g. Hardebeck et al.*, 1998; *Mallmann & Parsons*, 2008). These observations highlight the actual complexity of earthquake triggering processes, that are likely to go beyond the mere coseismic stress perturbations. Another recognised triggering mechanism, especially at longer distances from the source fault, is represented by the transient perturbations due to the passage of seismic waves known as dynamic stress changes (*e.g. Gomberg et al.*, 1998; 2001; *Brodsky & van der Elst*, 2014). In addition, post-seismic processes are likely to contribute toward the observed longer term aftershock patterns (*e.g. Freed*, 2005). These include afterslip (*e.g. Perfettini*, 2004; *Ross et al.*, 2017), poro-elastic effects (*e.g. Cocco & Rice*, 2002) and, at decadal scale, viscoelastic relaxation of the lower crust (*Wang et al.*, 2012, *Diao et al.*, 2014).

Although static stress changes alone do not always fully explain triggered earthquakes patterns, they represent the basis to produce physics-based short-term aftershock forecasts. In particular, stress changes are coupled to laboratory-derived friction laws describing the seismicity response to an earthquake perturbation (*Dieterich*, 1994) to define a family of models known as Coulomb rate-and-state (CRS) forecasts (*Toda et al.*, 2005; *Cocco et al.*, 2010; *Toda & Enescu*, 2011; *Parsons et al.*, 2012, 2014; *Segou et al.*, 2016, *Cattania et al.*, 2018, among others).
During the last ∼25 years, the development of CRS models has almost gone hand-in-hand with the evolution in the production and quality of seismological data. In the early to mid '90s, earthquake catalogues were not large enough, and analyses of spatial correlations between stress changes and occurrence of subsequent

moderate-to-large magnitude events were mainly qualitative (*e.g. Stein et al.*, 1994; *Toda et al.*, 1998; *Stein*, 1999). In this regard, the 1992 Landers sequence is often reported to be the clearest example of how the static stress change signatures can drive aftershock sequences (*King et al.*, 1994).
To start supporting a more systematic testing of stress transfers within the specific CRS implementation, a critical period came in the early 2000s when the improvement of regional networks and the introduction of new algorithms to develop enhanced seismic catalogues started to bear fruit (*e.g. Waldhauser & Elsworth*, 2000). It was indeed not before the work of *Toda et al.* (2005) that CRS models could be more rigorously tested on southern California seismicity benefitting from the first large-scale relocated catalogue by *Richards-Dinger and Shearer* (2000).
However, this first generation of CRS models was usually rooted on simple descriptions of earthquake sources and hazardous faults leading to mixed successes, up to the point that *Woessner et al.* (2011) argued that their performance was nowhere near to that of competing empirical models such as ETAS and STEP (Short-Term Earthquake Probabilities; *Gerstenberger et al.*, 2005).
In the last $\sim$10 years, the level of science around CRS modelling has evolved. More complete global seismicity catalogues and increasingly detailed fault characterisations drove us into the development of a second generation of physics-based models that in preliminary experiments perform sensibly better than before (*e.g. Segou et al.*, 2013; *Cattania et al.*, 2018). These results motivate further testing of these models (1) for comparing their predicting skills against competing forecast techniques in transparent evaluation platforms such as the Collaboratory for the Study of Earthquake Predictability (CSEP, *Jordan*, 2006; *Michael & Werner*, 2018), and (2) ultimately, for future consideration in OEF protocols.

In spite of the recent improvements in physics-based modelling, a major challenge regards quantifying their actual real-time performance. Indeed, most of the past literature regarding CRS models assesses their ability to reproduce the spatiotemporal evolution of aftershock sequences in retrospective mode, that is, taking advantage of best-quality datasets (especially in terms of high-quality source models) that are extremely unlikely to be available in real-time conditions following a major earthquake. The need for a model parameterisation that goes beyond an

input earthquake catalogue (as opposed to ETAS models) and that brings together geological, seismological and geophysical information about the Earth system is still an obstacle to integrate all these data products into an operational protocol. To evaluate and quantify the predictive power of CRS models in operational contexts, and more importantly to understand which are the most critical pathways to improve their near real-time skills, an increasing number of pseudo-prospective and purely prospective experiments are needed. Both these 'blind' forecasting modes allow a so-called 'out-of-sample' validation, since they do not use any feature of the data sample that the model is intended to forecast and avoid possible biases introduced by the incorporation of refined input data that would not be achievable in near real time. In this way, essential questions may be pursued: how do CRS forecasts perform when their parameterisation is driven by real-time data availability and quality? Which model components and modelling choices are the most critical and by how much?

1.3 Forecast Models of Induced Seismicity

The interest of the seismological community toward human-induced seismicity has appeared and dramatically grown over the last few years. The main reason lies in the increased development of subsurface geo-energy reservoirs, including unconventional shale gas development, enhanced geothermal energy systems, wastewater injection, and underground storage of liquid carbon (*Ellsworth*, 2013).
In recent years, fluid-induced seismicity with moderate magnitudes (M5-5.7) in regions such as the central United States and South Korea has led to significant damages and losses (*Keranen et al.*, 2013; *Ellsworth et al.*, 2019; *Lee et al.*, 2019). In this kind of environment, injection operations entail pumping pressurised fluid at depth that promotes seismicity in previously low seismic hazard regions or further increases existing high seismic rates. As a consequence, these activities raised severe concern from governments and harsh criticism from the general public.
For specific applications such as hydraulic fracturing (HF) and geothermal exploitation, several countries around the world adopt the so-called 'traffic light' system (*Bommer et al.,* 2006) as a mitigation strategy. According to these protocols, operators are required to reduce or even stop injection if an earthquake with

magnitude larger than a set threshold occurs. Variable magnitude limits are in force in different countries, from the $M_L = 4.0$ of Canada (*Kao et al.,* 2018) down to the $M_L = 0.5$ of the UK (*Clarke et al.,* 2019).

Given this emerging type of anthropogenic hazard, and notwithstanding the huge economic interests behind the exploitation of geo-energy reservoirs, researchers are more frequently asked to provide operators and regulators with forecast models of induced seismicity. The most studied aspects regard the maximum expected magnitudes (*e.g. Clarke et al.*, 2019) and forecasts of earthquake rates in response to injected fluid volumes, rates, and pressures.

While several hypotheses about the interplay of physical mechanisms controlling the seismic response to subsurface fluid injection are under investigation, our current approach in modelling such short-lived transient hazard mostly relies on probabilistic methods providing a framework for epistemic and aleatory uncertainties. Passing the natural earthquake forecast problem to the induced seismicity environment introduces new challenges. (1) Forecasting event with generally low to very low magnitudes; (2) creating forecasts with shorter time horizons: the comparative analysis of Omori's p-values of tectonic and induced seismicity rates shows that the latter often decay more quickly and can vary dramatically even within minutes; (3) developing models that need to account for additional sources of seismicity forcing (*e.g.* external fluid pumping) and their complex relationships with the induced event rates.

One consequence arising from the points illustrated above is the importance of enhanced, dense monitoring systems in fluid-induced seismicity contexts allowing for high-quality data products to be adequately developed and promptly processed within extremely short time windows. The aim is to integrate such input data into forecast models to provide operators with reliable and easily implementable tools for real-time hazard estimates.

1.4 Objectives of the Thesis and Outline

This thesis explores the modelling elements that improve the performance of short-term stress-based and statistical earthquake forecasts in operational contexts. Its goal is to provide quantifiable evidence on the type and quality of datasets required

for future model developments, as well as to assess guidelines for their near real-time implementation in cases of both tectonic and induced seismicity. To that end, we explore how individual modelling choices driven by the actual data availability affect the predictive skills of the forecasts.
Additional motivations for this dissertation are provided by a unique chance offered by stress-based modelling of aftershock sequences, that is, testing how the increasingly improving real-time data products help us to validate the existing physical hypotheses for earthquake triggering.

In Chapter 2 we present the theoretical framework and briefly discuss the state-of-the-art of Coulomb rate-and-state and ETAS models, as well as their limitations in terms of model uncertainties and parameterisation. We also introduce the statistical metrics used to evaluate model performance.

Chapters 3 is dedicated to assessing the pseudo-prospective performance of Coulomb rate-and-state models benchmarked against a standard ETAS during the 2016-2017 Central Italy aftershock cascade. We test a wide range of CRS models with increasing level of complexity tied to real-time conditions in terms of data quality and availability. Guided by a rigorous comparative model evaluation, we quantify how the out-of-sample forecasting skills are affected by the gradual incorporation of: spatially variable background seismicity rates, optimised model parameters, crustal structural heterogeneities describing hazardous faults informed by past earthquakes, spatially variable source models, and the contribution of smaller magnitude earthquakes in reshaping the co-seismic stress field (*i.e.* 'secondary triggering' effects).

We start Chapter 4 by applying a similar experimental framework to the 2019 Ridgecrest (California) earthquake sequence. While prior studies in literature tend to apply different models to separate case studies, without much evidence of replicable and robust inferences between different seismic sequences, here our aim is to explore whether the same conclusions drawn from the Central Italy experiment are also valid in a completely different tectonic setting.
Also, the data-wealthy environment of southern California allows us to expand the analysis on some critical points, specifically the influence of (1) artefacts and

errors in preliminary finite-fault slip models (2) the choice among different data sources to characterise earthquake ruptures (*i.e.* uncertainty in focal mechanisms), and (3) the incorporation of unfolding aftershock ruptures to better resolve the evolving co-seismic stresses in the near source region.

In Chapter 5, we move the focus of our models to an induced seismicity environment, specifically to the microseismicity recorded during and after hydraulic fracturing operations at the Preston New Road site in UK. Here, we probe the forecasting skills of the standard (tectonic) ETAS model and compare them against those of a modified ETAS that accounts for external forcing in the form of time-dependent fluid injection rates. Given the rich datasets coming from operations carried out at two different wells during two distinct time windows, we could also evaluate the comparative performance of ETAS models parameterised using both in-sample and out-of-sample data. The results of this study let us draw conclusions on the operational applicability of ETAS in hydraulic fracturing contexts and on which are the modelling strategies to be adopted for its successful future implementations.

Finally, in Chapter 6 we examine the significance of our findings, their limitations and discuss future directions.

Chapter 2

Formulation and Evaluation of Short-term Earthquake Forecast Models

Part of the text included in this chapter appears in the following article:

- Mancini, S., Segou, M., Werner, M. J., and Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016-2017 Central Italy Earthquake Cascade. *J. of Geophys. Res. Solid Earth*, 124, 8626-8643. `https://doi.org/10.1029/2019JB017874`.

In this chapter, we introduce the mathematical framework that we use to pass from the Coulomb theory to the creation of testable models of earthquake clustering. We also present the main features of the ETAS model, as well as the formulation of the most common statistical performance evaluation metrics implemented in CSEP.

2.1 Coulomb Rate-and-state Modelling

2.1.1 Coulomb Stress Changes

Earthquake slip on faults is approximated by tensile and shear dislocations that induce strain in the surrounding crustal volume, whose resulting deformation de-

pends on the specific rheology assumed. For an elastic half-space, the perturbation of the stress field (*i.e.* the change in the shear and normal components of the stress tensor) imparted by discrete rectangular dislocations is calculated using the analytical coefficients derived by *Okada* (1992).
According to the Mohr-Coulomb failure criterion, brittle ruptures such as earthquakes are encouraged when either (i) the shear stress (τ) acting on a fault plane increases, or (ii) a fault is unclamped following a normal stress (σ_n) reduction. The stress tensor obtained by means of the Okada solutions is then resolved on a given fault geometry (commonly known as "receiver fault") to compute τ and σ_n. These values are used to calculate the change in the value of the Coulomb Failure Function (CFF), which is a common physical quantity used to estimate the variation of the state of stresses on neighbouring faults.
Rice (1992) formulates the ΔCFF as:

$$\Delta CFF = \Delta\tau + \mu'(\Delta\sigma_n), \tag{2.1}$$

where $\Delta\tau$ is the change in shear stress resolved on a receiver fault and set positive in direction of fault slip, $\Delta\sigma$ is the change in normal stress (positive when the fault is unclamped), $\mu' = \mu(1 - B_k)$ is the effective coefficient of friction, with B_k the Skempton's coefficient describing pore pressure changes in response to a change in applied stress.

2.1.2 The Rate-and-state Framework

As the Coulomb stress hypothesis alone does not account for the time dependency of seismicity, Coulomb rate-and-state (CRS) forecast models couple the coseismic static stress change calculations with rate-and-state friction constitutive laws to estimate the expected rates of earthquake occurrence. Although more recent reviews of the rate-and-state model have been proposed (*e.g. Heimisson & Segall*, 2018), the approach presented in this study implements the standard formulation by *Dieterich* (1994). According to the latter, the spatiotemporal seismicity rate evolves as:

$$R(t,x,y) = \frac{r_0(x,y)}{\gamma(t)\dot{\tau}}, \tag{2.2}$$

where r_0 represents the background seismicity rate in space (x, y), $\dot{\tau}$ is the secular shear stressing rate (that is, it is assumed to remain constant) and γ is a variable that under stable conditions reaches the steady state with a value given by:

$$\gamma_0 = \frac{1}{\dot{\tau}}. \tag{2.3}$$

In the absence of stress perturbations, the seismicity rate R equals the background rate r_0. When a stress step is applied to the population of receiver faults, the state variable instantaneously assumes a new value:

$$\gamma_n = \gamma_{n-1} \exp\left(\frac{-\Delta S}{A\sigma}\right), \tag{2.4}$$

where $A\sigma$ expresses the effective normal stress acting on the receiver fault, ΔS is the stress imparted by the earthquake, and γ_{n-1} and γ_n represent the values of the γ variable before and after the stress change, respectively. While in the *Dieterich* (1994) formulation the applied stress change is the shear stress change, CRS modelling usually assumes it to be a "modified" Coulomb stress change (*Dieterich et al.*, 2000) that also includes the contribution of the effective normal stress changes. This is achieved by considering $S = \tau - (\sigma - \alpha)(1 - B_k)\sigma$, with α a positive non-dimensional constitutive parameter controlling perturbations in normal stress (*Linker & Dieterich*, 1992). To approximate S in equation (2.4) to the Coulomb stress change as traditionally defined in equation (2.1), CRS models assume that $\mu' = (\sigma - \alpha)(1 - B_k)$.
Dieterich (1994) and *Dieterich et al.* (2000) show that the state variable evolves as:

$$d\gamma = \frac{1}{A\sigma}[dt - \gamma dS], \tag{2.5}$$

Following equation (2.4), a positive stress change causes a drop of the γ value and

consequently a higher earthquake rate according to equation (2.2). However, the seismicity rate eventually recovers as the state variable evolves in time according to *Dieterich* (1994):

$$\gamma_{n+1} = \left(\gamma_n - \frac{1}{\dot{\tau}}\right) \exp\left(\frac{-\Delta t \dot{\tau}}{A\sigma}\right) + \frac{1}{\dot{\tau}}, \tag{2.6}$$

where Δt is the time step.
In the Dieterich's (1994) rate-and-state framework, the ratio between the normal stress $A\sigma$ and the secular shear stressing rate $\dot{\tau}$ is the aftershock recovery time (t_a) required for the seismicity rate R to return to the background value r_0 through an Omori-like decay:

$$t_a = \frac{A\sigma}{\dot{\tau}}. \tag{2.7}$$

Given this inverse correlation between the stressing rate and the aftershock duration, it is evident that the seismicity rate on the most slowly stressed faults takes more time to decay towards background values (*Stein & Liu*, 2009).

2.1.3 Sources of Uncertainty in CRS Modelling

The propagation of epistemic and aleatory uncertainties in CRS models has been investigated by several authors (*Hainzl et al.*, 2009; *Woessner et al.*, 2012; *Cattania et al.*, 2014 among others).

First, the static stress change calculations usually do not take into account any spatial variability in the elastic properties of the crust. Given our current limitations in treating rheological variations at a fault-specific level, elastic parameters (*i.e.* Lamé parameter, shear modulus and Poisson's ratio) are commonly assumed to be homogeneous throughout the study region, with values representing the average elastic properties of the upper crustal seismogenic layer. Although this source of aleatory uncertainty likely affects our capability to resolve the small-scale patterns of coseismic stresses, it has been shown that it is not a primary factor of uncertainty when compared to other elements, such as the existence of diverse

receiver fault populations, even for high model resolutions (*Cattania et al.*, 2014). Therefor, in the absence of further constraints, using a simplified elastic medium seems acceptable from a modelling perspective.

The spatial heterogeneity of the receiver planes has been indeed reported to be the main source of epistemic uncertainty in CRS forecasts by many authors (*e.g. McCloskey et al.*, 2003; *Steacy et al.*, 2005; *Hainzl et al.*, 2010). In the literature, the receiver fault formulation usually follows two approaches: the "optimally oriented planes" (OOP; *King et al.*, 1994), that assumes that earthquakes nucleate on hypothetical faults in favourable orientation with respect to the regional plus coseismic stress field, and the geological receiver plane (GRP) approach, where receivers are informed from mapped faults. Both approaches received extensive criticism in recent years; the OOP for relying on the knowledge of the largely unknown regional stress tensor to resolve stress changes on hypothetical planes that might not even exist (*Segou & Parsons*, 2016), whereas GRP may miss unmapped faults (*Jackson*, 2018), when even in well-studied regions such as California approximately 30% of seismicity occurs on previously unidentified structures (*Field et al.*, 2014).

Furthermore, the array of receiver faults in a region likely have different apparent friction coefficients (μ'), with small, limited offset faults having higher values than more evolved higher slip faults (*e.g. Parsons et al.*, 1999). However, in an operational forecast setting where it is not possible to assess the frictional state of every fault, an average intermediate value is a usually taken as a reasonable modelling approach for a broad region.

Stress-based aftershock models are also critically affected by the epistemic uncertainties behind complex source representations, that have more pronounced effects in the near-source region of a forecast. It is in fact common that many, and potentially conflicting, slip model versions for the same causative earthquake are available to modellers. Such slip models can be obtained by inverting different kind of datasets picking up the coseismic and/or post-seismic deformations. In this regard, *Cattania et al.*, 2014 show that ensemble forecasts based on sets of alternative slip distributions potentially outperform individual models.

Cocco et al. (2010) also show how these models are particularly sensitive to the definition of the background rate (r_0), specifically, to the use of a declustered (*i.e.* purely background) or non-declustered ("reference rate") precursory seismicity and to its spatial variability. Spatially variable background (or reference) seismicity rates are clearly supported by observations (*e.g. Toda & Enescu*, 2011); however, their application within CRS models is still debated, since the spatial correlation between background rates and the pattern of calculated stress changes can locally produce a forecasted seismicity rate (R) that strongly diverges from the observations.

Furthermore, it is widely accepted that faults constitutive properties used in the rate-and-state forecast implementation, namely $A\sigma$ and $\dot{\tau}$, are poorly constrained (*Cocco et al.*, 2010). While assigning predetermined values to the rate-and-state variables is one possible modelling choice (especially at local scales, where they are hardly resolvable for each single fault), fitting the parameters to a learning phase catalogue of past seismicity can represent a more reliable estimate and guarantees a more objective model parameterisation. One approach to the latter method consists in maximising a log-likelihood function (*Zhuang et al.*, 2012) that measures the goodness of fit between a model $R(x, y, t)$ and a catalogue made of N earthquakes at x_i, y_i, z_i, t_i, with $i = 1, ..., N$. For a point process, it is defined as:

$$LL = \sum_{i}^{N} \log\left(R\left(x_i, y_i, z_i, t_i\right)\right) - \int_{t_0}^{t_1} \int_{V} R(x, y, z, t) dx dy dz dt, \tag{2.8}$$

where $R(x, y, z, t)$ is the seismicity rate at the time and location of the observed events in the catalogue, while t_0 and t_1 represent the learning phase start and end time.

2.2 A Statistical Approach: the ETAS Model

The Epidemic-Type Aftershock Sequence (ETAS) model was first introduced by *Ogata* (1988) to describe the occurrence times and magnitudes of clustered seismicity and it was successively extended into the spatial domain by *Ogata* (1998).

Given their relatively simple formulations and considerable performance, different versions of ETAS models are currently employed by government agencies in several countries, including California (*Field et al.*, 2017), Italy (*Marzocchi et al.*, 2014), New Zealand (*Gerstenberger & Rhoades*, 2010), and Japan (*Omi et al.*, 2019). Previous applications of ETAS show a good performance both in retrospective tests (*Helmstetter et al.*, 2006; *Werner et al.*, 2011; *Woessner et al.*, 2011; *Marzocchi et al.*, 2012; *Cattania et al.*, 2018) and during unfolding sequences (*e.g. Marzocchi & Lombardi*, 2009; *Marzocchi et al.*, 2017). However, the model presents some weaknesses in earthquake sequences affected by considerable early catalogue incompleteness (*Omi et al.*, 2016; *Segou & Parsons*, 2016) and due to its purely statistical nature it does not include interaction effects between specific faults at short, intermediate and long off-fault distances.

The ETAS seismicity corresponds to a point process with a stochastic spatiotemporal branching evolution, where each earthquake triggers its own offspring events, whose number depends on the parent's magnitude and follow an Omori law decay. In the ETAS model, triggered earthquakes can have a larger magnitude than their parent event. The total space-time seismicity rate λ (or "conditional intensity") is defined as:

$$\lambda(x, y, t|H_t) = \mu(x, y) + \sum_{i:t_i<t} g(t - t_i, x - x_i, y - y_i; M_i), \tag{2.9}$$

where $\mu(x, y)$ is the background rate, a time-independent and spatially heterogeneous Poisson process, while the summation term represents the triggering history (H_t) from all preceding earthquakes occurring at $t_i < t$. The triggering function is expressed by empirical relations, according to the form of *Ogata* (1998):

$$g(t, x, y; M) = K_0 e^{\alpha(M-M_{cut})} \times c^{p-1}(t + c)^{-p}(p - 1) \times f(x, y|M), \tag{2.10}$$

with normalised temporal and spatial distributions as second and third terms on the right-hand-side, respectively. The parameter K_0 regulates the short-term aftershock productivity by a parent event with magnitude M equal or above a

minimum triggering magnitude (M_{cut}); α establishes the efficiency of earthquakes in triggering aftershocks as a function of magnitude. The second term on the right-hand side of equation (2.10) is the modified Omori law (Utsu, 1961) describing the temporal distribution of triggered earthquakes. The term $f(x, y|M)$ represents the probability distribution function (*pdf*) of the spatial decay of triggered seismicity near the parent event given the parent's magnitude. Although spatially anisotropic kernels have been proposed and tested in literature (*e.g. Helmstetter et al.*, 2006; *Werner et al.*, 2011; *Savran et al.*, 2020), here we adopt a spatially isotropic power law (*e.g. Ogata & Zhuang*, 2006; *Seif et al.*, 2017) as it is the standard ETAS spatial distribution and still represents a very common modelling approach in most of those OEF-implemented ETAS models that are the benchmark for our stress-based counterparts:

$$f(x, y|M) = (d\, e^{\gamma(M-M_{cut})})^{q-1}/\pi(x^2 + y^2 + d\, e^{\gamma(M_i-M_{cut})})^{-q}(q-1), \qquad (2.11)$$

where q describes how triggered events decay in space, and the term $d\, e^{\gamma(M_i-M_{cut})}$ scales the size of the M_{cut} aftershock zone with the magnitude of the parent earthquake.

2.2.1 ETAS Parameterisation

The robust estimation of input parameters has been reported in many cases to be one of the hardest challenges within the ETAS framework (*e.g. Seif et al.*, 2017; *Zhang et al.*, 2020). The most used method to estimate the ETAS parameters is the Maximum Likelihood Estimation (MLE) approach, following which the modeller obtains the set of parameters that, given the observations (*i.e.* a seismicity catalogue with N events, the "learning catalogue"), maximise the following point process log-likelihood function (*Zhuang et al.*, 2012):

$$\log L(K, c, p, \alpha, d, q, \gamma) = \sum_{i=1}^{N} \log \lambda\,(t_i, x_i, y_i|H_t) - \int_S \int_{T_0}^{T_1} \lambda\,(t, x, y)\, dtdxdy, \quad (2.12)$$

where T and S represent the selected time and space windows to fit the data,

respectively. Within the MLE criterion, a number of possible ways to reach stable global optima of the function have been proposed in the literature, including simulated annealing (*Lombardi*, 2015), expectation-maximisation (EM) algorithms (*Veen & Schoenberg*, 2008), and algorithms based on grid search methods (*Lippiello et al.*, 2014). In this work we adopt the latter approach; in particular, we apply the iterative algorithm by *Zhuang et al.* (2002) that simultaneously estimates the background rate, as the knowledge of $\mu(x, y)$ is required for the inversion of ETAS parameters.

A particular debate in the ETAS community regards whether the α value of the productivity law should be fixed a-priori or let converge to the Maximum Likelihood value. While both approaches can be found in the literature, several lines of evidence suggest that α should be set equal to β (*e.g. Helmstetter et al.*, 2005, 2006; *Hainzl et al.*, 2008; *Zhang et al.*, 2020). Among them: (1) it reproduces the Bath's law (*e.g. Felzer*, 2002); (2) when the spatial kernel is isotropic, the value of α is generally underestimated and K_0 is overestimated (*Hainzl et al.*, 2008; *Seif et al.*, 2017) and to reduce the bias in the productivity parameters a common choice is to re-estimate parameters fixing α=2.3 (assuming a b-value =1; *Helmstetter et al.*, 2006; *Hainzl et al.*, 2013); (3) α is very close to β when considering incomplete aftershock sequences in combination with a time variable background rate (*Hainzl*, 2013); (4) $\alpha \approx \beta$ would be in agreement with static stress triggering models (*Hainzl et al.*, 2010). On the other hand, setting α=β often leads to the undesired effect of obtaining a branching ratio, defined as the mean number of triggered events per earthquake averaged over all magnitudes, larger than 1. In practical terms, this means that the ETAS earthquake-generating process becomes supercritical (or 'explosive') and the triggered seismicity projected by the model never dies out. To overcome this potential issue, a popular solution (also adopted in this work) is to re-scale the estimates of the productivity parameters K_0 and α such that they return a stable process (*e.g. Seif et al.*, 2017).

2.2.2 ETAS Simulations

Forecasts of the ETAS model require simulations because the rate is conditional on the history. Here, we follow the simulation algorithm by *Zhuang and Touati*

(2015): (i) the number of unknown events is a random variable drawn from a Poisson probability mass function with parameter controlled by the productivity law, (ii) the occurrence times are sampled from the modified Omori law, (iii) the spatial locations follow the isotropic spatial kernel of equation (2.11), (iv) magnitudes are drawn from the Gutenberg-Richter distribution. For higher generations of triggered events, the algorithm is repeated until the simulation process eventually runs out of potential parent shocks.

In addition to triggered seismicity, the forecast has to account for background (*i.e.* spontaneous) events. One way to achieve this is by simulating those events assuming that (i) their number is Poisson distributed with mean equal to the rate of background events identified after declustering of the learning catalogue, (ii) their location is consistent with the smoothed spatial distribution of background seismicity in the study area, (iii) their occurrence times are sampled from a uniform distribution rather than from the Omori law, and (iv) magnitudes are drawn from the Gutenberg-Richter distribution. An alternative approach is to account analytically for the background seismicity occurring inside the forecast window. In this study we adopt the latter approach; we assign 2D Gaussian kernels with variable bandwidth around each background event identified after the stochastic declustering (*Zhuang*, 2002) of the ETAS learning catalogue. We set a minimum bandwidth value consistent with a reliable estimate of the average location error of the implemented ETAS learning phase catalogue. Finally, we integrate the contribution of all kernels over each grid cell, obtaining a background spatial density. The expected total mean rate (R_{tot}) due to the occurrence of potential background events and their related generations of triggered earthquakes in each cell is given by:

$$R_{tot} = \mu \left(\frac{1}{1-n} \right), \tag{2.13}$$

where μ is the local value of background rate and n is the branching ratio.

2.3 Performance Evaluation Metrics

2.3.1 CSEP tests based on likelihoods

A number of statistical tests (*Schorlemmer et al.*, 2007; *Zechar et al.*, 2010; *Werner et al.*, 2011; *Rhoades et al.*, 2011; *Marzocchi et al.*, 2012) are implemented within the Collaboratory for the Study of Earthquake Predictability (CSEP) in order to efficiently evaluate the spatiotemporal performance of short and long-term earthquake forecast models.
The most commonly used metrics are: (1) the modified N-test (*Zechar et al.*, 2010) to compare the total number of observed vs. forecasted earthquakes over a specific time horizon; (2) The S-test log-likelihood scores of the forecasts (*Schorlemmer et al.*, 2007; *Zechar et al.*, 2010) to measure the consistency between the observed and expected spatial distribution of the events; (3) the T-test (*Rhoades et al.*, 2011) to compare the relative performance of the models in terms of information gain per earthquake.

N-test

The N-test makes use of two metrics, under the assumption that the tested forecast is correct: δ_1 to assess the probability of observing at least N_{obs} earthquakes given a forecast of N_{fore}, and δ_2 to evaluate the probability of observing at most N_{obs} earthquakes given N_{fore}. To compute these two quantiles, the test implements a Poisson cumulative mass function F:

$$\delta_1 = 1 - F\left((N_{obs} - 1)\,|N_{fore}\right), \tag{2.14}$$

$$\delta_2 = F\left(N_{obs}|N_{fore}\right). \tag{2.15}$$

The N-test is then evaluated by applying a one-sided significance test and the forecast is "rejected" if either $\delta_1(\mathrm{t}) < \alpha$ or $\delta_2(\mathrm{t}) < \alpha$, where $\alpha = 0.025$ is the effective significance value (*Zechar et al.*, 2010).

S-test

To determine the spatial consistency of the models, their log-likelihood scores are used. Given a forecast and under the assumption that it is correct, we ask what the likelihood of the observation is in each bin of the testing region. For a single spatial cell, the log-likelihood (LL) of observing ω earthquakes given a forecast of λ events is defined as (*Zechar*, 2010):

$$LL(\omega|\lambda) = \log(Pr(\omega|\lambda)) = -\lambda + \omega \log \lambda - \log(\omega!), \tag{2.16}$$

where $Pr(\omega|\lambda)$ is the probability of observing ω assuming that λ is correct. The log-likelihoods used in the S-test (LL_S) make use of normalised λ rates in order to isolate the spatial component of the forecasts, under the assumption that the total number of expected events equals the total number of observed events. Because of the bold Poissonian assumption on the independency of the number of events in different spatial bins, the joint S-test log-likelihood scores (jLL_S) are obtained by summing the LL_S of all the (i, j) cells:

$$jLL_S(\Omega|\Lambda) = \sum_{(i,j)\in R} (-\lambda(i,j) + \omega(i,j)\log(\lambda(i,j)) - \log(\omega(i,j)!)), \tag{2.17}$$

where Ω and Λ are the observed and forecasted catalogues.
Log-likelihoods are negative by definition, with higher values (that is, values closer to zero) indicating better predictive skills.

T-test

The T-test (*Rhoades et al.*, 2011) quantifies the relative performance of competing models by ranking the forecasts according to their information gain per earthquake (IG) over a selected benchmark model. The IG score can be simply defined as the average log-likelihood difference, per earthquake, between a model A and a benchmark B:

$$IG(A, B) = \frac{jLL_A - jLL_B}{N}, \tag{2.18}$$

where N is the number of observed events. T-test's likelihoods are calculated from unnormalised rates so that both the spatial aftershock distribution and the forecasted seismicity rates influence the score. The 95% confidence interval over the mean IG are calculated from a paired Student's t-test (*Rhoades et al.*, 2011). A positive information gain per earthquake over the benchmark indicates a model performance improvement, which we deem significant if the confidence interval does not enclose zero (or if the confidence bounds of two competing forecast benchmarked over the same model do not overlap).

2.3.2 Non-likelihood based alternatives

Because of their formulations, the current likelihood-based CSEP validation metrics penalise false negative forecasts (*i.e.* bins where no earthquake is predicted but at least one occurs) much more than false positives (*i.e.* locations where seismicity is expected to increase but no event is observed). While this characteristic does not have a marked effect on "all positive" models like ETAS, it can be detrimental for models that forecast local seismic rate suppressions, such as the Coulomb stress shadows. To introduce fairer consistency and comparison tests among models rooted in different approaches, CSEP is now committed to reformulating some of the existing evaluation techniques to establish new standard procedures for future experiments. For example, *Savran et al.* (2020) proposed a strategy that overcomes the usage of standard grid-based probability maps and compares specific characteristics of competing forecasts on the basis of the similarity between synthetic catalogues simulated by the model and the observations. Importantly, transitioning from a grid-based to a simulation-based model evaluation would allow exploring the true model variability by relaxing the simplifying assumption that earthquakes are Poisson distributed in discrete space-time-magnitude bins.

A variety of other evaluation metrics have been applied to the validation of earthquake forecasts in recent years. Among these, it is worth mentioning the method applied by *Schneider et al.* (2014), who evaluated a set of forecasts for the 2010 El Mayor-Cucapah (Southern California) sequence using point process residual scores (*Clements et al.*, 2011). These residuals measure how much a model underpredicts

(high positive residuals) or overpredicts (high negative residuals) the number of target events in a bin. *Marzocchi et al.* (2012) introduced a Bayesian framework to assess the relative performance of multiple models. Instead of being based on a binary acceptance/rejection criterion, the Bayesian approach emphasises model comparison by providing the posterior probability of each competing model, that is, the probability for each forecast to be the data-generating model.

Regarding a slightly different family of earthquake models, the alarm-based forecasts, performance evaluation usually hinges on a binary framework, that is, it depends on the number of models hits ('Yes') and misses ('No'). In this context, the most popular evaluation method is the Area Skill Score (*Zechar & Jordan*, 2008), which quantifies the performance of an 'alarm function' relative to a reference model using the Molchan diagrams (*Molchan*, 1991; *Molchan & Kagan*, 1992), where the model's miss rate is plotted against the fraction of space-time occupied by an alarm. *Zechar and Zhuang* (2014) proposed a scoring metric that can be applied also to cases where it is not possible to define a miss rate and that does not require selecting a specific reference model (which can often represent a controversial choice); such method, the Parimutuel Gambling Score, is also applicable to both fully binary and probabilistic predictions.

For an overview of some of the tests listed above, including the CSEP-style scores of grid-based forecasts, the reader is referred to *Zechar* (2010) and the material available within the *Community Online Resource for Statistical Seismicity Analysis* (CORSSA) project.

Chapter 3

Improving Physics-based Aftershock Forecasts during the 2016-2017 Central Italy Earthquake Cascade

The material presented in this chapter appears in the following article:

- Mancini, S., Segou, M., Werner, M. J., and Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016-2017 Central Italy Earthquake Cascade. *J. of Geophys. Res. Solid Earth*, 124, 8626-8643. `https://doi.org/10.1029/2019JB017874`.

3.1 Introduction

The 2016-2017 Central Apennines earthquake sequence is one of the most recent examples of how damages from subsequent aftershocks can exceed those caused by the initial mainshock. Recent studies reveal that physics-based aftershock forecasts hold potential for reaching comparable skills to their statistical counterparts, but their performance remains a controversial subject.
In this study, we employ physics-based models that combine the elasto-static stress transfer with rate-and-state friction laws, and short-term statistical ETAS models

to describe the spatiotemporal evolution of the Amatrice-Visso-Norcia (hereinafter, AVN) earthquake cascade. We then track the absolute and relative model performance using CSEP's log-likelihood statistics over different time windows (and for total forecast horizon of one year) after the 24th August 2016 Mw=6.0 Amatrice earthquake. We propose a pseudo-prospective experimental framework with the goal of (1) assessing how data quality and individual model choices driven by real-time conditions affect the performance of physics-based forecasts benchmarked against a standard statistical ETAS model and (2) evaluating the comparative performance of the forecasts in critical time windows, such as the period following the 26th October Visso earthquakes leading to the 30th October Mw=6.5 Norcia mainshock.

In particular, we present 7 classes of physics-based models with gradual complexity increase. The simplest forecasts include preliminary data available a few minutes after each Mw $\geq$ 5.4 event, featuring synthetic source models with empirically derived fault length and previously determined fault constitutive parameters. More complex models incorporate: (i) optimised rate-state parameters, (ii) spatially heterogeneous receiver fault planes, (iii) best available slip models, and (iv) secondary triggering effects.

Our results show that the preliminary assumptions made to fill the early post-disaster knowledge gap severely hamper the absolute performance of CRS forecasts. When we incorporate revised data, optimise the rate-and-state parametrisation on previous regional seismicity and account for the multi-level heterogeneities as (1) 3D spatial distribution of receivers, (2) spatially variable fault slip models, and (3) smaller magnitude aftershocks that reshape the local stress field, we obtain a dramatic performance improvement of physics-based models.

Results suggest that CRS performance is comparable to ETAS only when secondary triggering is taken into account.

3.1.1 The Central Apennines Earthquake Sequence

The Central Apennines are among the highest seismic hazard zones in Europe (*Woessner et al.*, 2015). Present day deformation is taken up by NW-SE trending normal fault systems, expressing the post-orogenic extension at a rate of $\sim$3

mm/yr (*Serpelloni et al.*, 2005). Historical and modern seismicity indicate moderate (M5+) to large (M6+) magnitude earthquake cascades associated with heavy damage patterns in the broader area (*Rovida et al.*, 2016) such as the 1703 Norcia cascade where three M6.2+ earthquakes occurred within less than 20 days (*Boschi et al.*, 2000), the six M5.2+ events that struck Colfiorito between September and October 1997 (*Amato et al.*, 1998; *Chiaraluce et al.*, 2003; 2004), the 2009 Mw=6.3 L'Aquila sequence with an accelerating seismicity pattern observed few days before the mainshock (*Chiarabba et al.*, 2009; *Chiaraluce et al.*, 2011).

The 24^{th} August 2016 Amatrice earthquake activated a 60-km long normal fault system (Figure 3.1a) and was followed by an Mw=5.4 aftershock within less than one hour at the northern up-dip part of the mainshock rupture (*Chiaraluce et al.*, 2017 and references therein). Two months later, on October 26^{th}, two Mw=5.4 and Mw=5.9 earthquakes occurred further north near the village of Visso, closer to the southernmost aftershock zone of the 1997 Colfiorito sequence. The October 30^{th} Mw=6.5 Norcia event remains the largest event of the sequence to date and the strongest in Italy since the 1980 Mw=6.9 Irpinia earthquake. Few months later, on January 18^{th} 2017, four M5+ earthquakes occurred within a 4-hour window to the south of the Norcia mainshock rupture, coinciding with the northern aftershock zone of the 2009 L'Aquila sequence. The January 2017 seismicity raised immediate concerns about the Campotosto artificial dam lake, the second largest man-made lake in Europe.

In Table 3.1 we summarise the preliminary and revised source parameters of the major earthquakes of the AVN sequence.

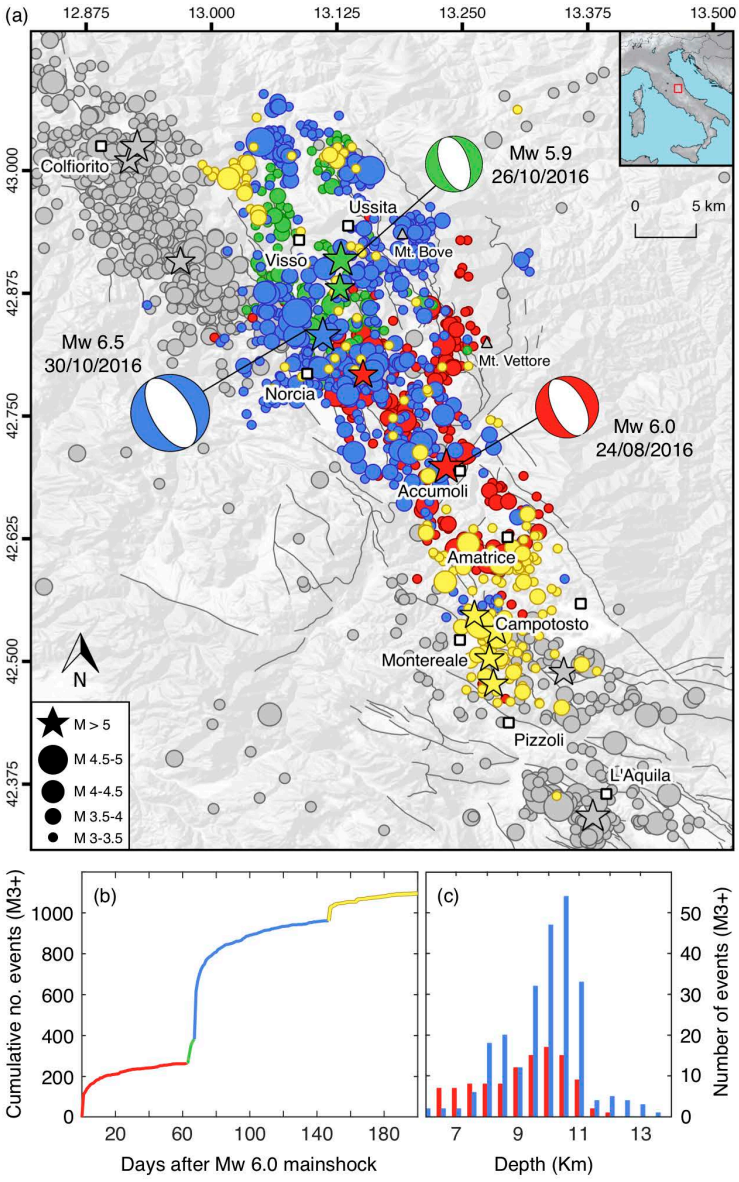

Figure 3.1: The 2016-2017 AVN sequence (M3+). (a) Map view. Colours indicate different seismicity phases: red August 24th 2016 - October 25th 2016, green October 26th - October 29th 2016, blue October 30th 2016 - January 17th 2017, yellow January 18th - August 24th 2017. The focal mechanisms of the three largest events are also displayed. Grey symbols indicate the 1997 Colfiorito and 2009 L'Aquila seismic sequences (M3+). We report the mapped active faults of the region (*EMERGEO Working Group*, 2016); (b) Cumulative seismicity (M3+) with time; (c) depth distribution of M3+ events in the first 24 hours following the Mw=6.0 Amatrice (red bars) and Mw=6.5 Norcia (blue bars) mainshocks.

Table 3.1: Source parameters for the M5+ events of the AVN sequence. Timestamps and magnitudes are available at `http://cnt.rm.ingv.it`, while the revised hypocentral depths are taken from the corresponding slip models where available. The period between the moment an event is recorded and the time when its rupture model is estimated varies from few days to several weeks. For the AVN sequence, finite-fault slip models were computed for six out of the nine M5+ events, and for most of them a period of 2-3 weeks from their occurrence passed before the release of a first robust version.

			Preliminary data		Revised data		
Event	Date	Time (UTC)	M_L	Z (Km)	M_w	Z (Km)	Slip model
Amatrice I	24/08/2016	01:36:32	6.0	4.0	6.0	8.0	Yes
Amatrice II	24/08/2016	02:33:30	5.4	9.0	5.4	8.0	Yes
Visso I	26/10/2016	17:10:37	5.4	9.0	5.4	9.0	No
Visso II	26/10/2016	19:18:07	5.9	8.0	5.9	8.0	Yes
Norcia	30/10/2016	06:40:18	6.1	10.0	6.5	9.0	Yes
Campotosto I	18/01/2017	09:25:40	5.3	9.0	5.1	9.0	No
Campotosto II	18/01/2017	10:14:09	5.4	9.0	5.5	11.0	Yes
Campotosto III	18/01/2017	10:25:23	5.3	9.0	5.4	11.0	Yes
Campotosto IV	18/01/2017	13:33:36	5.1	10.0	5.0	10.0	No

3.2 Data

We implement our models within a 3D grid (0-12 km of depth) with 2 km spacing in a ∼150 x 150 km testing region (see Appendix A, Figures A1-A2). We use: (1) the seismicity catalogue of the Italian Seismological Instrumental and Parametric Data-Base (ISIDe) for the period between January 1^{st} 1990 - August 23^{rd} 2016 (learning phase, 1533 M3+ events), (2) the near-real time catalogue for the 1 year after the Mw=6.0 Amatrice earthquake (testing phase, 1160 M3+ events), (3) Centroid Moment Tensors (CMT) catalogues for the learning phase (Figure A1) and the Database of Individual Seismogenic Sources (*Basili et al.*, 2008; *DISS*

Working Group, 2018) for constraining the active seismogenic sources and regional rupture styles of the Central Apennines.
For the primary events (Mw $\geq$ 5.4), we estimate the coseismic stress changes using the finite-fault slip models of *Tinti et al.* (2016) and *Scognamiglio et al.* (2016) for the Amatrice I-II events, respectively, and *Chiaraluce et al.* (2017) for the Visso II and Norcia events. For the January 18^{th} 2017 Mw=5.5 and Mw=5.4 Campotosto earthquakes, we use the only preliminary slip models available, as no refined versions have been issued. For the sources that lack a rupture model, we construct a synthetic slip distribution from their moment tensor solution, with empirically-derived fault dimensions (*Wells & Coppersmith*, 1994) and uniform slip (from the moment relation of *Hanks & Kanamori*, 1979) tapered at the edges.

3.3 Methods

In this section, we describe forecast characteristics and parameters. All models share a 1-year forecast horizon, with model update frequency of 24-hour time windows (dt) and M3+ target seismicity magnitude.

3.3.1 Development of CRS Forecast Models

All the Coulomb rate-and-state models presented below share some common characteristics, such as the implementation of an elastic half-space with a shear modulus of 30 GPa and Poisson's ratio ν=0.25, and the adoption of an average value of the coefficient of effective friction ($\mu' = 0.4$). However, in Appendix A.1 we also tested the effect of implementing a lower μ' in the stress change calculations.

We first develop a reference model, CRS-1, based on: (1) real-time preliminary catalogue data, including hypocentres and focal mechanisms available within few minutes to 1 hour, (2) stress perturbations from Mw $\geq$ 5.4 events estimated using a uniform slip model with kinematic parameters from focal mechanisms with nodal planes selection constrained by the predominant rupture geometries reported in the DISS database, (3) a spatially homogeneous background rate obtained by stochastic declustering (*Zhuang et al.*, 2002) of the CRS learning phase catalogue

and averaged over the testing region (r_0=0.034 M3+ events/day), and (4) spatially uniform receiver planes (SUP) for ΔCFF estimation expressing the NW-SE striking, SW dipping main trend of the Central Apennines normal fault systems (*Basili et al.*, 2008). This latter assumption is valid when we do not know enough about the complex structures of the neighbouring faults (*McCloskey et al.*, 2003). This preliminary forecast realisation features rate-and-state parameters $A\sigma = 0.04$ MPa and $\dot{\tau} = 10^{-3}$ MPa/yr ($t_a = 40$ years), previously used to characterise the active faults of the 1997 Colfiorito sequence (*Catalli et al.*, 2008).

The only difference between models CRS-1 and CRS-2 is that the latter implements a heterogeneous background rate, smoothed in space according to the adaptive kernel method proposed by *Helmstetter et al.* (2007).

From CRS-3 onwards we use: (1) revised hypocentral locations and moment tensor solutions available within 1-3 hours after a large earthquake (Table 3.1), and (2) fault constitutive parameters derived from a log-likelihood optimisation on the learning phase catalogue (see Section 2.1.3). During the optimisation procedure, the grid search for $A\sigma$ ranges between [0.01-0.1] MPa enveloping all estimated values in previous experiments (*e.g. Toda et al.*, 1998; *Console et al.*, 2006; *Catalli et al.*, 2008) with aftershock durations (t_a) of [1-1000] years. The best-fit $A\sigma$-$\dot{\tau}$ couples for each model are shown in Table 3.2. Otherwise, CRS-3 features the same implementation parameters as model CRS-2.

In model CRS-4, we introduce spatially variable receiver planes (SVP) to achieve a better representation of the structural heterogeneity of the fault system. In particular, we adopt the *Segou et al.* (2013) approach where geological receiver planes (GRP) are informed by pre-sequence focal mechanisms and active fault maps. We start by randomly selecting a nodal plane for each of the focal mechanisms included in the learning phase catalogue (Figure A1) and then we assign a rupture plane to each grid point through a 3D nearest neighbour spatial association. In regions where no previous focal mechanism exists, the assignment of a receiver fault is aided by the DISS database of active seismogenic structures. The resulting discrete fault grid is shown in Figure A2.

In CRS-5, we include the finite-fault rupture models for events with Mw $\geq$ 5.4 to implement a representation of spatial slip variability together with the structural

heterogeneity expressed by the SVP receiver matrix in CRS-4.
The CRS-6 and CRS-7 models incorporate secondary triggering effects of smaller magnitudes with different thresholds (CRS-6, 41 M4+ events; CRS-7, 1167 M3+ events). In CRS-6, we represent 35 M4+ events using uniform slip distributions (similar to CRS-1 to CRS-4) with random selection of rupture planes from Time Domain Moment Tensor (TDMT) solutions of the Italian CMT database and peer-reviewed slip models for 6 events with Mw $\geq$ 5.4. In CRS-7, we have rupture characterisations for only 4% of the M $<$ 4.0 earthquakes. Therefore, for those events we assign an isotropic distribution of coseismic stress changes (*Helmstetter et al.*, 2005; *Marsan*, 2005), following the formulation of *Chen et al.* (2013):

$$\Delta CFF = \frac{M_0}{6\pi r^3}, \tag{3.1}$$

with M_0 the seismic moment and r the hypocentral distance. While not physically realistic, the assumption of an isotropic stress field for the smaller events is reasonable in a modelling perspective; indeed, our lack of knowledge about the nodal plane parameters for most of those events would likely introduce a major source of uncertainty in the calculation of the full anisotropic stress field. Furthermore, this method is similar to the approach adopted in ETAS or in smoothed seismicity models (*Helmstetter et al.*, 2007).

We also produce two additional models, CRS-6s and CRS-7s, that isolate the sole contribution of secondary triggering (M4+ and M3+ stress sources, respectively) on model performance, where we omit the rate-and-state optimisation by implementing the same fault constitutive parameters previously derived for CRS-5. We take these two additional models into consideration during the model validation stage.

The implementation of CRS models is based on a parallel computer code for calculating seismicity induced by time dependent stress changes (*Cattania & Khalid*, 2016).

Table 3.2: Main features of CRS models. M_{min} = minimum magnitude for stress sources; USD = uniform slip distribution; FFM = finite-fault rupture model; I = isotropic stress field; SUP = spatially uniform receiver planes; SVP = spatially variable planes; Ho = homogeneous; He = heterogeneous.

Model number	Seismic Parameters	Stress Calculations					Rate-and-state Parameters			
		M_{min}	Secondary Triggering	Slip Distribution	μ'	Receivers	r_0	$A\sigma$ (MPa)	$\dot{\tau}$ (MPa/yr)	LL optimization
CRS-1	Preliminary	5.4	No	USD	0.4	SUP	Ho	0.04	0.001	No
CRS-2	Preliminary	5.4	No	USD	0.4	SUP	He	0.04	0.001	No
CRS-3	Revised	5.4	No	USD	0.4	SUP	He	0.03	0.00016	Yes
CRS-4	Revised	5.4	No	USD	0.4	SVP	He	0.02	0.00018	Yes
CRS-5	Revised	5.4	No	FFM	0.4	SVP	He	0.02	0.00018	Yes
CRS-6	Revised	4.0	Yes	FFM (M ≥ 5.4) USD (M ≥ 4.0)	0.4	SVP	He	0.03	0.00014	Yes
CRS-6s	Revised	4.0	Yes	FFM (M ≥ 5.4) USD (M ≥ 4.0)	0.4	SVP	He	0.02	0.00018	No
CRS-7	Revised	3.0	Yes	FFM (M ≥ 5.4) USD (M ≥ 4.0) I (M ≥ 3.0)	0.4	SVP	He	0.015	0.00019	Yes
CRS-7s	Revised	3.0	Yes	FFM (M ≥ 5.4) USD (M ≥ 4.0) I (M ≥ 3.0)	0.4	SVP	He	0.02	0.00018	No

3.3.2 The ETAS Reference Model

As a benchmark for the CRS forecasts, we use a standard version of the ETAS model (*Seif et al.*, 2017). Although the focus of this study is on the improvements of CRS models, we acknowledge that other ETAS parameterisations may perform differently.

To estimate the ETAS parameters we use the Maximum Likelihood Estimation (MLE) method during the ETAS learning phase corresponding to the latter part of the ISIDe catalogue (2005-2016) after the Italian permanent seismic network was considerably improved (*Schorlemmer et al.*, 2010).
In our benchmark ETAS we set $\alpha = \beta = b \cdot log(10)$ with Gutenberg-Richter b-value = 1, requiring larger magnitude earthquakes to have a higher triggering

potential than the small ones. For the estimation of the ETAS parameters we use the M3+ ETAS learning phase seismicity within a polygon covering the entire Italian mainland (Figure A3a). To account for earthquake interactions outside the spatiotemporal boundaries of the inversion, we also use events in auxiliary spatial and temporal windows (Figure A3b). The inverted parameters (Table 3.3) are in close agreement with those of *Seif et al.* (2017) for the same areal extent, time window and M_{cut}. For the ETAS simulations we set a minimum triggering magnitude (M_{cut}) of 3.0 and an M_{max}=7.5, consistent with historical seismicity and modern regional strain rates (*Rong et al.*, 2016). We fix the ETAS parameters for the whole forecast horizon, and we use them to simulate 1,000 synthetic catalogues within each forecast window (dt).

Table 3.3: ETAS parameters used for the simulations, with M_{cut}=3.0 and $\alpha = \beta = b \cdot log(10)$. We report parameter uncertainties as 1σ standard deviations.

Parameter	Value	Description
K_0	0.0936 ± 0.0085	Productivity parameter
c (days)	0.01309 ± 0.005	Omori-Utsu c-value
p	1.1842 ± 0.05	Omori-Utsu p-value
d (km^2)	1.6855 ± 0.31	Spatial kernel parameter
q	2.7572 ± 0.20	Spatial kernel parameter
γ	1.1832 ± 0.058	Aftershock zone scaling factor
Bg-rate (events/day)	0.0355	ETAS background rate (M ≥ 3.0)
n	0.97	Branching ratio

3.4 Results

In this section, we present the forecast results in the form of (a) temporal evolution of expected seismicity, (b) maps of predicted earthquake occurrence for period-specific (between mainshocks) and long-term (1 year) time windows. We then focus on the performance evaluation using the N-, S-, T-test metrics (*Zechar et*

al., 2010; *Rhoades et al.*, 2011) and comment on the absolute and relative predictive power of the models.

3.4.1 Forecast Timeseries

Figure 3.2a compares the expected and observed daily rates for M3+ events. The preliminary CRS-1 and CRS-2 models systematically underestimate the daily seismicity rates by an order of magnitude over the entire period, whereas the first updated model CRS-3 presents a 10-fold increase of expected rates arising from the optimised values of the fault constitutive variables.
Refined source parameters together with the optimised rate-and-state variables (CRS-3), the systematic introduction of SVP receiver faults (CRS-4) and heterogeneity in slip models (CRS-5) bring stress-based models closer to the observed rates, especially during the first day following each mainshock.
CRS-5 and CRS-6 share very similar expected rates in the 24 hours following each large magnitude event, standing out among all physics-based models for fitting more closely the short-term (1 day) seismicity after the Mw=6.0 Amatrice, Mw=5.4 Visso and Campotosto earthquakes. These two model implementations share the same behaviour as ETAS after the Amatrice and Norcia mainshocks. CRS-7, with M3+ secondary triggering effects, overestimates the short-term ($\leq$ 1 day) seismicity rates after the larger magnitude events (*e.g.* 376 M3+ expected events against 237 observations in the 24 hours following the Mw=6.5 mainshock); however, we note that this result is likely biased by early incompleteness of the real-time catalogue, as the M_c was larger than 3.0 for at least 6 hours after the Norcia event (Figure A5). For t $>$ 10 days, CRS-6/7 and ETAS closely match the observed rates whereas CRS-3/4/5 generally underestimate.

In Figure 3.2b we show the cumulative number of expected events vs. the cumulative real-time observations. We see that (1) ETAS captures the seismicity decay of the AVN sequence, while CRS models show a faster decay within the first few months, (2) CRS-7 is the only model that adequately forecasts the total number of observations within Poissonian error, (3) ETAS and the stress-based models from CRS-5 onwards satisfactorily estimate the seismicity after the Campotosto earthquakes.

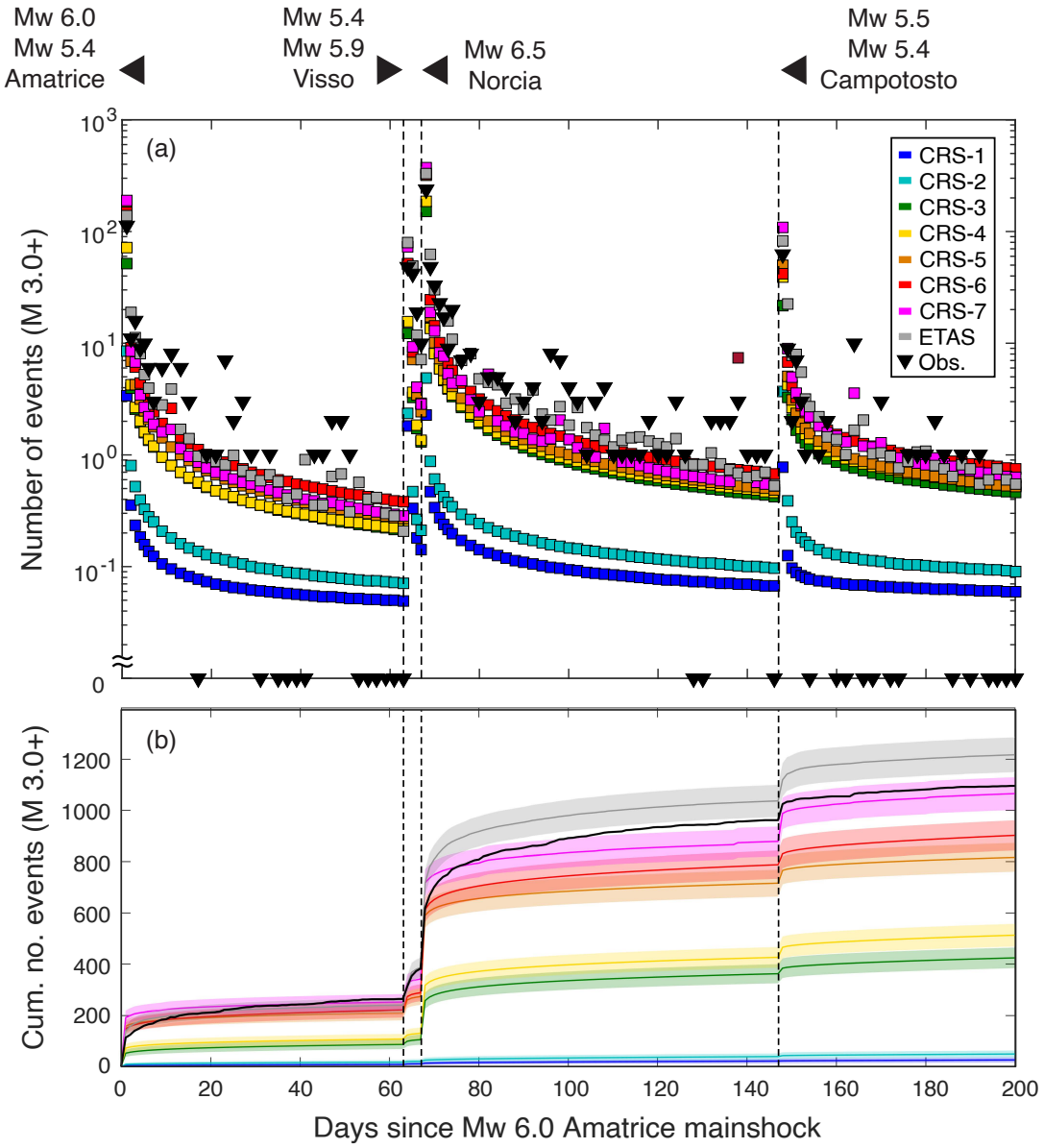

Figure 3.2: CRS and ETAS forecast time series. (a) Observed (triangles) and forecasted (squares) incremental number of M3+ events for 200 days following the Mw=6.0 Amatrice mainshock. For illustration purposes we plot values at 1-day intervals for each week after the primary events (vertical dashed lines), otherwise at 2-day intervals. (b) Cumulative expected and observed (black solid line) seismicity with shaded areas representing Poissonian uncertainties.

3.4.2 Seismicity Rate Maps

Figure 3.3 presents the forecast maps illustrating the difference in spatial distribution of expected rates between the preliminary (CRS-1) and the more elaborate physics-based models (CRS-7) against the statistical ETAS forecast. The full set of forecast maps, including first-day forecasts, is provided in Appendix A (Figures A8-A15).

Amatrice to Visso

The preliminary CRS-1 forecast presents low rates in the near-source region at the northern section of the Mt. Gorzano fault, as well as in the epicentral area of the Mw=5.4 aftershock occurred 1 hour after the Mw=6.0 mainshock (Figure 3.3a). The most advanced CRS-7 model includes SVP receivers, a spatially variable slip model for the Amatrice I/II events and secondary triggering effects from M3+ earthquakes; we observe good visual correlation between the observed and expected seismicity rates over the entire region with notable increased rates near the two epicentres and close to Amatrice and Mt. Vettore (Figure 3.3b). The ETAS model (Figure 3.3c) presents a smoother, more isotropic distribution of the higher expected rates over the entire epicentral area in comparison with CRS-7, including locations where no aftershock occurred in the first two months.
Comparing between CRS-1 and CRS-3 (Figures A8-A9, a-c), we find that optimised rate-and-state parameters introduce 1-2 orders of magnitude rate increase around the Mw=6.0 event. Comparing CRS-4 to CRS-5 (Figures A8-A9, d-e), shows that heterogeneous slip representations for the Amatrice I/II mainshocks lead to high expected rates in epicentral regions and towards Mt. Vettore. In CRS-6 (Figures A8-A9, e-f), we observe that secondary triggering effects from the 17 M4+ aftershocks do not exert an obvious difference in the spatial forecast from model CRS-5.

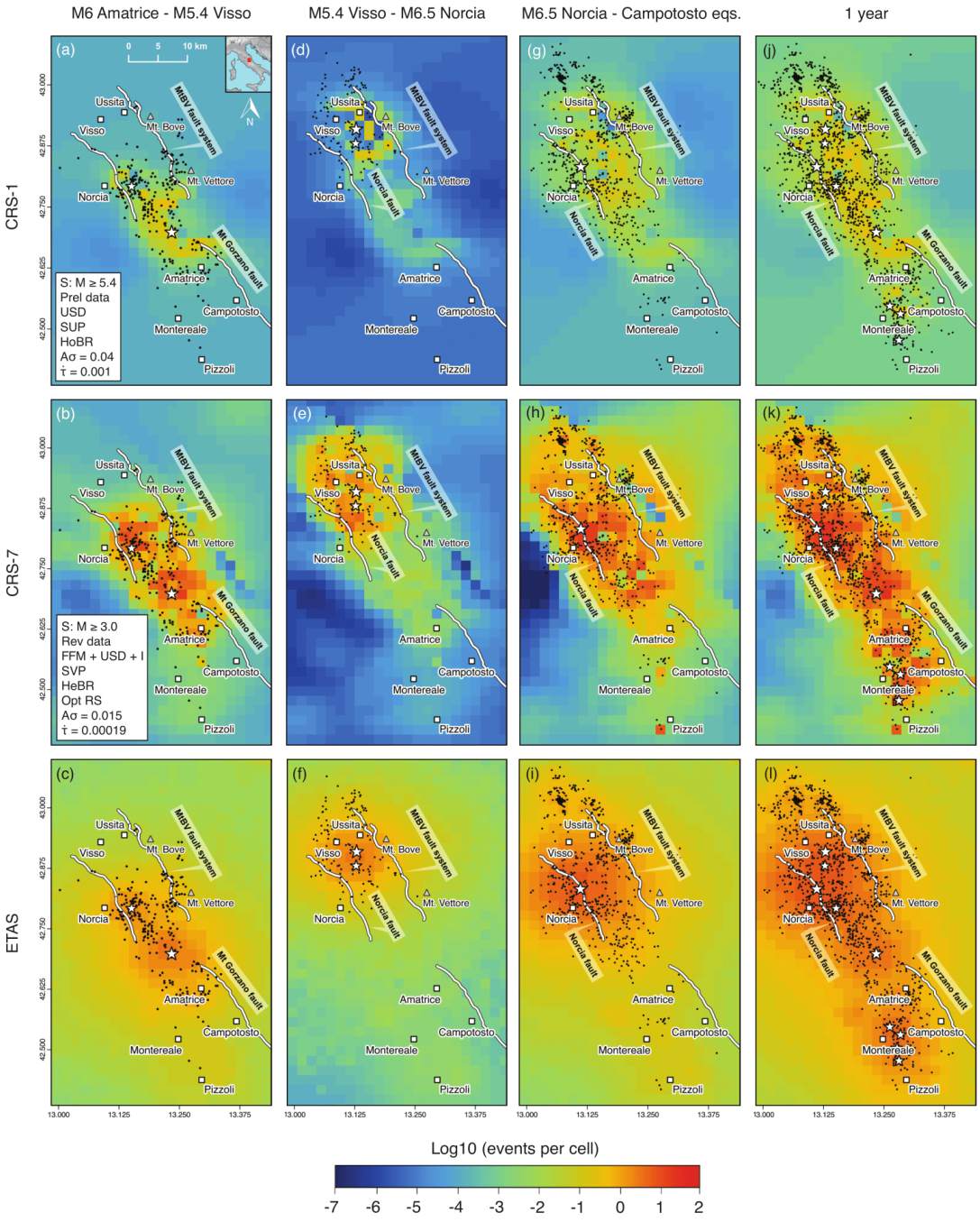

Figure 3.3: Maps of expected seismicity rates (M3+) for CRS-1/7 and ETAS at period-specific windows. We adopt a grid size of 2x2 km in all models. Black circles indicate observed events (M3+) within each specified time period, while stars indicate the primary earthquakes (not included among the target events of the respective time window). $A\sigma$ values are in MPa, $\dot{\tau}$ values are in MPa/year.

Visso to Norcia

Similarly to the previous time window, CRS-1 predicts particularly low rates immediately after and in the region of the Mw=5.4 (Visso I) and Mw=5.9 (Visso II) October events and fails to capture the aftershock activity north of Ussita (Figure 3.3d). In CRS-7 we observe a noteworthy increase of expected rates in near-source up to 5 orders of magnitudes with respect to CRS-1, and aftershock production is favoured on the northern tip of the Mt. Bove - Mt. Vettore (MtBV) fault system thanks to the spatially variable rupture model for the Visso II event and the inclusion of M3+ secondary triggering effects (Figure 3.3e). The ETAS model presents a good fit with the observed seismicity; the area of higher expected rates extends eastward to Mt. Bove, while the area of Norcia presents at least 10 times lower predicted rates compared to the Visso I/II epicentral region (Figure 3.3f).

Norcia to Campotosto

The preliminary CRS-1, in which stress perturbations are based on simplified uniform rupture distributions, fails to forecast the dramatic seismicity increase following the Mw=6.5 Norcia mainshock, presenting low rates in the near-source region and seismicity suppression on the Mt. Bove fault (Figure 3.3g). The incorporation of the complex Norcia mainshock rupture process, which involved a ∼30 km long fault plane, produces (1) near-source co-seismic stress heterogeneities, and (2) a halving of the stress shadows SW of Mt. Vettore and between Ussita and Mt. Bove (see Figure A6); this results in a systematic increase of the expected seismicity in the respective areas from CRS-5 onwards. The strength of the complex physics-based model CRS-7 is that it better captures the triggered seismicity south of Amatrice due to the incorporation of M3+ secondary triggering effects (Figure 3.3h), while M4+ effects in CRS-6 (Figure A12f) present an expected seismicity pattern similar to CRS-5 (Figure A12e). From visual comparison, we find the main differences between ETAS and CRS-7 in the high clustering region around Mt. Bove, where the former expects more than one order of magnitude higher rates than its stress-based counterpart, and in the near epicentral area, where the aftershock rates predicted by the ETAS model are approximately 2-3 times lower than CRS-7 (Figure 3.3i). This latter observation is possibly due to the early

catalogue incompleteness following the Norcia mainshock.

One-year forecast

The preliminary CRS-1 (Figure 3.3j) widely underestimates the observed aftershock production; we see an acceptable spatial distribution of the expected seismicity, but the estimated rates are systematically < 1 event/cell over the entire testing region. Model CRS-7 (Figure 3.3k) addresses this underestimation and is characterised by higher rates, particularly at intermediate off-fault distances (5-10 km from the mainshock faults) where a significant amount of smaller magnitude aftershocks occurred ($\sim$45%), contrary to the M4+ events that were mostly located at shorter ranges from the main ruptures (Figure A7). The 1-year ETAS forecast (Figure 3.3l) captures the spatial distribution of the aftershock zone accurately but due to its branching nature and the lack of a seismicity suppression mechanism, it projects a larger aftershock zone.

3.4.3 Model Performance in the Testing Region

We evaluate the absolute and relative predictive skills of the forecasts for a 1-year time period. We seek to test: (1) the forecasted vs. observed number of M3+ events using the modified N-test scores (*Zechar et al.*, 2010) and rejection ratios (R_N), (2) the spatial consistency of the models through their S-test joint log-likelihood scores (jLL_S) at short and long-term (*Schorlemmer et al.*, 2007; *Zechar et al.*, 2010), and (3) the information gain per earthquake (IG) with respect to the preliminary CRS-1 model through the T-test (*Rhoades*, 2011).
In Table 3.4 we summarise model performance.

Figure 3.4 presents the 1-day incremental N-test scores for the most preliminary (CRS-1) and updated (CRS-7) physics-based models and for the ETAS forecast, where quantiles δ_1 or δ_2 lower than the 0.025 significance level indicate model rejection due to rate under- or over-estimation, respectively. We define the N-test rejection ratio (R_N) for our models expressing the percentage of time that a given model does not pass the N-test over the selected testing period.

Table 3.4: Short-term (24 hours after the primary events) and long-term (1 year) model performance. jLL_S = S-test joint log-likelihood; $N_{F/O}$ = ratio between forecasted (F) and observed (O) number of events; IG_{CRS-1} = information gain from the preliminary CRS-1 model. Log-likelihood values are negative by definition, and smaller absolute values indicate a better model performance. We note how CRS-3 performance is severely penalised in the 24 hours period after the Norcia mainshock by its poor spatial consistency, leading to a deterioration of the information gain on CRS-1 to a negative value.

Model	Short term (24 hours)												Long term (1 year)		
	after Amatrice			after Visso			after Norcia			after Campotosto					
	jLL_S	$N_{F/O}$	IG_{CRS-1}	jLL_S	$N_{F/O}$	IG_{CRS-1}	jLL_S	$N_{F/O}$	IG_{CRS-1}	jLL_S	$N_{F/O}$	IG_{CRS-1}	jLL_S	$N_{F/O}$	IG_{CRS-1}
CRS-1	-670	0.03	/	-396	0.04	/	-989	0.01	/	-386	0.01	/	-7104	0.03	/
CRS-2	-665	0.07	0.90	-401	0.05	0.15	-1035	0.02	0.56	-378	0.06	1.65	-7170	0.05	0.58
CRS-3	-760	0.46	1.49	-382	0.26	1.98	-1865	0.64	-0.12	-381	0.34	3.09	-8721	0.41	1.11
CRS-4	-619	0.64	2.89	-303	0.32	3.80	-1614	0.79	1.00	-305	0.62	4.60	-7171	0.49	2.55
CRS-5	-557	1.40	3.62	-130	1.05	7.86	-770	1.34	4.43	-481	0.79	1.88	-6149	0.75	3.47
CRS-6	-673	1.31	2.46	-188	1.08	6.65	-970	1.37	3.69	-422	0.67	2.76	-6771	0.82	3.01
CRS-7	-418	1.69	4.60	-110	1.53	8.17	-681	1.59	4.84	-249	1.73	5.40	-4911	0.97	4.57
ETAS	-222	1.22	6.51	-118	1.65	7.96	-408	1.39	6.06	-112	1.31	7.72	-4326	1.10	5.13

The incremental N-test results for the full set of models are provided in Appendix A (Figure A16). We observe that: (1) CRS-1 is rejected due to underestimation of the observed number of events (δ_1=0) at short-term (1 day) and at t > 1 week after each mainshock (Figure 3.4a), (2) CRS-7 and ETAS present ∼50% overestimation of the number of events immediately after the Amatrice, Visso, Norcia and Campotosto events (Figure 3.4, b-c), (3) the ETAS model features the lowest R_N (6%) and while it shows a good performance in the time period following the Norcia mainshock, it suffers rejection for three days following the Campotosto events. On the other hand, CRS-7 presents a mixed performance after the Amatrice, Visso and Norcia events, with overestimation within the first 24 hours and underestimation for t ≤ 1 week, but it passes the test for t > 24h after the Campotosto seismicity.

CRS-2 performance is similar to CRS-1, as these two preliminary models feature

the highest N-test rejection ratios ($R_N \approx 30\%$) during the 1-year testing period (Figure A16, a-b). Models CRS-3/4 show better overall R_N values (15% and 14%, respectively) but underpredict ($\delta_1 < 0.025$) during the two weeks following the two largest events of the sequence (Figure A16, c-d). We find that this is most probably due to the implementation of simple uniform slip models that project negative coseismic stress changes on the MtBV fault system after the Amatrice earthquake (see Figure A8, c-d), and suppress the near-source expected rates after the Norcia mainshock with a heavy stress shadow (Figures A11-A12, c-d).
Although over-forecasting in the first 24 hours following the Amatrice and Norcia earthquakes, CRS-5 and CRS-6 are the only two models that pass the test after the Mw=5.4 Visso event and further reduce the overall R_N to 12% and 8.5%, respectively (Figure A16, e-f).

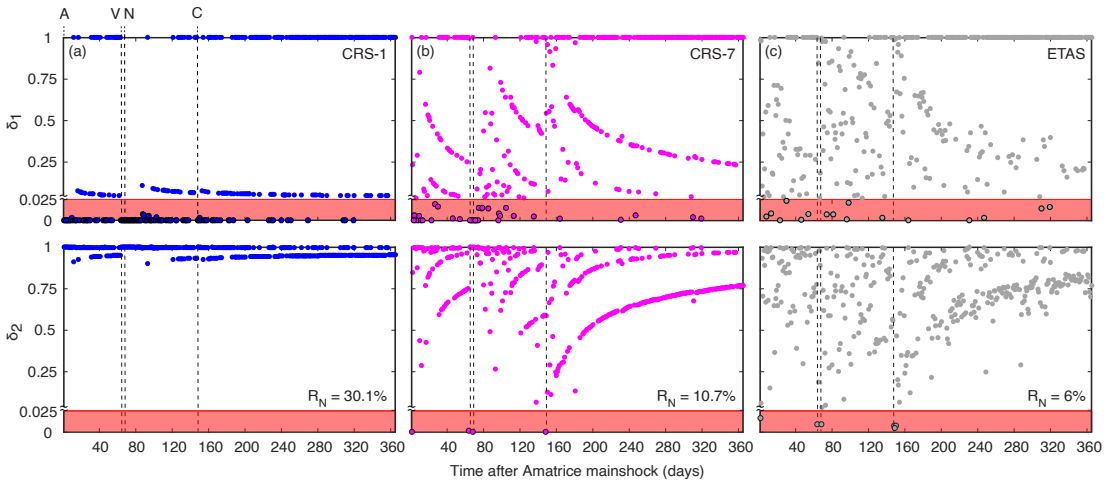

Figure 3.4: Incremental modified N-test scores. We show the δ_1 (top) and δ_2 (bottom) quantiles for CRS-1/7 and ETAS in the whole testing region, within 1-day intervals and for one year following the Mw=6.0 Amatrice mainshock. The shaded areas under the red horizontal lines indicate the values for model rejection ($\delta_{1,2} < 0.025$); we apply a vertical exaggeration to these areas to better highlight the rejection values. Vertical lines mark mainshock occurrence (A = Amatrice, V = Visso, N = Norcia, C = Campotosto events). R_N = modified N-test rejection ratio.

We use the S-test joint log-likelihood scores (jLL_S) to evaluate the predictive power of the forecasts in space. We present the earthquake forecasts expressed by the joint log-likelihood over the entire testing region, where the expected rate at each spatial cell is multiplied by N_{obs}/N_{fore}, the ratio of forecasted to observed events in the entire region, so that the normalised forecast matches the observed number of events. Higher jLL_S values indicate a better model performance (*Zechar et al.*, 2010). Results from the short-term performance test (Table 3.4) show that: (1) the ETAS model outperforms all the stress-based counterparts in the first 24 hours after the Amatrice, Norcia and Campotosto events, (2) the elaborate physics-based model CRS-7 presents the best jLL_S score following the Visso earthquakes and it constantly achieves a better short-term performance than all the other CRS models, (3) CRS-5 is the second best CRS model in short-term windows, highlighting the importance of including spatially variable slip sources for physics-based modelling; the only exception is represented by the first 24 hours following the Campotosto seismicity where two out of the four M5+ earthquakes (namely the Mw=5.1 and Mw=5.0) lack a spatially variable rupture model, while for the Mw=5.4 and Mw=5.5 events only a preliminary slip model release is available.

In Figure 3.5 we compare the evolving spatial performance of the forecasts for the 1-year testing period after the Mw=6.0 Amatrice occurrence, including the auxiliary models CRS-6s and CRS-7s that isolate the effect of secondary triggering from M4+ and M3+ sources, respectively. Results show that: (1) CRS-7/7s, including secondary triggering effects from M3+ aftershocks, are the best physics-based models, particularly when rate-and-state parameters are optimised on the learning phase seismicity (CRS-7), (2) the ETAS model has the highest spatial joint log-likelihood values over the entire testing period (jLL_S = -4326), with a striking similar performance with CRS-7 in the 4 days between the Visso-Norcia earthquakes. While models CRS-1/2/4 perform similarly throughout the evaluation period, CRS-3 is characterised by the lowest scores because of the oversimplification of coseismic slip for the Mw=6.5 Norcia mainshock. When we compare CRS-6 to CRS-6s, we note that the parameters optimisation degrades CRS-6 performance after the Norcia mainshock. The long-term scores of CRS-6s are only slightly bet-

ter than CRS-5, but a lower magnitude threshold for stress sources (CRS-7/7s) remarkably increases the short and long-term spatial consistency of the models.

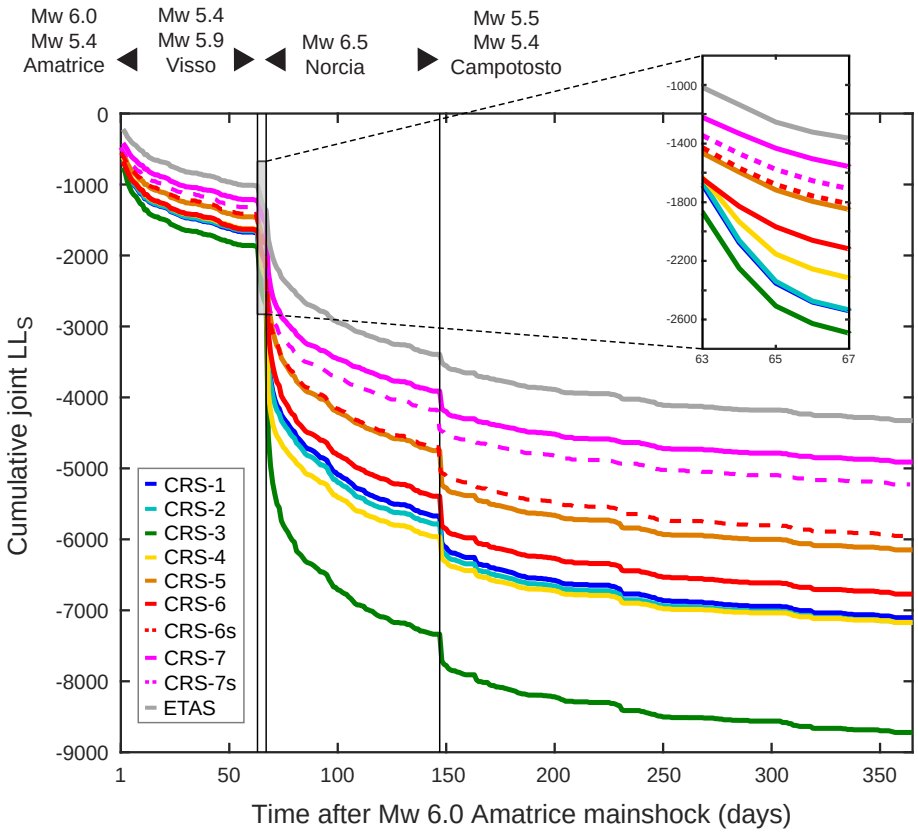

Figure 3.5: Cumulative spatial joint log-likelihood (jLL_S) versus time. The scores are obtained by summing the S-test log-likelihoods (LL_S) of each spatial cell and 1-day time step. Vertical lines mark the occurrence of the largest magnitude events.

We estimate the information gains per earthquake using the preliminary CRS-1 as benchmark model (IG_{CRS-1}). A positive IG score indicates that a model is more informative than the selected benchmark. In Figure 3.6, we present the results in the form of average daily information gains per earthquake and their 95% confidence bounds from a paired Student's t-test (*Rhoades et al.*, 2011) for CRS and ETAS forecasts. From the analysis of the IG ranks, we note that CRS-7 is systematically the best performing among all CRS classes at all time periods (Figure 3.6

and Table 3.4). The majority of CRS forecasts show positive information gains ($IG_{CRS-1} \geq 2$), indicating a significant improvement on the preliminary CRS-1 model. For the critical $\sim$ 4-day time window before the Norcia earthquake, CRS-7 and ETAS models present a comparable information gain ($\Delta IG_{CRS-1} < 0.3$) with respect to the reference CRS-1, and CRS classes 5/6/7 show a considerably good performance ($IG_{CRS-1} \approx$ 6-7, Figure 3.6b). Especially in the first day after the Visso events, we find that CRS-7 outperforms the ETAS forecast (Table 3.4); the observed poorer performance of the statistical model is most likely caused by the lack of M3+ precursory seismicity detected by the real-time catalogue in the near-epicentral region of the first Mw=5.4 Visso event (*Chiaraluce et al.*, 2017).

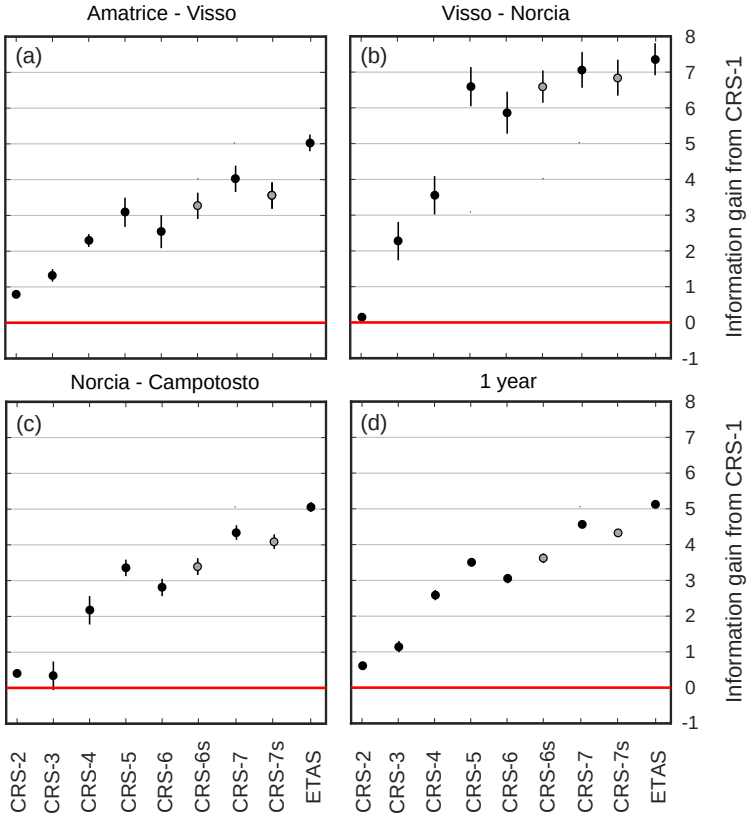

Figure 3.6: Average daily information gain per earthquake from the preliminary CRS-1 model for period-specific windows and for a 1-year forecast horizon. A forecast is significantly better than CRS-1 at 95% confidence if the error bars do not intersect the no gain level (red horizontal lines).

The 1-year performance evaluation reveals that the most advanced physics-based implementation, CRS-7, compares closely with the ETAS model ($\Delta IG_{CRS-1} \approx 0.5$, Figure 3.6d). The trend of improving information gain from CRS-5 to CRS-6s/7s shows the importance behind considering stress changes from lower magnitude aftershocks. These results suggest that including SVP receivers, finite-fault rupture models and secondary triggering from M3+ sources in CRS models leads to a gradual information gain increase.

We assess the spatial component behind the models' performance by information gain maps for a 1-year time horizon (Figure 3.7), while period-specific maps are provided in Appendix A (Figures A17-A19). We calculate the cumulative log-likelihood differences (ΔLL) at each grid point using CRS-1 (Figure 3.7, a-f) and ETAS (Figure 3.7, g-i) as reference models. In Figure 3.7, positive ΔLL values (green) highlight improved performance with respect to the reference model. Results show that: (1) variable-slip source representations lead to a sensible information gain increase in the Norcia-Visso near epicentral region (ΔLL up to 100) and on the Mt Bove-Vettore (MtBV) fault system ($\Delta LL > 60$; Figure 3.7d); (2) CRS-7 better captures the unfolding sequence, as considering the contribution of M3+ sources into the evolving stress field entails a net performance improvement especially at the extreme edges of the fault system ($\Delta LL > 250$ north of Visso and $\Delta LL > 50$ in the Montereale-Pizzoli region; Figure 3.7f); (3) updated source parameters description of mainshock faults through optimised rate-and-state parameters improve the forecast in many locations, but reduce the likelihood in a broader aftershock region (Figure 3.7b), particularly at Mt. Bove location (ΔLL $<$ -100); (4) model performance at the north-western MtBV fault termination greatly benefits from the combined inclusion of SVP receiver planes (Figure 3.7c) and, to a greater extent, from finite-fault slip models describing the Visso and Norcia coseismic slip (Figure 3.7d).

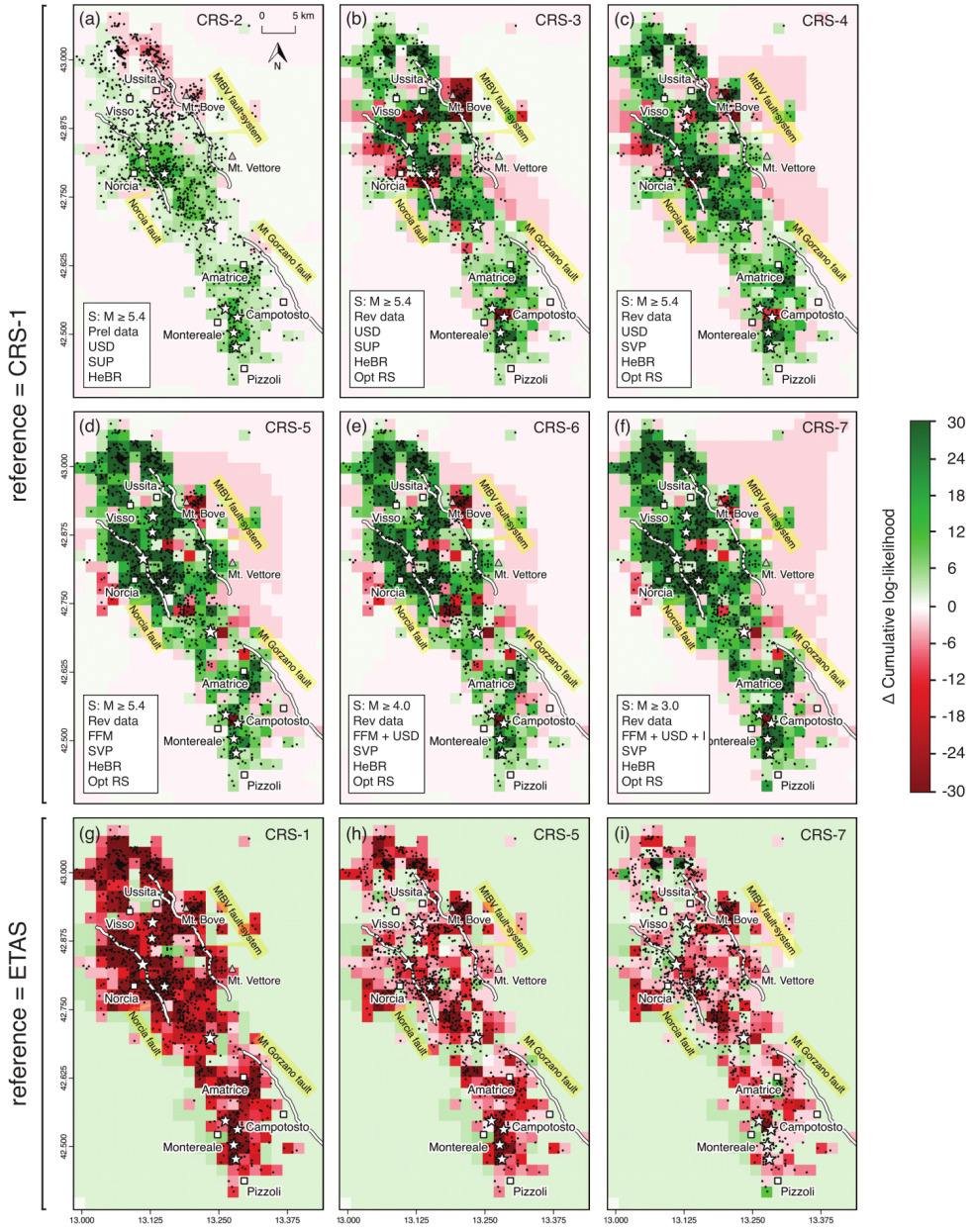

Figure 3.7: Spatially resolved cumulative log-likelihood differences (ΔLL) between pairs of models for a 1-year evaluation period. Maps show a subset of the testing region that includes the observed seismicity and the mainshock faults. As we do not need to isolate the spatial component of the forecasts, here we calculate the LL from unnormalised rates. LL values at each spatial bin are obtained summing over all time steps. Positive (green) values indicate improved performance with respect to the reference model specified to the left. Black dots represent the M3+ observations while the stars indicate the M5+ events. Values are saturated at ± 30 for visualisation purposes.

When we compare physics-based forecasts to ETAS we observe that, although CRS models generally present a slightly positive information gain in the off-fault areas of the testing region ($\Delta LL \approx 0.3$), the log-likelihood differences between pairs of models are more marked where aftershocks were actually recorded. We find that ETAS widely outperforms its physics-based counterparts in the near-source regions (Figure 3.7, g-i), where slip models present large variability (*Scognamiglio et al.*, 2016; *Chiaraluce et al.*, 2017). However, we find that secondary triggering effects significantly increase the predictive power of CRS models that outperforms ETAS in the high clustering zone north of Visso-Ussita and in the Montereale-Pizzoli region (Figure 3.7i).
The model parameters that exert most influence on the predictive power of physics-based models are the spatially variable slip distributions and the inclusion of the secondary triggering effects from small magnitude earthquake (M3+). The most evolved CRS model captures the triggered seismicity at the fault edges, therefore improving the forecast in the case of the Campotosto cluster further south than Amatrice.

In Figure 3.8 we isolate the ΔLL in the critical period between the Visso and Norcia events, and we compare CRS-7 performance to CRS-1 and ETAS. We notice that CRS-7 is a significantly more informative model than CRS-1 in the time window that precedes the Mw=6.5 mainshock (Figure 3.8a). Results also confirm the comparable performance between ETAS and CRS-7; the former better captures the near epicentral seismicity in the area of the Visso events, while the latter is better performing in locations of high aftershock production at fault extremities.

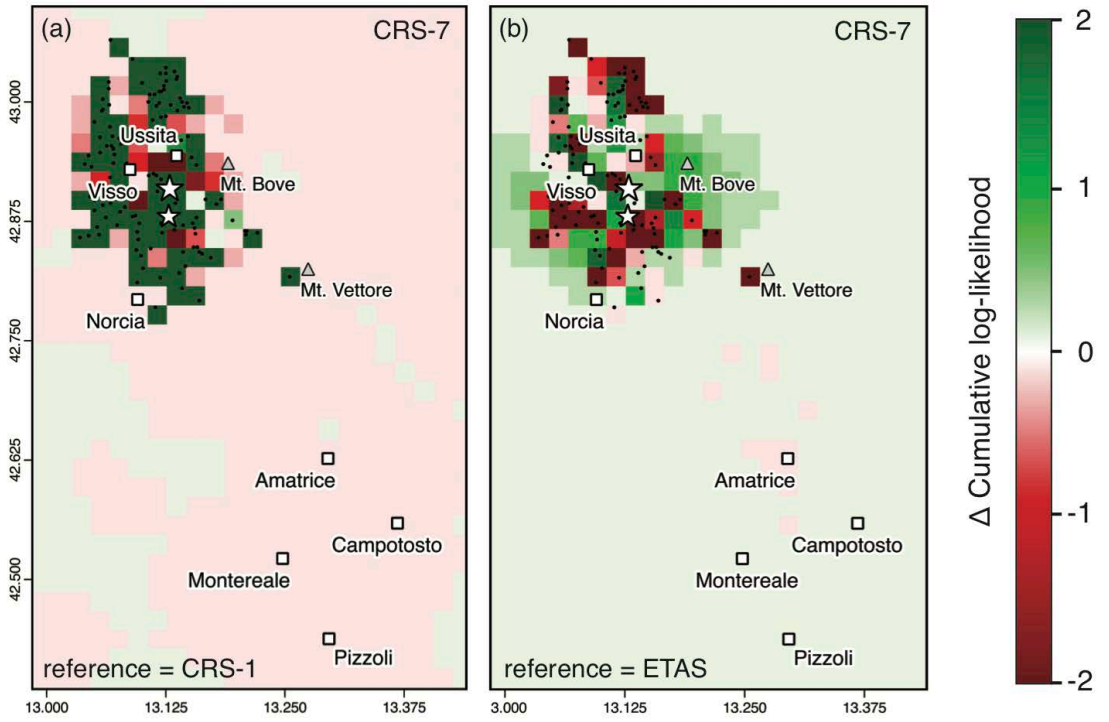

Figure 3.8: Maps of cumulative log-likelihood difference for the 4-day period between the Visso events and the Norcia mainshock between: (a) CRS-7 and CRS-1, (b) CRS-7 and ETAS. Positive values indicate a better performance of CRS-7. Values are saturated at ±20 for clarity.

3.5 Conclusions

The complex 2016-2017 Amatrice-Visso-Norcia (AVN) earthquake sequence in the Central Apennines represents a unique opportunity to test earthquake triggering hypotheses that are the basis for physics-based forecast models. Here we focus on how input data quality and model parameters influence the predictive power of physics-based forecast models. We design a pseudo-prospective experiment for the first year of the AVN sequence. We implement a benchmark statistical ETAS model and 7 classes of physics-based forecasts with gradually increasing level of input data quality and complexity. We then evaluate the absolute and relative model performance by means of the N (number), S (space) and T-test metrics

that are currently implemented within the CSEP initiative.

We find that the representation of the crustal and stress-field heterogeneity increases the information gains of physics-based models. The CRS model that predicts closely and in cases outperforms the statistical ETAS model combines best-available source models, spatially-varying receivers informed by pre-existing focal mechanisms and geological maps, and most importantly includes secondary triggering effects from M3+ events that enable us to describe the evolving coseismic stress field in greater detail. After the Mw=5.4 Visso event, the elaborate CRS-7 model outperforms ETAS, while in the critical 4-day time window before the Mw=6.5 Norcia mainshock and within the first year of the sequence, CRS-7 is as informative as ETAS.

The results support previous findings that in the near-source area, defined by the surface fault projection, ETAS models present higher predictive power (*e.g. Segou et al.*, 2013), although the lack of precursory seismicity (*e.g.* before the Visso events) locally hampers their short-term performance. On the other hand, triggered seismicity at intermediate off-fault distances due to static stresses within an evolving sequence is better described by physics-based approaches. The preliminary unrevised source parameters and empirical source models released within a few minutes to hours following large events result in poor performance of stress-based models that by definition require a more demanding physical parametrisation than their empirical counterparts. The comparative evaluation of the spatial consistency of CRS models suggests that the over-simplified uniform slip models implemented immediately after a large earthquake cannot adequately reproduce the early aftershock spatial distribution (< 1 day).

The Italian experiment reveals that in order for stress-based forecasts to reach and outperform ETAS models we need to use best-available data within heterogeneous fault and source representations. Although high-quality relocated catalogues are not yet readily available during the early stages of a seismic crisis, both CRS and ETAS models in our study will improve their predictive skills with enhanced detection and event characterisation techniques.

Data Source

The earthquake catalogue for the AVN sequence with event locations and magnitude chronologies was acquired through access to `http://cnt.rm.ingv.it`, while the moment tensor solutions from `http://cnt.rm.ingv.it/tdmt`. The Italian CMT dataset is available at `http://rcmt2.bo.ingv.it/Italydataset.html`. The Database of Individual Seismogenic Sources for Italy and surrounding areas is accessible at `http://diss.rm.ingv.it/diss`.

Chapter 4

The Predictive Skills of Elastic Coulomb Rate-and-State Aftershock Forecasts during the 2019 Ridgecrest, California, Earthquake Sequence

The material presented in this chapter appears in the following article:

- Mancini, S., Segou, M., Werner, M. J., and Parsons, T. (2020). The Predictive Skills of Elastic Coulomb Rate-and-State Aftershock Forecasts during the 2019 Ridgecrest, California, Earthquake Sequence. *Bull. Seismol. Soc. Am.*, 110(4), 1736-1751. `https://doi.org/10.1785/0120200028`.

4.1 Introduction

On 4 July 2019, a Mw 4.0 earthquake occurred in the Searles Valley (Southern California) and was followed within ∼30 minutes by a Mw 6.4 event. Just 34 hours later, on 6 July, a Mw 7.1 earthquake struck near the town of Ridgecrest approximately 10 km NE of the Mw 6.4 epicentre.
The 2019 Ridgecrest earthquakes belong to the Eastern California Shear Zone

(ECSZ), where large magnitude seismicity had not been observed since the 1999 Mw 7.1 Hector Mine event. The two earthquakes nucleated on a system of orthogonal strike slip faults (Figure 4.1): northeast-trending left-lateral for the Mw 6.4 event and northwest-trending right-lateral for the Mw 7.1 mainshock. The activated area is located in the vicinity of the Airport Lake and Little Lake fault zones, characterised by distributed faulting with mainly right-lateral strike slip and normal kinematics (*Bryant*, 2017). The resulting cascade of aftershocks involved several subparallel faults that cumulatively exceeded 75 km in length (*Barnhart et al.*, 2019; *Ross et al.*, 2019; *Chen et al.*, 2020).

The Ridgecrest area has previously experienced moderate magnitude earthquakes, including the 1982 Mw 5.2 Indian Wells Valley event and the 1995-1996 sequence with three M5+ shocks, the first two of which occurred five weeks apart. The 1995 earthquakes exhibited similar complexity to the 2019 events, with triggered seismicity on normal and strike-slip northwest and northeast trending faults (*Hauksson et al.*, 1995).

The tectonic setting of the epicentral region, bounded by the Garlock system to the south and extending towards the Owens Valley fault to the north (Figure 4.1), where moderate to large magnitude earthquakes occurred in the last $\sim$150 years (including the 1872 M$\approx$7.5 Owens Valley earthquake), immediately raised severe concerns on whether the occurrence of the July 2019 events could promote the nucleation of large events on nearby faults, as previously observed in the ECSZ (*e.g.* the 1992 Landers sequence).

This makes the 2019 Ridgecrest sequence a unique opportunity to further test the performance of CRS forecasts and advance our understanding of model features that improve short-term aftershock forecasts in high-hazard settings with complex rupture patterns and diverse population of triggered seismicity.

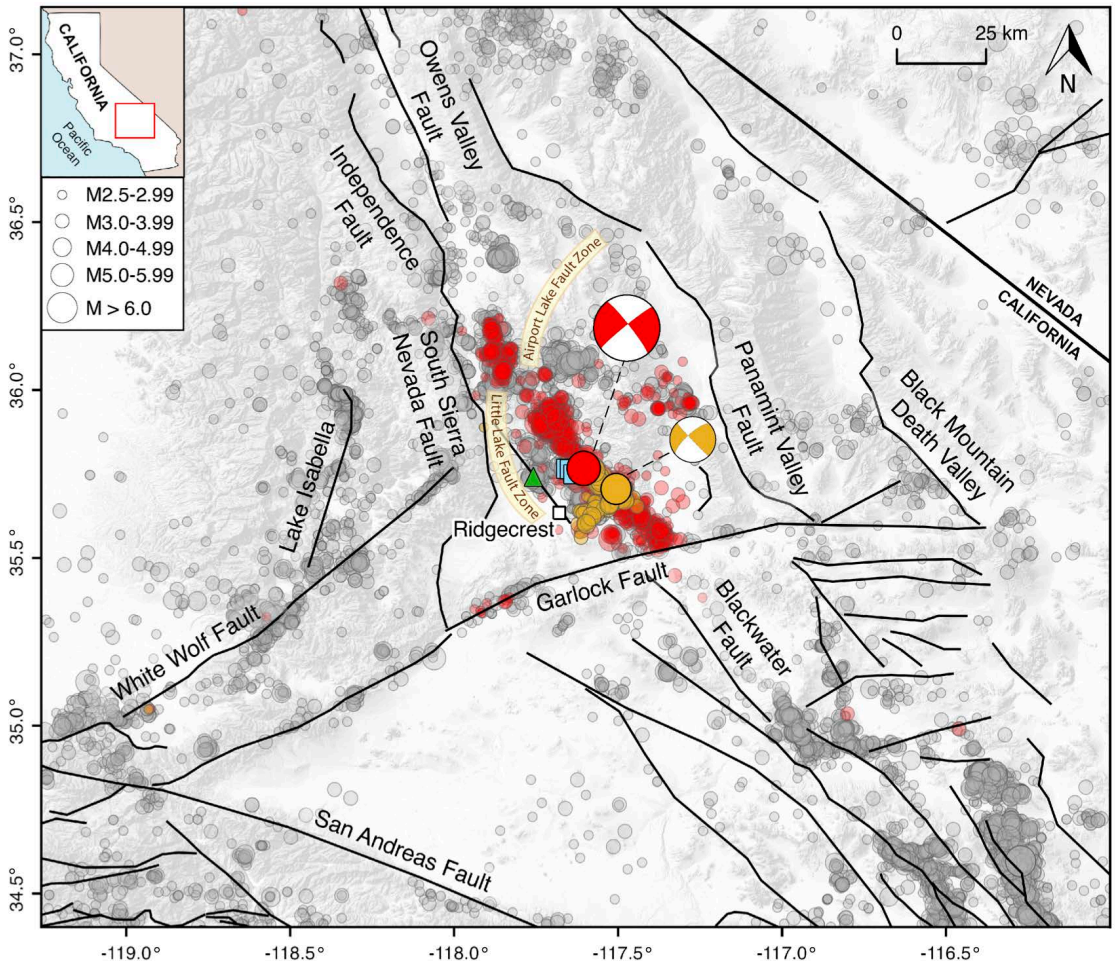

Figure 4.1: Testing region map. Earthquakes with M2.5+ are shown: pre-Ridgecrest (1981-2019, grey circles), post Mw 6.4 Searles Valley event (orange), post Mw 7.1 Ridgecrest mainshock (red). We report the focal mechanisms of the two mainshocks. The 1 October 1982 Indian Wells event (M=5.2) is indicated as a green triangle. Light blue squares represent the epicentres of the 1995-1996 Ridgecrest sequence mainshocks (M=5.4, 17 August 1995; M=5.8, 20 September 1995; M=5.2, 7 January 1996). The 1995 Ridgecrest sequence activated a number of normal, left and right-lateral faults. Black solid lines indicate the UCERF3 (*Dawson*, 2013) fault traces.

Here we perform a pseudo-prospective forecast experiment, where the issued models follow approximately the evolution of near real-time data availability and quality during the first month of the sequence.
We develop seven Coulomb rate-and-state models. Our model parametrisation supports a gradually increasing complexity; we start from a preliminary model implementation with simplified slip distributions and spatially homogeneous receiver faults to reach an enhanced one featuring optimised fault constitutive parameters, finite-fault slip models, secondary triggering effects, and spatially heterogenous planes informed by pre-existing ruptures.
In the era of machine learning catalogues that promise future improvements in real-time earthquake detection (*e.g. Ross et al.*, 2018; *Mousavi et al.*, 2019), the data-rich environment of Southern California allows us to test whether incorporating data collected in near real-time during an unfolding earthquake sequence boosts our predictive power. Hence, we extend the work presented in Chapter 3 by testing if the predictive power of Coulomb rate-and-state forecasts increases when we update fault models using evolving aftershock data.
Focusing on their overall predictive power and their spatial consistency, we present the absolute and relative performance of the forecasts for the first month following the Mw 6.4 Searles Valley earthquake, making use of the well-established evaluation metrics introduced by CSEP. To benchmark the CRS models, we produce a basic realisation of a statistical ETAS model and perform a comparative model evaluation for the short (24 hours after the two Ridgecrest mainshocks) and intermediate-term (one month).

Our stress-based forecasts expect heightened rates along the whole near-fault region and increased expected seismicity rates in Central Garlock Fault. The comparative model evaluation supports that faulting heterogeneities coupled with secondary triggering effects are the most critical success components behind physics-based forecasts, but also underlines the importance of model updates incorporating near real-time available aftershock data reaching better performance than standard ETAS. We explore the physical basis behind our results by investigating the localised shut down of some pre-existing faulting styles in the Ridgecrest near-source area.

4.2 Data

We develop the earthquake forecasts on a testing region (Figure 4.1) centred on the Mw 7.1 mainshock epicentre and extending equally E-W and N-S for ∼160 km (three time its rupture length). We discretise the models using a three-dimensional grid with 2 km spacing between 0-28 km of depth. Our target seismicity is the real time catalogue of 1812 M2.5+ aftershocks reported in the USGS Advanced National Seismic System Comprehensive Catalog (ComCat; *Guy et al.*, 2015) for the month following the Mw 6.4 Searles Valley event (testing phase, July 4^{th} - August 4^{th} 2019). As a pre-Ridgecrest learning phase (to calibrate models) we consider the seismicity between January 1^{st} 1981 - July 3^{rd} 2019 by merging the 1981-2018 relocated catalogue by *Hauksson et al.* (2012) with the ComCat events covering the last six months before the Ridgecrest sequence (43,986 events in total with M2.5+). While we adopt the catalogue of focal mechanisms by *Yang et al.* (2012) as evidence of the past local rupture styles in the testing region, we use the focal mechanism solutions from the Southern California Earthquake Data Center (SCEDC) as a real-time database. To constrain the faulting style on the main regional faults, we use the rupture parameters reported in the third version of the Uniform California Earthquake Rupture Forecast (UCERF3; *Dawson*, 2013), and we assign the larger scale off-fault rupture kinematics using the smoothed stress inversion from focal mechanisms and topography by *Luttrell and Smith-Konter* (2017), which is one of the available SCEC Community Stress Models (CSM).
To calculate the static coseismic stress changes after the Mw 7.1 mainshock we use the preliminary finite-fault slip model version issued on the USGS event information webpage, which provides near real-time automated source characterisation. For the Mw 6.4 event and for all those earthquakes with a focal mechanism solution, we create a synthetic uniform slip distribution within a planar surface implementing the empirical equations of *Wells and Coppersmith* (1994) to calculate the approximate fault dimensions and the relation by *Hanks and Kanamori* (1979) to estimate the amount of slip given the event magnitude.

4.3 Methods

4.3.1 Coulomb Rate-and-state Modelling

For the calculation of the Coulomb stress changes (ΔCFF) we assume an average an elastic half-space medium (*Okada*, 1992) with shear modulus of 30 GPa and Poisson's ratio =0.25 as representative values for the upper crust.
We issue successive CRS realisations by gradually introducing one or more levels of complexity in terms of model parameterisations and fault and source heterogeneities (*i.e.* each model preserves all the characteristics of the previous one, changing/introducing only those specified in its denomination). In other words, we conduct a 'pseudo-prospective' experiment where we test the effectiveness of different Coulomb rate-and-state forecasts that evolve from preliminary to progressively more elaborated parameterisations according to near-real time data availability. We update all the forecasts at time windows (dt) of 24 hours or when a M6+ event occurs (whichever comes sooner), for a total forecast horizon of 1 month.
It is worth pointing out that we do not parameterise models using the same data sample they are meant to forecast and that we estimate the next day seismicity rates out of sample. This means that models neither know nor use any next-day seismicity information to tune their components. Here, all CRS models are developed simultaneously, so that our modelling choices for the more enhanced realisations are not biased by the performance of the earlier versions. CRS model characteristics are summarised in Table 4.1.

The first CRS model, CRS1-*basic*, is the most preliminary version featuring: (1) stress changes imparted only by the Mw 6.4 and Mw 7.1 events, for which we implement a uniform slip distribution tapered at the edges of the fault from the real-time kinematic parameters provided by the SCEDC; (2) simplified receiver plane geometry, spatially uniform (SUP) and parallel to the Mw 7.1 fault which is consistent with the main regional regime; (3) spatially variable background rate (r_0) after stochastic declustering of the learning phase seismicity catalogue (1981-2019), smoothed in space using the adaptive kernel method of *Helmstetter et al.* (2007); (4) rate-and-state parameters averaged from the previous work of *Toda et*

al. (2005), who investigated the fingerprint of stress transfer by tuning parameters to Southern California seismicity during a subset (1986-2003) of our learning phase window.

Table 4.1: Main features of CRS models. M_{min} = minimum magnitude for stress sources; FM = focal mechanism; USD = uniform slip distribution; FFM = finite-fault slip model; $eFFM$ = edited finite-fault slip model; I = isotropic stress field; SUP = spatially uniform receiver planes; SVP = spatially variable planes; SSI = smoothed stress inversion; He = heterogeneous.

Model	Stress Calculations					Rate-and-state Parameters			
	M_{min}	Secondary Triggering	Slip Distribution	μ'	Receivers	r_0 (1981-2019)	$A\sigma$ (MPa)	$\dot{\tau}$ (MPa/year)	LL optimisation
CRS1-*basic*	6.4	No	USD	0.4	SUP	He	0.05	0.0018	No
CRS2-*optimised*	6.4	No	USD	0.4	SUP	He	0.015	0.00025	Yes
CRS3-*FFM/SSI*	6.4	No	FFM (M ≥ 7) USD (M ≥ 6)	0.4	SVP (UCERF3 + SSI)	He	0.02	0.00027	Yes
CRS4-*secondary*	2.5	Yes	FFM (M ≥ 7) USD (M ≥ 2.5) I (M ≥ 2.5)	0.4	SVP (UCERF3 + SSI)	He	0.02	0.00027	Yes
CRS5-*past_FMs*	2.5	Yes	FFM (M ≥ 7) USD (M ≥ 2.5) I (M ≥ 2.5)	0.4	SVP (UCERF3 + SSI + past FMs)	He	0.02	0.00027	Yes
CRS6-*eFFM*	2.5	Yes	eFFM (M ≥ 7) USD (M ≥ 2.5) I (M ≥ 2.5)	0.4	SVP (UCERF3 + SSI + past FMs)	He	0.02	0.00027	Yes
CRS7-*new_FMs*	2.5	Yes	eFFM (M ≥ 7) USD (M ≥ 2.5) I (M ≥ 2.5)	0.4	SVP (UCERF3 + SSI + updating FMs)	He	0.02	0.00027	Yes
CRS7-*usgs*	2.5	Yes	eFFM (M ≥ 7) USGS-USD (M ≥ 2.5) I (M ≥ 2.5)	0.4	SVP (UCERF3 + SSI + updating FMs)	He	0.02	0.00027	Yes

In CRS2-*optimised* we optimise the constitutive parameters during the learning phase catalogue by maximising the log-likelihood function of *Zhuang et al.* (2012). The grid search spans [0.01-0.1] MPa for $A\sigma$ and [1-300] years for t_a and includes

the stress changes due to the past M4+ events within the testing region during the 1981-2019 window. To account for early catalogue incompleteness, the fitting routine ignores the first 20 minutes after each stress-perturbing earthquake, corresponding to the best fitting Omori c-value obtained for our ETAS model (see Table 4.2).

CRS3-*FFM/SSI* introduces two features, namely the implementation of the near real-time USGS finite-fault slip model (FFM) for the Mw 7.1 mainshock (while no slip inversion was available for the Mw 6.4 event) and a first order structural heterogeneity of the receiver faults (spatially variable planes - SVP) in the form of: (1) in off-fault regions, planes informed from the smoothed stress inversion (SSI) from focal mechanisms and topography by *Luttrell and Smith-Konter* (2017), and (2) at the fault-specific scale, mapped UCERF3 fault geometries with kinematic parameters assigned following the USGS Fault Database in polygons extending ±2.5 km around the rupture traces, with the exception of the Garlock Fault System where we consider a 5 km buffer. The inclusion of UCERF3 rupture parameters allows accounting for well-known extensional normal faults, such as the Kern Canyon and Tank Canyon, oblique-normal faults such as the Independence Fault, right-lateral faults such as the Owens Valley, and left-lateral faults related with the Garlock Fault System.

To resolve the evolving coseismic stress field in greater detail, CRS4-*secondary* incorporates secondary triggering effects due to the stress changes following each M2.5+ aftershock. Except for the Mw 7.1 mainshock, which has an associated finite-fault model, we implement uniform slip distributions from the SCEDC real-time catalogue of focal mechanisms with random selection of nodal planes, and we adopt a magnitude-dependent isotropic stress field for all those events without any available rupture characterisation.

In CRS5-*past_FMs* we introduce the representation of the second order structural complexities by resolving the stress changes on the diverse small-scale receiver fault populations of the area informed from pre-sequence focal mechanisms (1981-2019). We assign kinematic rupture parameters to each grid cell hosting at least one focal mechanism following a 3D nearest neighbour association and using the focal plane provided by the *Yang et al.* (2012) catalogue.

The preliminary slip model version available on the USGS event webpage reaches

a depth of 28 km and is explicitly affected by (1) a deep (∼23 km) slip patch artefact on the north-western fault edge, and (2) an excessive fault length of about 160 km, extending well beyond the Garlock Fault to the SE and the Independence Fault to the NW. To overcome the limitations imposed by this near real-time data product, in model CRS6-*eFFM* we edit the USGS model (eFFM) by setting the rupture length according to the ShakeMap fault trace and extrapolating a vertical extension of 17 km using the empirical relations of *Wells and Coppersmith* (1994). We chose to include both results to highlight the importance of real-time data quality control for automated operational forecasting.
CRS7-*new_FMs* is the most complex among our physics-based forecasts. In this model, we make use of those fault planes that are gradually revealed by aftershocks to resolve the evolving near-source coseismic stress changes. Here, the available real-time SCEDC focal mechanism of a given aftershock replaces the receiver plane earlier assigned to the relative grid point following a criterion of proximity to the centre of its cell. We update the 3D receivers' matrix by performing such nearest neighbour re-assignment of aftershock nodal planes every time the forecast is updated.

We also present three additional sensitivity tests. First, to evaluate the effect of real-time data selection, we produce an alternative model version, CRS7-*usgs*, where we use the USGS catalogue of focal mechanisms for both the computation of the synthetic slip models and the receiver faults update instead of the SCEDC one. This choice is motivated by the discrepancies between the kinematic parameters of the mainshocks reported in real-time by the SCEDC and the USGS, with special concern about the Mw 6.4 earthquake (∼20° difference in strike).
Moreover, we assess how the overall spatial performance of our more complex CRS realisation changes when the rate-and-state r_0 value is defined by means of an undeclustered seismicity catalogue (known as "reference rate"). We perform this test in the wake of rederivations of the Dieterich's model suggesting that initial conditions for populations of seismic sources should also account for the long-term seismicity interactions (*Heimisson*, 2019). Finally, we test the effect of implementing a different coefficient of effective friction in model CRS7-*new_FMs*.

4.3.2 The ETAS Benchmark

Past and present initiatives, such as the Working Group of California Earthquake Probabilities (WGCEP), the Uniform California Earthquake Rupture Forecast (UCERF) and the Collaboratory for the Study of Earthquake Probability (CSEP) tested the predictive skills of ETAS models applied to Southern California seismicity under different implementations for short- and long-term time horizons. The application of ETAS to Californian seismicity catalogues made it possible to conclude that moderate events in California occur near locations of small earthquakes (*e.g. Werner et al.*, 2011) and that the inclusion of the triggering potential of small magnitude events improves forecast performance (*Helmstetter et al.*, 2006). Since our goal is to measure any improvement in the stress-based models, here we implement a standard ETAS version (*Seif et al.*, 2017) to be used as benchmark and we acknowledge that a better performance may be reached by other ETAS formulations.

We estimate the maximum likelihood ETAS parameters (Table 4.2) in the testing region considering a subset of the learning phase catalogue from 1983 to 2019. We use the first two years of the learning phase (1981-1982) as auxiliary seismicity to account for event interactions outside the target time window. The fitting process also considers earthquake triggering coming from outside the spatial boundaries of the target region by including the contribution of the M2.5+ seismicity occurred within the entire Southern California Seismic Network (SCSN) authoritative region. We set $M_{cut} = 2.5$ and $\alpha = \beta = b \cdot log(10)$ (with Gutenberg-Richter b-value=1) to improve ETAS' productivity forecasts (*Hainzl et al.*, 2008; *Seif et al.*, 2017; *Zhang et al.*, 2020). We keep the ETAS parameters fixed during the whole 1-month horizon and simulate 1,000 catalogues in each forecast time window (dt).

Table 4.2: ETAS parameters, with $M_{cut} = 2.5$ and $\alpha = \beta = b \cdot log(10)$. Our productivity parameter ($K_0 = 0.07$) is comparable to the one used by *Seif et al.* (2017) inverted on Southern California data between 1981-2014 ($K_0 = 0.08$).

Parameters	Value	Description
K_0	0.0709 ± 0.0007	Productivity parameter
c (days)	0.0137 ± 0.0007	Omori-Utsu c-value
p	1.1179 ± 0.003	Omori-Utsu p-value
d (km^2)	0.0089 ± 0.0004	Spatial kernel parameter
q	1.4834 ± 0.0082	Spatial kernel parameter
γ	1.8222 ± 0.0129	Aftershock zone scaling factor
Bg-rate (events/day)	0.161	ETAS background rate ($M \geq 2.5$)
n	0.897	Branching ratio

4.4 Results

Here, we present (1) the stress interaction results considering the UCERF3 faults within a ∼120 km radius from the mainshocks and (2) the physics-based and statistical forecasts expressed as expected number of events in the whole testing region within 1-day time intervals for a 1-month time horizon.

4.4.1 Coulomb Stress Interactions

As a first order picture of coseismic stress perturbations, we estimate Coulomb stress change values on (1) the surfaces of UCERF3 mapped faults and (2) on the Mw 7.1 fault plane using the geometry reported in the USGS finite-fault model. For these calculations, we implement a slightly coarser discretisation considering 5-km depth intervals between 0-25 km (Figure 4.2) and a wide range of friction coefficients (0.2-0.8). The 4 July Mw 6.4 earthquake moment tensor calculation has northwest trending right-lateral, and northeast trending left-lateral solutions. We choose to simulate the left-lateral plane based on observed deformation from

InSAR (Appendix B, Figure B1). We conclude that the 4 July Mw 6.4 shock likely triggered the 6 July Mw 7.1 earthquake based on calculated 0.08 to 0.2 MPa stress increases in the area of the Mw 7.1 hypocentre (Figure 4.2a), with the range depending on assigned friction coefficients. While failure stress at the hypocentral area of the Mw 7.1 shock was increased, the eventual rupture areas around that region had calculated stress decreases between -0.09 to -0.25 MPa (Figure 4.2a). The 4 July Mw 6.4 generally reduced stress or caused very small increases on most nearby surrounding faults (as defined by UCERF3) with the exception of the central Garlock fault, which had a more significant stress increase of 0.03 to 0.07 MPa (Figure 4.2a). It does not appear that the Mw 7.1 slip distribution was affected by these stress decreases because it shows relatively uniform slip despite the stress change variations (see Figure 4.3).

The combined stress change effects of the 4 July Mw 6.4 and 6 July Mw 7.1 earthquakes are calculated on surrounding UCERF3 faults (Figure 4.2b). The Garlock fault is the longest fault in the region and is believed by some to have the potential to host the largest earthquakes; the central segment of this fault had a maximum 0.006 - 0.338 MPa stress increase caused by the combined Mw 6.4 and Mw 7.1 Ridgecrest earthquakes. Significant (ΔCFF $\geq$ 0.01 MPa; *Harris & Simpson*, 1992; *Hardebeck et al.*, 1998) stress increases are also noted on the Southern Sierra Nevada Fault (maximum ΔCFF = 0.258 MPa), Owens Valley (0.116 MPa), Tank Canyon (0.090 MPa), Panamint Valley faults (0.048 MPa), Lake Isabella (0.042 MPa) and Blackwater (0.036 MPa).

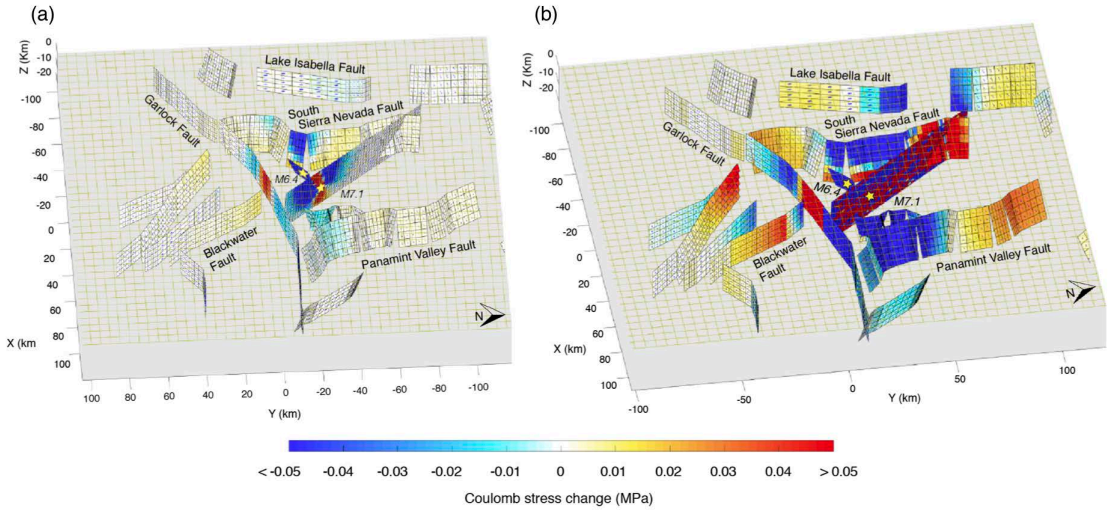

Figure 4.2: Example of calculated combined coseismic stress changes on mapped UCERF3 faults following (a) the 4 July 2019 Mw 6.4 event and (b) both the Mw 6.4 and 6 July Mw 7.1 earthquakes near Ridgecrest, CA. Hypocentres of the Mw 6.4 and Mw 7.1 earthquakes are shown by yellow stars. Displayed stress changes were calculated using a friction coefficient of 0.4. Stress increases ($\Delta CFF \geq 0.01$ MPa) are calculated on the Central Garlock, South Sierra Nevada, Owens Valley, Tank Canyon, and Panamint Valley faults.

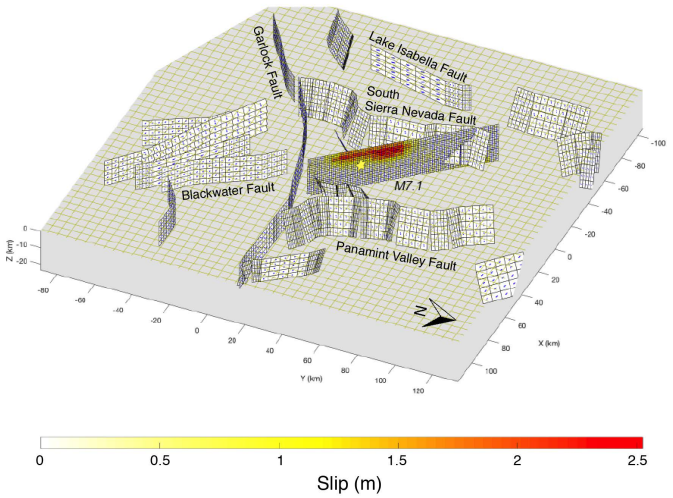

Figure 4.3: Modelled slip distribution for the 6 July Mw 7.1 earthquake. While the Mw 7.1 hypocentre correlates spatially with a calculated stress increase from the 4 July Mw 6.4 shock (Fig. 4.2a), much of the slip distribution occurred where coseismic stress changes are calculated to have been reduced.

4.4.2 Earthquake Forecasts

Here, we present the results in terms of (a) time evolution of expected seismicity, (b) spatial maps of M2.5+ expected earthquakes within specific time periods starting after the Mw 6.4 Searles Valley event, for the short (24 hours after the mainshocks), and intermediate-term (1 month) time windows. We also present the model validation for the first month of the Ridgecrest sequence using the S and T-test metrics (*Zechar et al.*, 2010; *Rhoades et al.*, 2011) implemented in the CSEP initiative, which perform a model-data consistency check and an inter-model predictive skill comparison, respectively. All forecast model results are provided in Appendix B (Figures B2-B5).

Earthquake Rate Forecasts

In Figure 4.4a we present the M2.5+ observed vs. expected daily occurrences.
The preliminary and oversimplified CRS1-basic underestimates the seismicity rates by an order of magnitude, which is in agreement with the results of the Central Apennines experiment (see Chapter 3). We observe that the introduction of optimised fault constitutive parameters in CRS2-*optimised* reverses the severe underprediction of CRS1-*basic*, making all the successive realisations comparable to the real-time catalogue. While we find that all the physics-based models match well with the number of M2.5+ events in the 24 hours following the Mw 6.4 event, they mostly overpredict in the short-term after the Mw 7.1 shock, for a maximum of 120% in CRS2 (1579 M2.5+ expected vs. 713 observed) and a minimum of 26% in CRS7 (899 expected events), with a general good agreement over the entire first month of the sequence.
The ETAS model, with rates expressed from the mean of the simulations, strongly overpredicts after both mainshocks. However, we expect the early incompleteness of the real-time catalogue following the Ridgecrest main events to affect the apparent overprediction of most of the models.

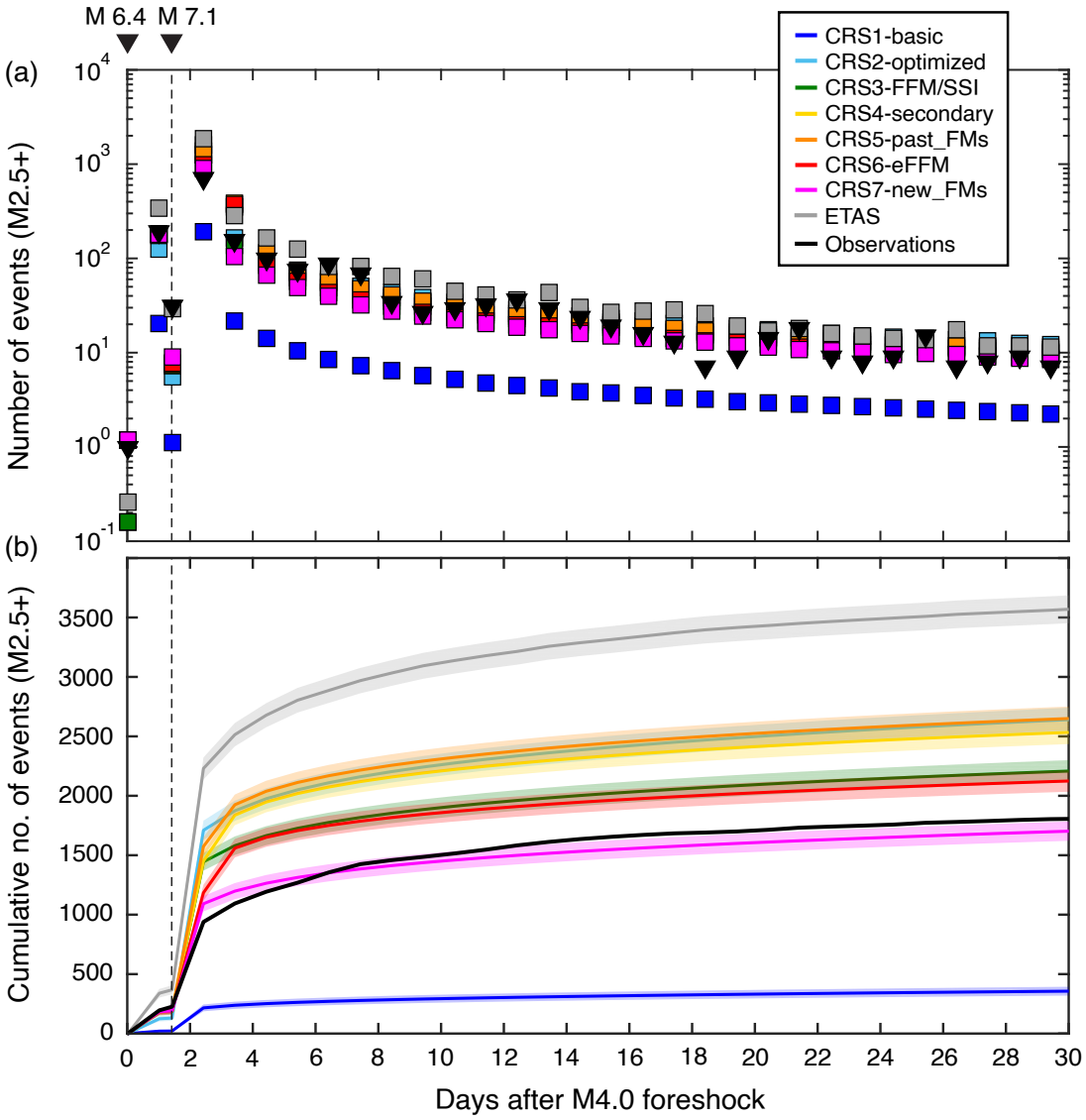

Figure 4.4: Forecast time series for physics-based and statistical models for the first month of the Ridgecrest sequence. (a) Incremental time series: black triangles indicate the observed number of M2.5+ events, while squares represent the expected numbers. (b) Comparison between the cumulative expected vs. observed (black line) rates. The shaded areas indicate Poissonian uncertainties.

Figure 4.4b compares the cumulative number of expected vs. observed earthquakes. We find that (1) the inclusion of secondary triggering effects from CRS4-*secondary* onwards leads to a 15% increase in the cumulative number of forecasted aftershocks during the first month, almost entirely due to the short-term triggering expected from the early Mw 7.1 aftershocks (days 2-3); (2) CRS6-*eFFM*, based on the edited USGS finite-fault model (eFFM), reduces the 24 hours post-mainshock over-prediction; (3) the update of the receivers using the unfolding aftershock rupture parameters (CRS7-*new_FMs*), although appearing not critical immediately after the Ridgecrest mainshock ($<$ 1 day) due to the limited number of available early focal mechanisms, brings an important improvement between days 2 and 3 by reducing the overprediction seen in models CRS4/5/6; (4) although ETAS fits the seismicity decay well in the 34-hour window between the main events, it presents the poorest performance immediately after the Mw 7.1 shock; (5) stress-based models fit adequately the seismicity decay from the third day onwards, with model CRS7-*new_FMs* better approximating the total number of events within Poissonian uncertainty.

We finally test how the implementation of a different coefficient of friction in the Coulomb calculations affects the output of the best performing model. When we use μ' values of 0.2 and 0.6, we find a variability of the expected rates after the mainshocks of about $\pm$20% (Figure B2), which reflects the importance of coseismic normal stress changes.

Forecast Maps

Figure 4.5 shows the seismicity rate maps of the most preliminary (CRS1-*basic*) and the most enhanced (CRS7-*new_FMs*) physics-based forecasts against the statistical ETAS realisation for the 24 hours following the Mw 6.4 Searles Valley and the Mw 7.1 Ridgecrest events and for the whole 1-month horizon. Although we formally assess the performance of the models for the entire testing area, in Figure 4.5 we show the sub-region characterised by the highest aftershock productivity. Similar maps for the complete set of models, including the alternative CRS7-*usgs*, are available in Appendix B (Figures B3-B5).

The expected CRS and ETAS seismicity patterns in the 24 hours following the

Mw 6.4 Searles Valley event (Figure B3) mostly miss the observed L-shaped aftershock distribution. The visual comparison between the stress-based models (Figures 4.5a,d and B3a-h) and ETAS (Figure 4.5g) shows how both forecasting methods suffer from the lack of a finite-fault model that describes the complex slip distribution along either the NE-SW left-lateral or NW-SE right-lateral fault sections involved. The most striking feature of these maps is the misalignment between the expected vs. observed seismicity along the left-lateral fault (Figure 4.5a), arising from the selection of the kinematic parameters assigned to the Mw 6.4 rupture. For the CRS models, including CRS6-*eFFM*, we initially use the real-time catalogue of focal mechanisms by the SCEDC where a strike = 69° was reported. However, the USGS strike of 48° better matches visually the distribution of triggered aftershocks along the Mw 6.4 left-lateral and also part of the right-lateral rupture before the Mw 7.1 event (CRS7-*usgs*, Figure B3h). This result highlights the critical role of real-time rupture characterisation for operational earthquake forecasting purposes, especially since the uncertainties behind fault strike angles in modern networks reach 20° (*Kagan*, 2003). Figure B3 shows that this misalignment only partially recovers when rates are enhanced by the implementation of optimised rate-and-state variables (CRS2; Figure B3b) and off-fault receiver planes are based on regional faulting styles (CRS3; Figure B3c). The lack of M2.5+ seismicity in the real-time ComCat catalogue in the 31 minutes between the Mw 4.0 foreshock and the Mw 6.4 Searles Valley earthquake results in minimal differences in the spatial distribution of expected rates between CRS3 and the remaining stress-based models.

The forecast maps for the first 24 hours after the Mw 7.1 Ridgecrest mainshock (Figure B4) show that: (1) the highly clustered seismicity at the northwestern fault edge is captured even in the preliminary CRS1-*basic* model, although the uniform slip model results in misaligned aftershock distributions (Figure 4.5b); (2) from CRS2 onwards, seismicity rates increase across the Ridgecrest fault, marking the importance of an optimised rate-and-state parameterisation; (3) the finite-fault slip model incorporation leads to high near-source rates in agreement with the distribution of early aftershocks (CRS3, Figure B4c) but also increased rates east of the South Sierra Nevada Fault: the latter likely are an artefact due to the noisy

preliminary USGS source model; (4) the early post-mainshock seismicity NW of the Coso Volcanic Field (CVF) is partially underestimated initially (Figure B4, c-d) but recovers when coseismic stresses are resolved on pre-existing ruptures taken from past focal mechanisms (Fig B3e); (5) the edited USGS finite-fault slip model improves the expected patterns east of the South Sierra Nevada Fault and reduces the overestimation in the southern CVF region; (6) the isotropic ETAS model adequately captures the triggered seismicity in the near source, but overpredicts in the off-fault region and underestimates observed rates northwest of the CVF (Figure 4.5h).

The 1-month cumulative maps (Figure B5) illustrate that: (1) the preliminary model suffers from underestimation within stress shadows and the previously described misalignment resulting from the use of the SCEDC preliminary focal mechanism (Figure 4.5c); (2) the near-source forecast improves when using the finite-fault slip model though in its preliminary non-edited version (CRS3-*FFM/SSI*, Figure B5c) while the visual comparison suggests further local improvements when secondary triggering effects are considered (CRS4-*secondary*, Figure B5d); (3) the small-scale rupture heterogeneity, represented by pre-existing ruptures taken from past focal mechanisms, provides benefits to the off-fault representation (CRS3 vs. CRS5, Figure B5c,e); (4) updating the receiver fault representation to include evolving aftershock planes presents localised differences in expected rates that become also notable on the SE fault termination near the central Garlock Fault (Figure 4.5f); (5) the ETAS model (Figure 4.5i) accurately reproduces the high observed rates in the near-source area and around the CVF but, given its basic parameterisation that does not incorporate fault information, it projects too wide an aftershock zone that leads to overprediction at intermediate distances.

The most advanced CRS7-*new_FMs* model predicts heightened rates on the northern section of South Sierra Nevada (SSN), and less heightened rates on southern Garlock, around the southern Owens Valley, Lake Isabella and White Wolf faults (Figure B6a). We do not predict important triggered seismicity on the Panamint Valley Fault, Tank Canyon and on the southern SSN section. A common output from all the physics-based forecast models is the increased expected rate along the Central Garlock Fault which is yet to be observed as of the time of writing.

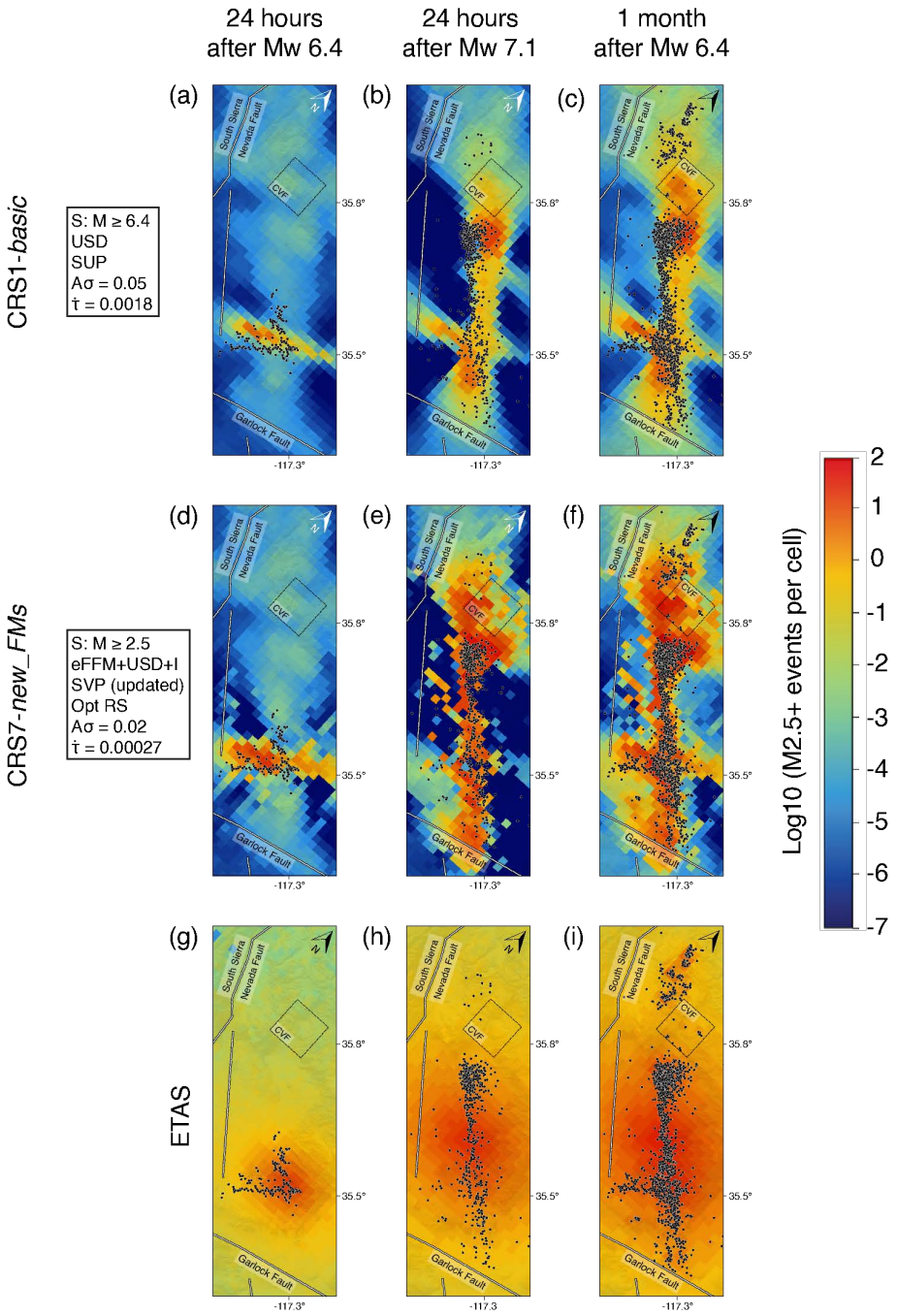

Figure 4.5: Maps of expected seismicity rates for CRS1/7 and ETAS in the area of main aftershock productivity for the first 24 hours following the two mainshocks and for the first month of the Ridgecrest sequence. Observed events (M2.5+) in each time window are represented as circles. The dashed-line square indicates the area of the Coso volcanic field (CVF). $A\sigma$ values are in MPa, $\dot{\tau}$ values are in MPa/year.

Statistical Evaluation of Model Performance

We first assess the absolute spatial performance of the forecasts by calculating the S-test joint log-likelihood scores (jLL_S; *Zechar et al.*, 2010) for the 24 hours following the two mainshocks and for the 1-month cumulative forecast horizon. We then carry out a comparative analysis of model performance through the T-test metrics (*Rhoades et al.*, 2011) describing the information gains per earthquake (IG) with respect to the simple model CRS1-*basic*.
Table 4.3 summarises the statistical scores of the physics-based and ETAS models.

Table 4.3: Summary of short-term (24 hours) and intermediate-term (1 month) model performance of CRS and ETAS models during the Ridgecrest sequence.

Model	24 hours after Mw 6.4			24 hours after Mw 7.1			1 month (cumulative)		
	jLL_S	$N_{F/O}$	IG_{CRS1}	jLL_S	$N_{F/O}$	IG_{CRS1}	jLL_S	$N_{F/O}$	IG_{CRS1}
CRS1-*basic*	-1378	0.10	-	-2905	0.27	-	-8849	0.20	-
CRS2-*optimised*	-1129	0.64	2.55	-3310	2.21	-0.42	-9451	1.46	0.28
CRS3-*FFM/SSI*	-1538	0.91	2.71	-2076	1.77	1.53	-8404	1.22	1.19
CRS4-*secondary*	-1071	0.92	2.92	-2046	1.77	1.57	-7784	1.40	1.34
CRS5-*past_FMs*	-1054	0.94	3.01	-1796	1.94	1.84	-7003	1.46	1.70
CRS6-*eFFM*	-1054	0.94	3.01	-1566	1.39	2.38	-6760	1.17	1.88
CRS7-*new_FMs*	-1054	0.94	3.01	-1482	1.26	2.55	-6440	0.94	2.25
CRS7-*usgs*	-599	0.62	5.65	-1627	1.28	3.35	-6146	0.92	2.50
ETAS	-361	1.75	6.41	-1982	2.61	1.22	-5699	1.98	2.32

We compare the ability of models to reproduce the spatial aftershock patterns by expressing the forecasts in term of cumulative joint log-likelihood (jLL_S) vs. time over the entire testing region. Figure 4.6 shows the cumulative temporal evolution of jLL_S. We find that: (1) ETAS and the most enhanced CRS7 achieve the best overall spatial consistency; (2) CRS7-*usgs* presents similar spatial performance to ETAS within the first week of the sequence, with the Mw 6.4 USGS

focal mechanism implementation increasing significantly its likelihood score (in the first 24 hours of the experiment, CRS7-$usgs_{jLL_S}$ = -599 vs. ETAS$_{jLL_S}$= -361; (3) stress based forecasts from CRS5-*past_FMs* onwards outperform the isotropic ETAS model after the Ridgecrest mainshock (Table 4.3), underscoring the importance of updating the receiver plane representation using past (CRS5) or both past and aftershock focal mechanism planes (CRS7-*new_FMs*); (4) the systematic log-likelihood increase with the growing CRS model complexity illustrates how different components (*e.g.* the Mw 7.1 slip model, secondary triggering effects, receiver updates) improve the overall model performance.

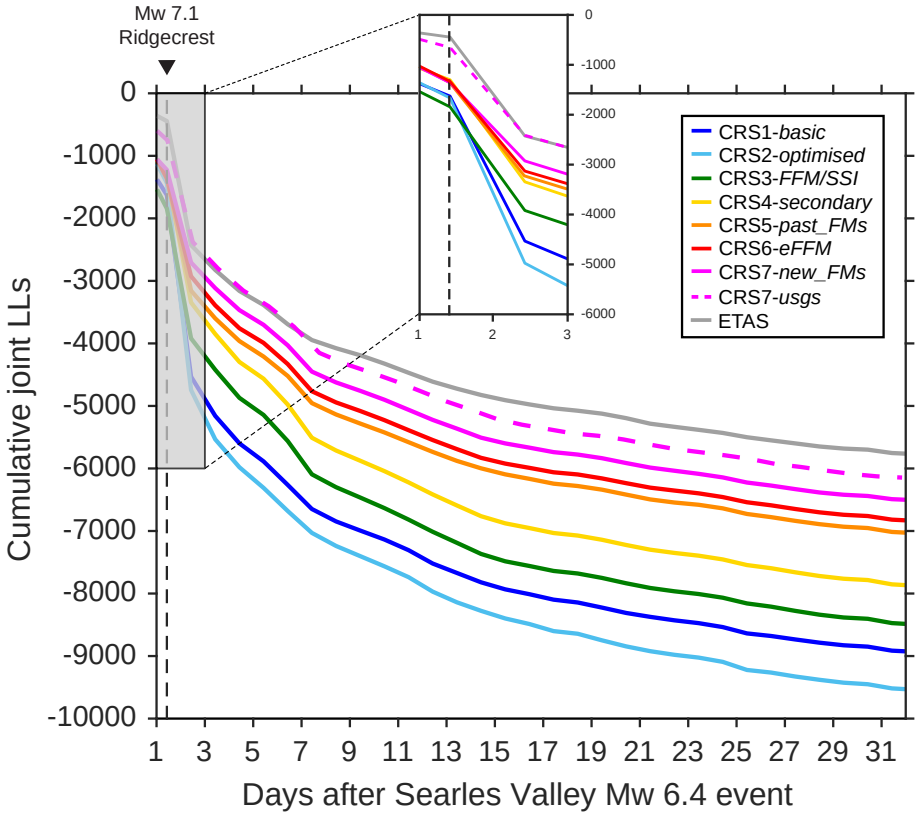

Figure 4.6: Cumulative S-test joint log-likelihood (jLL_S) timeseries. The scores are obtained by summing the S-test log-likelihoods (jLL_S) of each spatial cell and 1-day time step. The vertical dashed line marks the occurrence of the Ridgecrest mainshock.

We also find that the implementation of a reference seismicity rate in CRS7-*new_FMs* improves the joint log-likelihood in the 24 hours after the Mw 6.4 event (Figure B6c). However, the score deteriorates following the Mw 7.1 mainshock reaching a slightly worse performance in the intermediate term when compared to the model version implementing a background rate. This result is due to the higher rates projected by the model in zones of high clustering of past triggered seismicity, such as the area of the 1995 Ridgecrest aftershock sequence and the regions within and NE of the Coso field (Figure B6d); in the latter three regions, no significant clustering of M2.5+ aftershocks was observed during the first month of the 2019 sequence.

In Figure 4.7 we compare the average daily information gains when CRS1-*basic* is taken as benchmark. The short-term results for the 24 hours after the Mw 6.4 suggest that although all physics-based models are genuinely more informative than CRS1 ($IG_{CRS1} \geq 2.5$) none of them except the most enhanced one (CRS7-*usgs*, $IG_{CRS1} = 5.65 \pm 0.49$; grey square in Figure 4.7a) perform as well as ETAS ($IG_{CRS1} = 6.40 \pm 0.41$).
Following the Mw 7.1 Ridgecrest mainshock (Figure 4.7b) ETAS is outperformed by most of the stress-based forecasts as shown by the low IG values. Here, the decisive factors behind the CRS performance improvement are the edited fault slip model and the receiver updates. We also find a small overall performance improvement (ΔIG ≈ 0.15) when receiver planes are updated using the first 34 hours aftershocks (CRS7-*new_FMs*); rupture parameters for this time window are taken from the admittedly limited number of early aftershock focal mechanisms but, as we show further on, this improvement presents a significant spatial component. The cumulative 1-month evaluation window (Figure 4.7c) reveals similar information gain patterns. Here, CRS7-*new_FMs* outperforms CRS5-*past_FMs* highlighting the medium-term effect of receiver plane updates within the evolving sequence. Finally, the enhanced physics-based model CRS7-*usgs* achieves a higher mean average information gain per earthquake than ETAS.
By plotting the log-likelihood differences in the space domain (Figure B7), we observe that the fault-based CRS forecasts are more localised along the ruptures when compared to the standard ETAS model and outperform the statistical counterpart

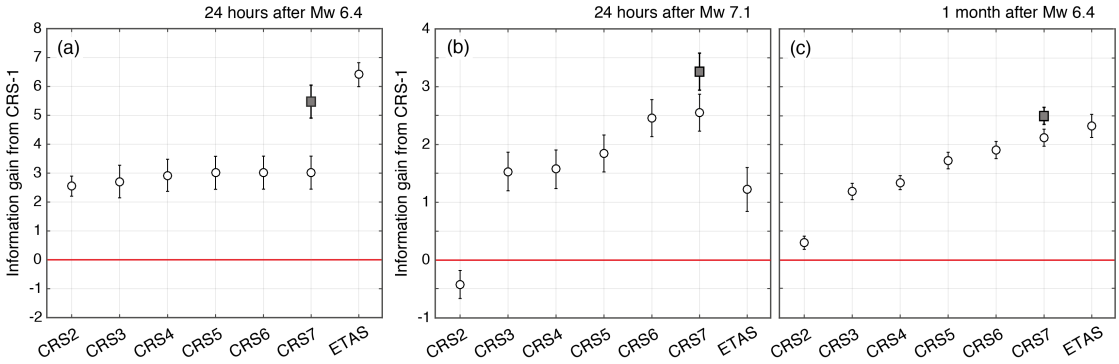

Figure 4.7: Average daily information gain per earthquake from the preliminary CRS1-basic model for: (a) 24 hour after the Mw 6.4 Searles Valley event, (b) 24 hour after the Mw 7.1 Ridgecrest mainshock and (c) for a cumulative 1-month forecast horizon. The filled grey squares indicate the information gain score of the alternative CRS7-*usgs* model. The horizontal lines mark the no-gain level.

in the broader region by predicting low off-fault rates (Figure B7b). However, if we look at smaller distances, we notice that the ETAS model is more robust on the strictly near-fault area (Figure B7c) in agreement with similar previous experiments (*e.g. Segou et al.*, 2013) and with our forecasting experiment in the Central Apennines (Chapter 3), although the enhanced physics-based model CRS7-*usgs* significantly outperforms ETAS in the region of high aftershock clustering around the north-western edge of the Ridgecrest rupture.

To better evaluate the effect of updating the receiver planes during the unfolding aftershock sequence, we show in Figure 4.8a the T-test's log-likelihood differences the for the 1-month forecast between CRS6, updated by past focal mechanisms, and CRS7, updated by past and evolving focal mechanisms. We see two regions characterised by a clear performance improvement (green cells) arising within an otherwise noisy ΔLL signal. We exclude from this discussion the wider area of the Coso field since triggering mechanisms within this active volcanic region may be influenced by other phenomena (*e.g.* fluid flow; *Martinez-Garzon et al.*, 2018). In Figure 4.8b-d, we plot the distributions of pre- and post-Ridgecrest focal mechanisms in the identified regions using the ternary diagrams of *Frohlich and Apperson* (1992), with the addition of a third zone of interest close to the Garlock Fault characterised by lower aftershock rates and mostly unvaried CRS model performance.

To facilitate our interpretation, we present focal mechanisms of cells with notable cumulative log-likelihood difference $|\Delta LL| \geq 6$. In the southern-edge zone, the significance behind the update using evolving focal mechanisms (Figure 4.8d) is smaller since the pre- and post-Ridgecrest focal mechanism populations remains similar. However, the results suggest a shift between pre- and post-Ridgecrest focal mechanism distributions in the two areas (Zone 1 and 2; Figure 4.8b-c) where the receivers update with evolving aftershocks leads to a robust improvement, with promoted strike-slip ruptures (from 54% to 69% and from 53% to 67% in zones 1 and 2, respectively) and suppressed normal fault aftershocks (from 25% to 10% and from 24% to 13%).

To determine whether the pre-existing normal faulting (pre-Ridgecrest) is in fact discouraged within the evolving sequence, we resolve the Ridgecrest coseismic stress changes on the average plane of the pre-Ridgecrest normal focal mechanisms (Figure B8). Indeed, we find that the ΔCFF estimates support a near source stress shadow on pre-existing normal faults that is more evident between 2-12 km depth in Zone 2 (Figure B8, b-f) and below 4 km in Zone 1 (Figure B8, c-f).

The latter observation provides a physical basis for the shift in the focal mechanism population during the unfolding Ridgecrest sequence but also points out the importance of forecast updates using aftershock data.

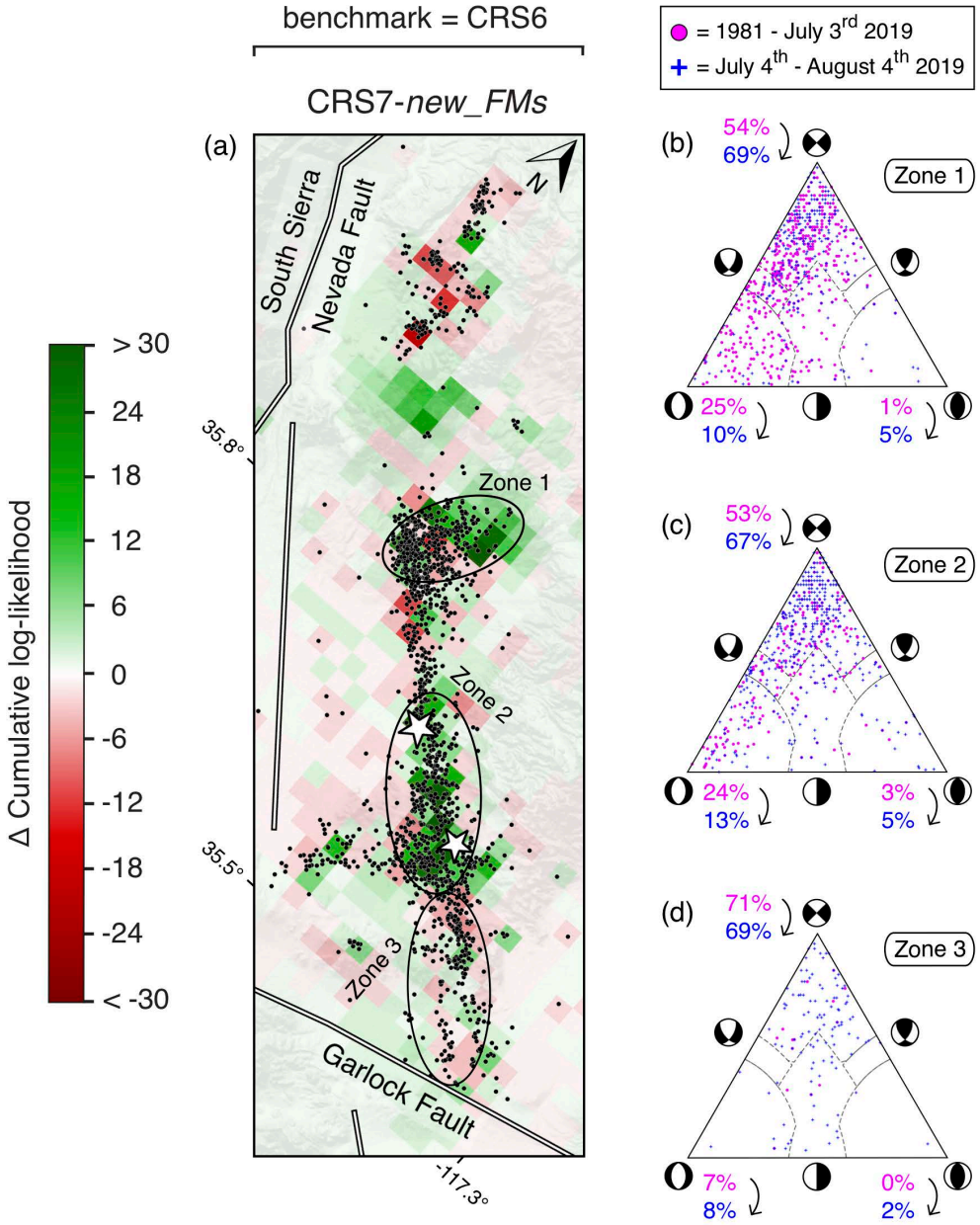

Figure 4.8: Influence of pre-existing and evolving rupture populations in stress-based forecasts. (a) Cumulative ΔLL map for the 1-month forecast horizon between CRS6-*eFFM* and CRS7-*new_FMs*. Positive (green) values indicate a better performance of CRS7-*new_FMs*. Black points indicate the locations of M2.5+ aftershocks between 4 July 2019 and 4 August 2019, white stars indicate the two mainshocks. Values are saturated at ±30 to facilitate visualisation. (b-d) Ternary diagrams showing the focal mechanisms distribution during the pre-sequence (1981-2019, magenta circles) and post-Ridgecrest (blue crosses) time windows.

4.5 Discussion and Conclusions

We tested the predictive skills of seven Coulomb rate-and-state (CRS) forecasts developed within a pseudo-prospective experiment covering the first month of the 2019 Ridgecrest sequence. Our models progressively evolve in their implementation: from an over-simplified parameterisation, based on uniform slip representation and parallel receiver faults, to the most complex physical model incorporating optimised rate-and-state fault constitutive parameters, secondary triggering effects, the USGS Mw 7.1 finite-fault slip model and receivers that consider the UCERF3 faults, off-fault rupture patterns based on pre-existing ruptures, and finally near-source rupture planes revealed by unfolding aftershocks. The forecast results suggest high expected rates along the whole ∼75 km long near-fault region, as confirmed by the observed events. All physics-based models expect increased seismicity rates in Central Garlock Fault, though not significant reactivation has occurred at the time of this writing other than the observed triggered creep (*Barnhart et al.*, 2019). However, delayed aftershocks may be expected on low-stressing rate faults, which highlights the challenges that short- and long-term forecasts must address (*Toda & Stein*, 2018).

When we validate models by means of the formal statistical tests currently implemented within the CSEP community, we see that our results agree with recent works suggesting that advances in the implementation of short-term physics-based earthquake forecasting (*e.g. Segou and Parsons*, 2016) show significant performance increases and can approach, or at times outperform, simple benchmark ETAS models. Specifically, the results confirm those discussed in Chapter 3 that critical components such as finite-fault rupture models, secondary triggering effects, optimised rate-and-state parameters and spatially variable receiver faults significantly enhance the predictive skills of Coulomb stress-based models (*Cattania et al.*, 2018). Our conclusions are further supported by recent modelling developments that illustrate the importance of past focal mechanism data in the estimation of aftershock rupture styles (*Segou & Parsons*, 2020).

Importantly, in this study we evaluated the significance of updating critical components of physics-based models, such as the receiver planes, using aftershock data from the unfolding Ridgecrest sequence. The observed evolving spatial and tem-

poral diversity between the pre-Ridgecrest and within-sequence focal mechanism populations offers a physical interpretation for the estimated local performance improvement, reflected in higher information gains in different regions across the fault. We document a shift in the faulting styles of local triggered seismicity illustrated by a decrease in the percentage of normal fault earthquakes (25% pre-Ridgecrest vs. 10% within the aftershock sequence). In that context, earthquakes on specific pre-existing faulting styles at a local fine scale might be suppressed while others may be enhanced. Therefore, updating the modelled source and receiver populations as aftershock data unfolds is an important step for improving the performance of short-term stress-based earthquake forecasts.

On the other hand, our experimental design showcases one of the modelling caveats that currently affect physics-based aftershock forecasts. We clearly see how, even in a data-rich environment for real-time earthquake products such as California, uncertainties of early focal mechanisms and slip models can be detrimental for operational stress-based forecast models. In particular, the variability of the kinematic parameters associated to the Mw 6.4 Searles Valley event from different providers reveals the influence of data choices among multiple authoritative sources. However, it is extremely encouraging that, although subject to assumptions regarding epistemic and aleatory uncertainties, the most enhanced CRS models that make use of aftershock data can generate informative forecasts that are beginning to compare well to those of statistical models.

Data Source

Some data used in this study were collected by the California Institute of Technology (Caltech) and U.S. Geological Survey (USGS) Southern California Seismic Network (`https://doi:10.7914/SN/CI`) and distributed by the Southern California Earthquake Data Center (SCEDC). The *Hauksson et al.* (2012) and *Yang et al.* (2012) catalogues of seismicity and focal mechanisms can be acquired through access to the SCEDC website (`https://scedc.caltech.edu/research-tools/altcatalogs.html`), as well as the catalogue of focal mechanisms for the Ridgecrest sequence (`https://service.scedc.caltech.edu/eq-catalogs/FMsearch.php`). The ANSS Comprehensive Earthquake Catalog (ComCat) can be searched

at https://earthquake.usgs.gov/earthquakes/search/. The stress inversion by *Luttrell and Smith-Konter* (2017) is available on the SCEC Community Stress Model webpage (https://www.scec.org/research/csm).
The preliminary slip model by G. P. Hayes (USGS) for the Mw 7.1 Ridgecrest mainshock is available at the USGS event webpage: https://earthquake.usgs.gov/earthquakes/eventpage/ci38457511/finite-fault. UCERF3 fault section data is accessible through the open-file report at http://pubs.usgs.gov/of/2013/1165/pdf/ofr2013-1165_appendixC.pdf.
The code "CRS" (*Cattania & Khalid*, 2016) can be downloaded at https://github.com/camcat/crs. Coseismic stress change on 3D individual UCERF3 faults are calculated using Coulomb 3.3 (*Toda et al.*, 2011; https://earthquake.usgs.gov/research/software/coulomb/) and the software "DLC" by R. Simpson (USGS) based on the subroutines of *Okada* (1992).

Chapter 5

Probabilistic Forecasting of Hydraulic Fracturing Induced Seismicity Using an Injection-Rate Driven ETAS Model

The material presented in this chapter appears in the following submitted article.

- Mancini, S., Werner, M. J., Segou, M., and Baptie, B. J. (2020). Probabilistic Forecasting of Hydraulic Fracturing Induced Seismicity Using an Injection-Rate-Driven ETAS model. *Submitted to Seismol. Res. Lett.*.

5.1 Introduction

Reliable forecasts are desirable to mitigate disturbance or damage also from human-induced seismicity. In particular, seismicity induced by fluid injections is a growing concern (*Schultz et al.*, 2020 and references therein).
Statistical models of injection-induced seismicity have shown some skill in capturing the complex range of seismic responses to fluid injections (*e.g. Shapiro et al.*, 2010; *Kiraly-Proag et al.*, 2016; *Verdon & Budge*, 2018). The Epidemic-Type Af-

tershock Sequence (ETAS) model (*Ogata*, 1988), originally developed to reproduce the short-term clustering of tectonic earthquakes, was tested under different fluid-induced seismicity scenarios including natural circulation of fluids at depth (*Hainzl & Ogata*, 2005) as well as human-related activities, such as natural gas extraction (*Bourne & Oates*, 2017), Enhanced Geothermal Systems (EGS - *e.g. Bachmann et al.*, 2011; *Mena et al.*, 2013; *Asanuma et al.*, 2014), hydraulic fracturing for unconventional shale gas development (HF - *e.g. Lei et al.*, 2017; 2019; *Jia et al.*, 2020), and wastewater disposal (*Llenos & Michael*, 2013). These studies concluded that fluid-driven seismicity has distinctive spatiotemporal characteristics, some of which are different from the "regular" tectonic seismicity dominated by earthquake-to-earthquake triggering mechanisms. While the standard ETAS features a stationary background rate due to slower tectonic loadings, *Bachmann et al.* (2011) introduced an ETAS model with a background rate linearly proportional to the injection rate and found that this model performed best in forecasting the seismicity induced in Basel (Switzerland) due to the stimulation of a deep geothermal energy reservoir.

In its limited number of applications to HF environments, the ETAS model was mostly used explore the behaviour of HF-induced seismicity and to show that time-varying background rates positively correlate with injection operations (*Lei et al.*, 2019; *Jia et al.*, 2020). *Lei et al.* (2017) showed that an ETAS model featuring a non-stationary background rate better reproduces the observed features of seismicity when an external forcing is applied (*e.g.* fluid flow or aseismic slip in cases of induced and natural seismicity, respectively), but they did not assess ETAS performance in a formal forecasting experiment.

In this study, we probe the suitability of the ETAS model as a statistical tool for near real-time forecasts of the seismic rates during and after HF operations. We expand on previous applications of the ETAS model to HF by quantitatively assessing the predictive skills of a suite of temporal ETAS models that (i) are calibrated and tested on a much richer microseismicity dataset, (ii) seek to reproduce seismic rates from a wider magnitude range (from M$\sim$3 down to M=-1.5), (iii) explore how the forecast performance changes under different modelling assumptions (standard vs. modified model formulations) and parameterisations (in-sample vs. out-of-sample forecasts), and (iv) test the influence of expressing the non-stationary background

rates by using either averaged or sleeve-specific fluid pumping parameters.
We take advantage of a rich microseismicity dataset recorded at Preston New Road, Lancashire (UK), during unconventional shale gas development by Cuadrilla Ltd in two wells, PNR-1z in 2018 (*Clarke et al.*, 2019) and PNR-2 in 2019. First, we implement the ETAS model in its original tectonic formulation and assess whether (1) it captures the temporal evolution of the microseismicity, and (2) parameters optimised using the available data improve model performance. Second, we implement a modified ETAS model featuring a background seismicity rate proportional to the injection rate following *Bachmann et al.* (2011) but here applied in the context of HF. This presents a particular challenge as HF operations feature short injection episodes along different sleeves, while EGS injections are continuous with gradually changing flow rates at a single injection point. Within the modified ETAS class, we (1) assess model performance against the standard ETAS model, and (2) quantify the influence of using an average (bulk) constant of proportionality between seismicity and injection rates calculated over the entire period of operations at each well versus constants specifically calibrated on individual injection periods. For both ETAS classes we also perform an out-of-sample experiment where we calibrate the ETAS model on PNR-1z data and then use it to independently forecast microseismicity during PNR-2. We rank the forecasts by means of the well-established metric of the likelihood scores.
The comparative performance evaluation illustrates the predictive skills of injection-rate driven ETAS models and how these may inform real-time decision-making by operators and regulators during HF operations.

5.2 Operations and Seismicity at Preston New Road, UK

Hydraulic fracturing operations at the PNR-1z well occurred between 15 October and 17 December 2018. A total of 17 sleeves were hydraulically fractured (Figure 5.1a) with an additional 18 mini fracs, consisting of a few tens of m^3 of fluid pumped. Overall, a total of $\sim$4600 m^3 of slick water fluid was injected (Figure 5.2a) with an average volume per sleeve of 234 m^3 (and a maximum $V_{MAX} = 431$

m^3). Hydraulic fracturing was paused between 3 November and 4 December 2018 as flow-back from the well took place.

The microseismicity at PNR-1z was recorded by a downhole array in the adjacent PNR-2 well consisting of 12 three-component geophones that detected over 38,000 events. The largest event occurred on 11 December 2018 ($M_L = 1.5$). Here, we use the available earthquake catalogue that includes origin times and moment magnitudes (M_w) as determined by *Schlumberger Ltd.*, the geophysical processing contractor. The limited dynamic range of the downhole geophones leads to problems in magnitude estimation for $M_w \geq 0.0$ events due to clipping. To avoid a potential bias, we matched these with events in the catalogue obtained from broadband surface stations operated by the British Geological Survey (BGS) that reported 172 events with local magnitudes (M_L). We replaced the moment magnitudes for all $M_w \geq 0.0$ events in the downhole catalogue with the corresponding local magnitude estimate. We estimate a magnitude of completeness (M_c) between -1.2 and -1.5 (see Appendix C, Figure C1). We chose M_c = -1.5 for our analyses to increase the number of events to around 20% of the entire dataset.

Figure 5.2a shows a histogram of the hourly number of events during operations along with the cumulative volume of injected fluid. The observed seismicity at PNR-1z shows multiple peaks that visually correlate well with the pumping periods and then decay rapidly with time after injection stops. We find evidence of considerable variations in seismic responses despite comparable injection rates across sleeves (*e.g.* Figure 5.2 c-d). For instance, at sleeve #2 (injection stage S02) event rates increase as soon as injection starts and remain relatively stable (Figure 5.2c), while at sleeve #40 (injection stage S17) there is a delayed onset of seismicity followed by substantially higher rates (Figure 5.2d).

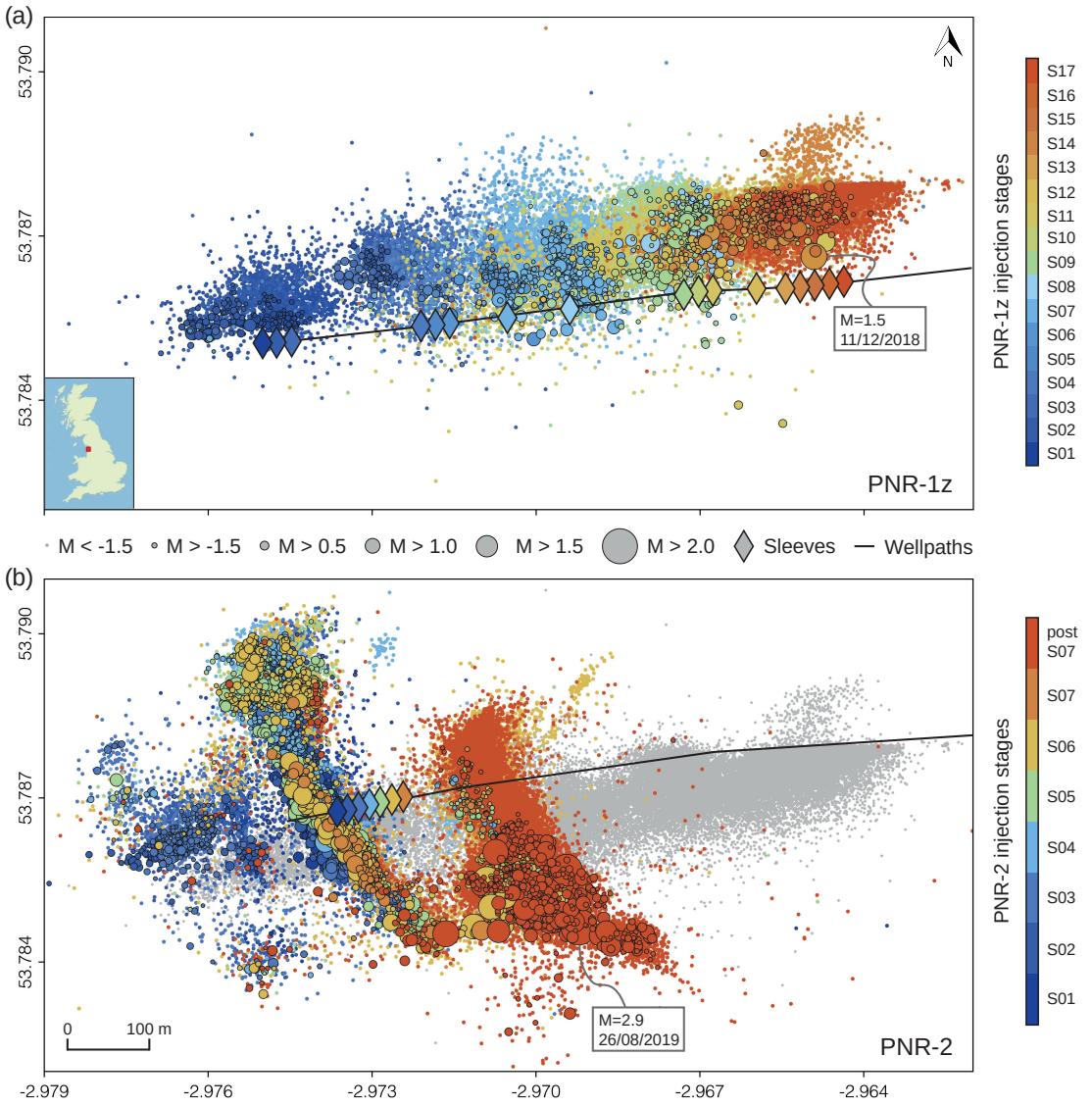

Figure 5.1: Map view of earthquakes recorded during hydraulic fracturing at the Preston New Road unconventional shale gas site. Events are colour-coded by the associated injection stage and their size scales with magnitude. (a) Seismicity between 15 October and 17 December 2018 during and after injection at the PNR-1z well. (b) Seismicity between 15 August and 2 October 2019 during and after injection at the PNR-2 well; grey dots indicate the epicentres of events occurred during operations at PNR-1z. The black lines represent the surface projection of the two wellpaths. Diamonds illustrate the position of the sleeves worked during the operations at the two wells and are coloured by the corresponding injection stages.

The horizontal PNR-2 well runs roughly parallel to PNR-1z offset by approximately 200 m. Operations started on 15 August 2019 but were suspended on 26 August following a $M_L = 2.9$ earthquake that was felt up to a few kilometres from the epicentre (*Cremen & Werner*, 2020). Seismicity was recorded by a downhole array of 12 geophones in the adjacent PNR-1z well, and the final catalogue, extending up to 1 October 2019, consists of over 55,000 microseismic events (Figure 5.1b) with magnitudes reported as M_w. We added a correction of 0.15 magnitude units to the downhole moment magnitudes following *Baptie et al.* (2020). Furthermore, the PNR-2 catalogue suffers from brief but critical data gaps that result in a loss of otherwise recorded seismic events, including the largest event in the sequence ($M_L = 2.9$) and presumably its early aftershocks. We filled these gaps with events recorded by the combined surface network of the BGS and the operator (*Baptie & Luckett*, 2019). Although we estimate a magnitude of completeness below -1.5 (Figure C1b), we use M_c = -1.5 for comparability with the PNR-1z catalogue. The early earthquake productivity at PNR-2 appears an order of magnitude larger than that observed during the initial injection stages at PNR-1z, even under similar injected volumes (Figure 5.2b). The complexity of the seismic response to injection is similar to PNR-1z (Figure 5.2e). As at PNR-1z, we observe a general positive co-dependency between seismicity and fluid injection at PNR-2.

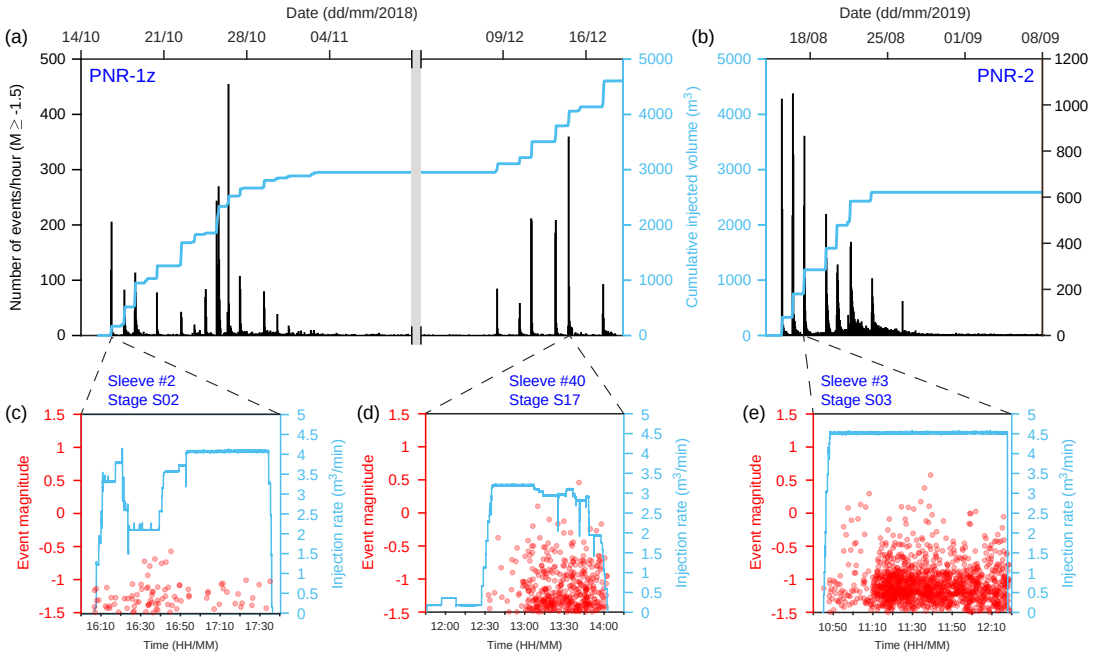

Figure 5.2: Seismicity response to hydraulic fracturing at the Preston New Road site. (a-b) Histograms of the number of M $\geq$ -1.5 events per hour (black bars) as a function of time during operations along with the cumulative volume of injected fluid (light blue line) at PNR-1z and PNR-2, respectively. For illustration purposes, we inserted a time gap during the pause of operations at PNR-1z, which is indicated by the grey area. (c-e) Examples of seismic productivity and earthquake magnitudes vs. time (red circles) in response to the injection history (light blue line) at selected sleeves.

5.3 Methods

5.3.1 The Standard ETAS Model

We create three versions of the standard ETAS model (the "ETAS1" class). In ETAS1-*optimised* we estimate ETAS parameters from the target catalogue (either PNR-1z or PNR-2) and thus perform an in-sample (best-case) forecast evaluation. In ETAS1-*unoptimised* we use the parameters estimated from PNR-1z data to forecast the PNR-2 seismicity out-of-sample. ETAS1-*global* serves as an alternative benchmark model with ETAS parameters (except for the background rate)

estimated from global subduction zones by *Zhang et al.* (2020).

5.3.2 The Modified ETAS Model for Injection-induced Seismicity

In the second forecast class ("ETAS2"), we modify the ETAS model to account for events forced by an external driver. We couple the background rate to the time-dependent fluid injection rate $I_r(t)$:

$$\lambda_m(t|H_t) = \mu(I_r) + \sum_{i:t_i<t} Ke^{\alpha(M_i - M_{cut})} \times c^{p-1}(t - t_i + c)^{-p}(p-1), \tag{5.1}$$

with λ_m a "modified" seismic rate and the background rate $\mu(I_r)$ now assumed to be linearly related to the injection rate via a constant of proportionality c_f (*Bachmann et al.*, 2011):

$$\mu(I_r) = c_f(I_r(t)). \tag{5.2}$$

To estimate c_f, we maximise:

$$\log L(c_f, K, \alpha, c, p) = \sum_{i=1}^{N} \log \lambda_m\left(t_i|H_t\right) - \int_{T_0}^{T_1} \lambda_m\left(t\right) dt. \tag{5.3}$$

Within the ETAS2 class, we develop three forecast versions. In ETAS2-*bulk* we estimate and use only a single value of c_f for each well, fit over the entire period of operations.

ETAS2-*specific* implements specific values of c_f for each sleeve, calibrated within the individual injection periods; in this model, we fix the triggering parameters (K, α, c, p) to the respective values previously obtained for ETAS2-*bulk* assuming that the contribution of event-to-event interactions does not change in different injection periods, when the external forcing is likely to be the dominant mechanism of earthquake production.

Finally, ETAS2-*unoptimised* uses the ETAS parameters estimated on the PNR-1z

catalogue (including its bulk proportionality constant) to forecast out-of-sample the expected seismic response at PNR-2.

Simulating ETAS2 models requires a different method for background events during injection periods. We apply the thinning algorithm (*e.g. Zhuang & Touati*, 2015): (i) estimate a mean expected number of forced events ($\overline{N_f}$) by multiplying c_f by the injection rate integrated over the duration of either the injection period or the forecast window (whichever is shorter); (ii) draw a random variable (N_f) from a Poisson distribution with mean equal to $\overline{N_f}$; (iii) distribute the N_f events in time according to a piece-wise linear, non-homogeneous Poisson process with rate $\mu(I_r)$ driven by the injection rate (smoothed using 1-minute moving windows); (iv) simulate all aftershock generations triggered by the directly forced events by means of the standard procedure.

For consistency, all six ETAS versions are: (1) updated hourly or when an injection period starts (whichever comes sooner), (2) estimated by 1,000 stochastic ETAS simulations with fixed M_{max}= 6.5 (the most likely regional maximum expected tectonic magnitude; *Woessner et al.*, 2015), and (3) forecasting the number of M $\geq$ -1.5 events.

In Appendix C, we report a summary of the tested ETAS versions (Table C1) and the values of the ETAS parameters (Table C2), including the bulk and sleeve-specific values of c_f (Tables C3 and C4 for PNR-1z and PNR-2, respectively).

5.3.3 Evaluation of Model Performance

Because each forecast consists of a probability distribution of earthquake numbers over the forecast period, we evaluate and rank forecast models using a probabilistic score, namely the log-likelihood values. The score quantifies the likelihood of the observed number if the models were the data-generator, specifically the logarithm of the probability Pr(ω|model) of observing ω earthquakes given the ETAS forecasts (*Zechar*, 2010):

$$LL(\omega|model) = log(Pr(\omega|model)). \tag{5.4}$$

To compensate for the limited number of simulations, which is likely to under-

sample the range of possible simulated ETAS rates, we approximate the simulation histogram of each forecast window with a Negative Binomial Distribution (NBD; *Harte*, 2015) (Figures C2 and C3). We choose the two-parameter NBD because it characterises earthquake clustering and process over-dispersion much better than the Poisson distribution (*Kagan*, 2010). We calculate the likelihood scores from the fitted NBD.

5.4 Results

5.4.1 Forecast Timeseries

In Figure 5.3, we present the incremental hourly timeseries of the three in-sample ETAS forecasts for PNR-1z and PNR-2. We select illustrative sub-periods characterised by (1) weak and strong seismic responses to injection, and (2) seismicity without injection. The panels compare the observed number of M $\geq$ -1.5 events per hour with the mean and 95% predictive interval of the ETAS model.

First, we find that the ETAS1 class projects the onset of increased rates with a 1-hour delay compared to observations. This is not an unexpected effect due to the scarcity of M $\geq$ -1.5 parent earthquakes prior to each injection period and the fact that ETAS1 does not account for external seismicity forcing. In post-injection conditions, when any earthquake clustering is likely driven by event-to-event triggering, ETAS1-*optimised* reproduces well the hourly seismicity within the model's 95% ranges at PNR-1z (Figure 5.3a) and PNR-2 (Figure 5.3b). During periods of no injection and low seismicity at PNR-1z, the 95% forecast range often encompasses the critical value of zero events, reflecting the intrinsic stochasticity of the ETAS model. The standard ETAS1-*optimised* severely underestimates the observed rates by an order of magnitude during the high seismicity periods, whether the seismic response is weak or strong.

The ETAS2 class, featuring an injection-rate-driven background rate, substantially reduces the discrepancies with the observed rates. ETAS2-*bulk*, which captures the average seismic response to injection, both under- and over-predicts during injection periods. This mixed performance is a result of the single proportionality constant for each dataset that does not sufficiently capture the complex relation-

ships between injection rate and seismicity. ETAS2-*specific*, which describes the seismicity response with sleeve-specific injection data, presents the best match during the periods of high seismicity rate due to pressurised fluid forcing. Here, the visual comparison is very encouraging, but hinges on in-sample, sleeve-specific proportionality constants between seismic rates and injection rates.

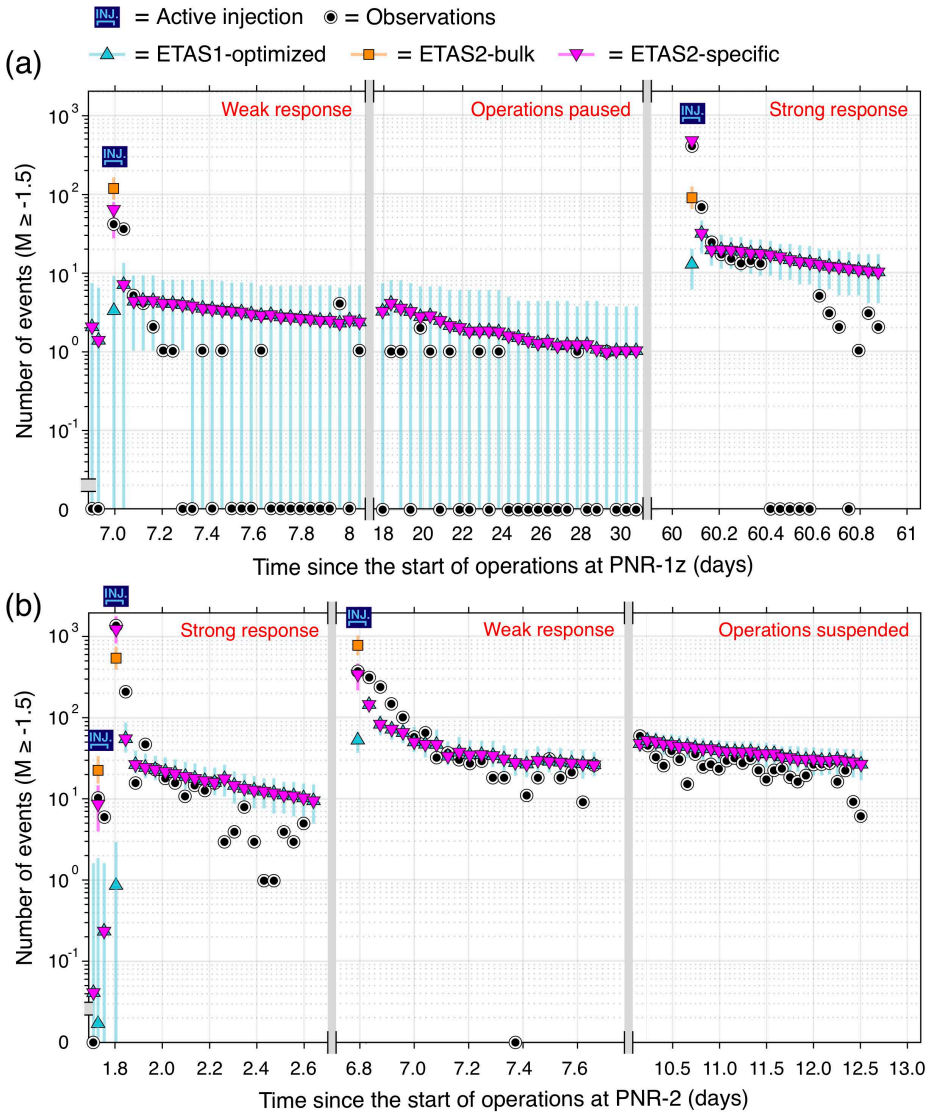

Figure 5.3: Incremental 1-hour timeseries of expected vs. observed number of M $\geq$ -1.5 events at PNR-1z (a) and PNR-2 (b). We report selected examples from injection sleeves characterised by weak and strong seismicity response as well as during the pause of operations. ETAS2-*bulk* model predictions are shown only during injection periods indicated by the "Inj." label (otherwise its forecasts are identical to ETAS1-*optimised* and ETAS2-*specific*). Black circles indicate the number of observed events in each forecast window. Other symbols represent the mean (expected) number from the simulations. Bars denote 95% ETAS model simulation ranges. For illustration purposes during periods of suspended/paused injection, data are plotted at 12-hour intervals for PNR-1z and 2-hour intervals for PNR-2.

We next analyse the performance of all ETAS models, including the out-of-sample versions, over the entire testing periods at PNR-1z (Figure 5.4 a-c) and at PNR-2 (Figure 5.4 d-f). Using a simple acceptance/rejection criterion, we consider a forecast accepted (green symbols) if the observations fall within the 95% model range, otherwise we mark it as rejected (red symbols). An ideal forecast, which predicts the observations perfectly, aligns along the diagonal lines of Figure 5.4. While the observations fall into the 95% forecast range of the ETAS1 models about 80% of the time, these matches correspond to periods of low seismicity: accepted forecasts occur only when less than 40 events are observed at PNR-1z (Figure 5.4 a,b) and less than 150 events are observed at PNR-2 (Figure 5.4 d,e). We also note that (1) at both PNR-1z and PNR-2 ETAS1-*global* overpredicts less frequently than models parameterised on well-specific seismicity when the seismicity rate is extremely low (Figure 5.4 a,b and Figure 5.4 d,e) but underpredicts more during high-rate windows, and (2) in PNR-2 the differences between ETAS1-*optimised* and ETAS1-*unoptimised* are negligible (Figure 5.4d), a result of the similar parameters estimated from the two wells (Table C2).

The performance of the ETAS2 class (Figure 5.4 c,f) differs from ETAS1 mostly during injection periods, and the improvement is appreciable. ETAS2-*specific* performs strikingly well, as the only model to forecast very productive periods with more than 300 events at PNR-1z (Figure 5.4c) and more than 1,000 events at PNR-2 (Figure 5.4f). Finally, the out-of-sample ETAS2-*unoptimised* model, which uses the bulk seismic response to injection at PNR-1z to forecast seismicity at PNR-2, persistently underpredicts injection-induced high rates (Figure 5.4f), but its underprediction is less severe than that of the ETAS1 class.

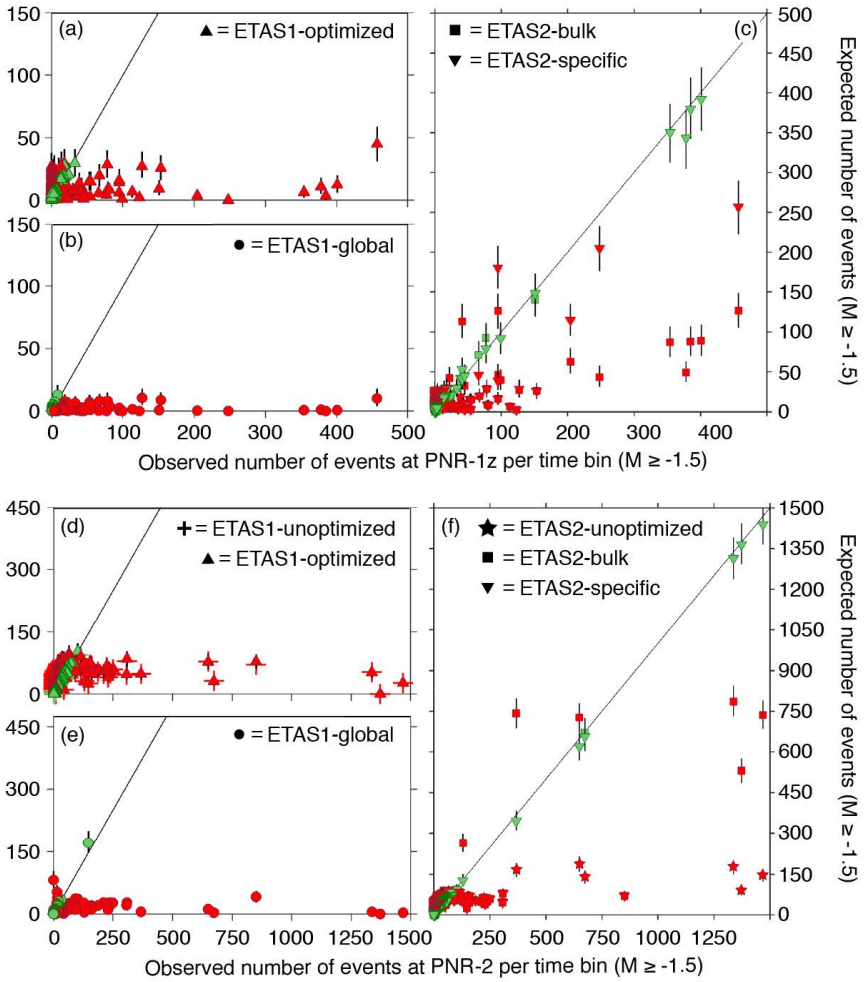

Figure 5.4: Observed vs. expected number of events per forecast period over all injection stages. Red symbols denote rejected forecasts (data outside model range); green symbols denote accepted forecasts. Bars denote 95% ETAS model simulation ranges.

5.4.2 Likelihood Scores

The cumulative log-likelihood scores of the models over the entire duration of the PNR catalogues show that the injection-rate driven ETAS2 realisations considerably outperform models belonging to the standard ETAS1 class (Figure 5.5).
In particular, ETAS2-*specific* has the highest likelihood scores at both wells and

thus ranks as the best performing model, followed by ETAS2-*bulk* as second-best. The latter performs unevenly in the two wells, with better predictive skill in PNR-1z (Figure 5.5a) than in PNR-2 (Figure 5.5b) during the first few days of operations. Encouragingly, the out-of-sample ETAS2-*unoptimised* model scores better than all ETAS1 models and performs similarly to ETAS2-*bulk* during the first week of treatment of PNR-2. In other words, a model calibrated on PNR-1z data could have provided informative forecasts for PNR-2.

ETAS1-*global* performs worse than the injection-rate driven ETAS2 class but compares well with the other ETAS1 models and even with the ETAS2-*unoptimised* and ETAS2-*bulk* models in the early stages of PNR-2 (inset of Figure 5.5b); this is a priori surprising for a model calibrated on moderate to large subduction zone earthquakes.

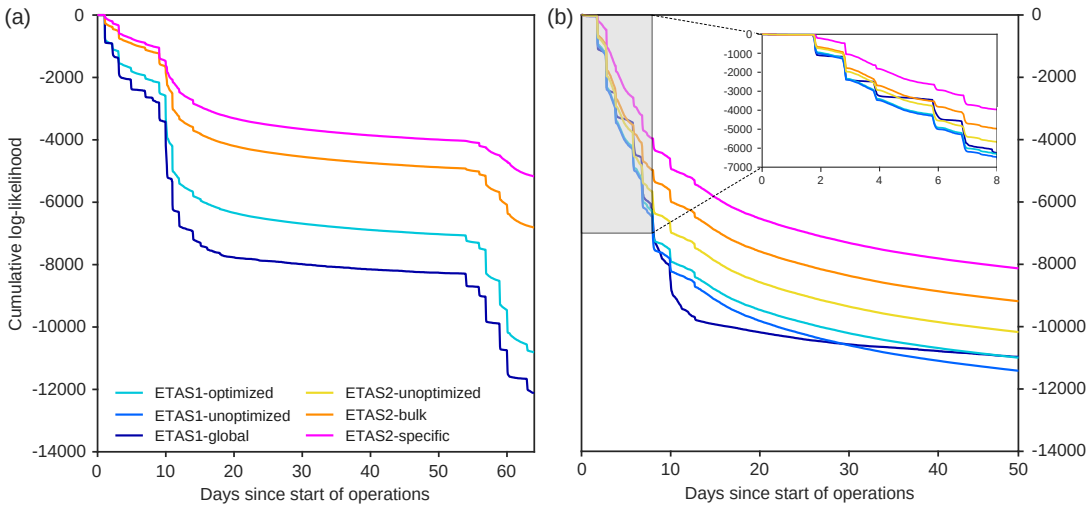

Figure 5.5: Cumulative log-likelihood timeseries. ETAS models tested on (a) PNR-1z and (b) PNR-2.

5.5 Discussion and Conclusions

The PNR microseismic datasets present a unique opportunity to develop and evaluate statistical forecasting models of hydraulic fracturing induced seismicity.

Notwithstanding the variability and uncertainties linking pumping data to the induced seismicity response at both PNR wells, we observe a generally positive co-dependency between seismicity and injection rates that supports the incorporation of operational parameters into the standard tectonic ETAS model.
In comparing the performance of the standard and injection-rate driven ETAS forecasts, we find that the seismicity decay after the operations, or between stages, is satisfactorily captured by the standard ETAS. We interpret this result as follows. During operations we witness the complex interplay of rapid pore pressure effects and earthquake clustering, expressing a variety of possible mechanisms (*e.g.* elasto-static stress transfer, poro-elastic effects, aseismic creep) (*Schultz et al.*, 2020), while external forcing ceases in inter- and post-injection periods and seismicity shows a more typical tectonic behaviour.
However, the log-likelihood scores of the ETAS models demonstrate that a non-stationary background rate tied to the injection rate is necessary to avoid severe underpredictions during injection periods, when the seismic productivity is high. Thus, even a simplistic linear relationship between injection rate and induced seismicity leads to informative ETAS forecasts in HF environments.
From the model comparison, we conclude that (1) bulk constants of proportionality do not accurately describe the variable seismic response to fluid injection, and (2) a sleeve-specific modulation of the seismic response to injection is the most critical element for producing reliable forecasts.
In our study, the best-performing ETAS model is an in-sample forecast that represents a best-case scenario. This performance may be difficult to attain out-of-sample. However, the sleeve-specific constants of proportionality could be estimated and fine-tuned in near real-time conditions from the initial seismic response at the sleeve, following the approach presented in published experiments to estimate other real-time model parameters at the two PNR wells, such as the seismic efficiency (S_{EFF}) and the seismogenic index (SI) (*Clarke et al.*, 2019; *Kettlety et al.*, 2020). In this regard, we acknowledge that a direct comparison between our sleeve-wise c_f values and the SI and S_{EFF} constants for PNR-1z and PNR-2 would be a valuable contribution to further explore whether there is one or more predominant geological or physical mechanism to explain the variable injection rate scaling between different sleeves (*e.g.* pre-stress heterogeneity, injection into dif-

ferent rock formations, crossing pre-existing critically stressed fracture zones or faults, etc.). Unfortunately, this comparison is currently hampered, as we evaluate incremental c_f constants for each single injection period, while *Clarke et al.* (2019) and *Kettlety et al.* (2020) use SI and S_{EFF} values obtained considering the cumulative PNR microseismicity. A further study with more focus on this point is therefore suggested.

Given the temporal variability of the seismic response to constant injection, the parameters will doubtlessly be more uncertain, and this additional uncertainty should be propagated into the forecasts. In this regard, the operator would have to assume that (1) the injection rate at each sleeve is known in advance and (2) the evolving sleeve-specific seismic response is continuously acquired and adequately detected to support frequent model calibrations.

To mimic real-time conditions (*i.e.* before data are available for parameter estimation), we also evaluate forecasts from three out-of-sample models. Although their performance is worse than the in-sample models, we also see encouraging results. The models present low log-likelihood scores in the longer term (*i.e.* more than 3-5 days after the start of operations), but they perform comparably to some in-sample models during the first few days of operations. This is true even for the ETAS model calibrated on data from global subduction zones. This is promising for operational conditions: operators could provide forecasts during the very early stages of operations using parameters that are either generic or previously calibrated on adjacent wells. As well-specific and stage-specific data become available, forecasts can be improved with re-estimated parameters and the operational injection data.

In light of the outcomes of the PNR experiments, we conclude that injection-rate driven ETAS models produce informative time-dependent probabilistic seismic rate forecasts. This result is likely significant also for more informed decision-making within the traffic light schemes. It is indeed worth remembering that not only does ETAS provide a model of the full distribution of seismicity as a function time, but it also simulates the associated magnitude distribution. As the model performs well in capturing the M $\geq$ -1.5 seismicity at PNR, the evolving Gutenberg-Richter b-values could be used to extrapolate the expected induced rates (along with their associated probability of occurrence) at larger (*i.e.* potentially damaging) magni-

tude thresholds. Although it was beyond the scope of this study, this approach would complement other magnitude forecasting strategies that *Clarke et al.* (2019) and *Kettlety et al.* (2020) have already tested on the PNR datasets.
Given the reasons above, injection-rate driven ETAS models represent a useful mitigation tool that, convolved with models of ground motion, exposure and vulnerability, can support time-dependent probabilistic seismic hazard and risk assessment. These forecast models may therefore provide useful information for operators, regulators, residents and other stakeholders in HF environments.

Data Source

The PNR-1z and PNR-2 microseismicity catalogues as well as the fluid injection rate data used in this study can be acquired through access to the UK Oil and Gas Authority website at `https://www.ogauthority.co.uk/exploration-production/onshore/onshore-reports-and-data/`.

Chapter 6

Discussion and Conclusions

In this dissertation we presented the results of three short-term earthquake forecast experiments, carried out in environments of tectonic as well as induced seismicity both in pseudo-prospective (out-of-sample) and retrospective (in-sample) modes. While summarising the key findings, in this chapter we evaluate their limitations and practical implications within the current research framework, and discuss their importance in motivating the development of future modelling strategies.

6.1 Findings and Future Directions for CRS Forecasts

In Chapters 3 and 4, we explored which are the most critical elements for improving the operational performance of short-term Coulomb rate-and-state (CRS) aftershock forecasts compared to a standard ETAS benchmark.
By testing models developed under a common physical assumption (*i.e.* stresses are statically transferred between faults) but validated considering different modelling choices, we found that their predictive skills considerably improve when incorporating increasingly high quality data products. We therefore could not reject the static stress transfer hypothesis, but rather highlight how its specific implementations in continuum mechanics affect the performance of CRS forecasts. In particular, we concluded that elaborate CRS forecasts reach in performance (and sometimes outperform) the standard ETAS models only when they include

(1) spatial and temporal crustal heterogeneities to characterise earthquake sources and hazardous faults, and (2) an evolving stress field.

The Representation of Faults

The Central Apennines and Ridgecrest experiments highlight how remarkable is the effect of the representation of fault fabrics on the predictive skills of stress-based forecasts. This is a consequence of the uncertain constraints usually available to represent faults in the broader sense, be it by means of geological evidence, geophysical profiling, or taking advantage of other observations from diverse data sources (seismological, interferometric, GPS, etc.).

First, this work shows that when CRS models feature over-simplified source fault representations, with empirically derived fault dimensions and uniform slip distribution, they fail to capture even the first order spatial features of near-source triggered seismicity. The results highlight that the incorporation of spatially variable finite-fault slip models is of dramatic importance to better resolve the near-fault coseismic stress patterns.
One drawback is that depending on the inversion algorithm and the type of inverted data, alternative kinematic models are often published for the same event. Unfortunately, there is no robust evaluation method to establish which one is the "true" slip distribution as these models represent a compromise between non-unique solutions and are affected by epistemic uncertainties that are unlikely to be resolved in the near future. Furthermore, propagating such uncertainties into CRS models is hard as it is not common practice to express them explicitly in standard slip model formats; this point is critically important to guide future interactions between data providers and modellers, as well as to agree on what should be the mutually accepted data formats. A possible way to tackle the non-uniqueness of slip models is to implement a Bayesian framework where slip distributions are used as priors, and higher resolution displacement models are progressively inverted using the observed aftershock patterns; however, tests published so far have not shown significant model improvements (*Strader*, 2014), not to mention the difficulties in finding a physical justification for such approach. Other studies in literature show how the ensemble averaging of forecasts based on different rupture models can

address the issue of multiple available slip distributions in a more promising way (*e.g. Marzocchi et al.*, 2012; *Cattania et al.*, 2014). However, we note that selecting among different slip models generally only affects the performance of purely retrospective forecasts. The most detailed descriptions of source faults are indeed often achieved after several weeks or months, once the processing of large amounts of data makes it possible to reconstruct complicated fault geometries (often made of a mesh of distributed sub-faults rather than one continuous linear rupture) and to image the small-scale features of slip distributions. In this regard, our pseudo-prospective test on the Ridgecrest sequence shows extremely encouraging results, as the incorporation of even the most preliminary near real-time finite-fault slip model for the Mw 7.1 mainshock contributes to a remarkable improvement of forecast performance (by 2 information gain units) with respect to the CRS versions featuring a simplified synthetic source.
On the other hand, the same case study also suggests how artefacts present in early inversions may affect the spatial consistency of the stress-based models. While this result may seem discouraging for those hoping to use near real-time slip models as a mere drop-in input data, minor amendments to declared model errors are possible in an operational context (*e.g.* constraining fault dimensions from alternative sources such as ShakeMaps or editing the actual amount of slip by means of empirical magnitude-dependent relations).
We find therefore critical that authoritative data providers put their efforts in automated or semi-automated inversion of real-time data to deliver finite-fault slip distributions (and the related uncertainties) fast enough to be incorporated in near real-time CRS models.

An even more challenging aspect concerns the representation of receiver faults at different spatial scales. While the optimally oriented planes assumption is becoming less popular for not being supported by observations (*e.g. McCloskey et al.*, 2003; *Ishibe et al.*, 2011; *Segou & Parsons*, 2020), CRS model performance improves when introducing the large-scale geological representation of receiver planes. Here, the critical question is: how much do we actually know about regional faulting?
Although lines of evidence suggest that the majority of large-magnitude after-

shocks are most likely to occur on well-expressed faults (*e.g. Parsons & Segou*, 2014), because of our lack of knowledge we continue to witness high magnitude earthquakes occurring on unknown faults. One of the clearest examples is provided by the 2019 Ridgecrest sequence, where a Mw 6.4 foreshock triggered a Mw 7.1 event on a 55-km long unmapped strike-slip rupture. This underpins the argument that even in the best studied regions on Earth fault inventories (*e.g.* UCERF3, DISS, etc.) are incomplete, and stress perturbations can activate or connect (*e.g.* the 1992 Landers sequence) large previously unidentified faults.

The modelling strategy that we implemented in Chapters 3 and 4 reveals that a blend of large-scale seismological data such as smoothed stress inversions and more localised information from past focal mechanisms can be successfully adopted as proxy of earthquake rupture styles where no direct geological information is available.

Another important finding emerges that faulting styles vary not only in space, but also in time following the unfolding coseismic stress conditions. In Chapter 4 we illustrated how some small-scale aftershock populations with certain orientations and rakes can be activated only in response to specific ruptures. In particular, during the Ridgecrest sequence static stress changes locally drove away normal faults from rupture and further promoted strike-slip faulting aftershocks. By using the incoming focal mechanisms to detect such shift in the relative abundances of ruptures styles, we find that updating receiver planes using evolving aftershock data improves model performance in the near-source region.

The above results should foster discussion within the physics-based earthquake forecasting community around the treatment of receiver faults. In particular, one matter of debate should regard focal mechanisms, as they certainly represent a data product CRS modelling heavily relies on to characterise faults. The quality and completeness of the currently available focal mechanism catalogues can potentially be enhanced by applications of deep learning techniques that increase the number of phase detections (*e.g. Ross et al.*, 2018; *Hara et al.*, 2019). Such improvement in data quality might make it easier to reduce their inherent epistemic uncertainties and propagate them into the model.

Furthermore, in Chapter 4 we used aftershock focal mechanisms to include evolving planes into stress calculations. Here, we note that interesting pathways can

be opened by the machine learning (ML) algorithms that have been developed in the last few years to support a range of issues from better real-time detection of early aftershocks (*Ross et al.*, 2018) to improved event characterisation. Coupled with existing real-time high-resolution earthquake location methods (*e.g.* RT-DD; *Waldhauser*, 2009), ML techniques hold the potential to greatly help defining fault structures as they are revealed by aftershocks with an unprecedented level of detail. Although fitting fault planes by only using events' hypocentres likely represents a hard challenge (especially in the presence of distributed and complex fault patterns), this auxiliary data product can potentially be useful in combination with early focal mechanism solutions to better constrain the geometry of activated ruptures and to help resolving the nodal plane ambiguity.
In summary, considering that (i) we have a solid enough theoretical understanding of which is the range of mechanically possible ruptures under the effect of the regional and coseismic stress fields, (ii) at a fine scale aftershocks can occur on a wider spectrum of possible ruptures than it was previously thought (*Segou & Parsons*, 2020), and (iii) fault populations also change in time, the idea of resolving coseismic stress changes using only large scale known faults with prescribed kinematic parameters should be relaxed. A desirable approach to make all the above elements coexist with our knowledge limitations (*e.g.* the currently large uncertainties in focal mechanisms) would be treating receiver planes in a probabilistic rather than deterministic way, so as to define an array of candidate receivers at each grid point onto which to resolve the coseismic stress changes. This paves the way to fascinating avenues of research and represents a challenge for modellers that will have to produce unified and globally consistent CRS protocols.

The Importance of Smaller Events

Not only are enhanced earthquake detection and high-quality locations needed for imaging aftershock ruptures, but also to better unravel the spatiotemporal features of the evolving stress field.
Secondary triggering effects have been disregarded by many of the previous forecasting tests in literature (*e.g. Toda et al.*, 2005; *Toda & Enescu*, 2011; *Woessner et al.*, 2011; *Segou et al.*, 2013 among others), but our experiments highlight how

they actually matter and improve model performance. Although smaller magnitude earthquakes are capable to generate only minor and localised stress perturbations, their cumulative fine-scale effect over the wider aftershock area controls earthquake sequences.
This means that in order to improve model performance in those intermediate off-fault regions currently characterised by low information gain, stress-based forecasts should behave more ETAS-like. It is indeed recognised that also in the ETAS model a critical role in controlling the overall aftershock patterns is played by the cascade of triggered events (*Helmstetter & Sornette*, 2003) and that considering the triggering contribution of smaller magnitude events increases the forecast skills (*Werner et al.*, 2011).
On the other hand, the most recent advances in ETAS modelling (*Field et al.*, 2017) suggest a more fault-informed approach, in a way rendering ETAS more CRS-like and opening the way for further testing of hybrid models.

Additional Modelling Elements

Further work is required to test the inclusion of additional model characteristics or the revision of current ones in future implementations.
For instance, we show the critical role played by the optimisation of the rate-and-state parameters to better match the observed seismicity response. In Chapters 3 and 4, we used rich learning phase catalogues with multiple large stress steps to calibrate the constitutive fault parameters through likelihood maximisation (*Cattania & Khalid*, 2016); this approach has potential limitations in regions with poorer seismic monitoring systems or very slow deformation rates (*i.e.* scarce seismicity), although modern reprocessing of global catalogues may partially close this gap. In similar cases, it may be worth exploring further how fault constitutive parameters can be estimated in near real time using the observed seismicity response to the evolving stress field (*e.g. Maeda*, 2006) and how this may influence the expected seismic productivity of CRS models (*i.e.* the N-test scores).
Moreover, spatially homogeneous rate-and-state parameters still represent an oversimplification of reality; however, *Cattania et al.* (2014) found that the spatial variability of rate-and-state parameters has only a moderate impact on model per-

formance, as they do not modify the spatial features of the stress field.

Other points remain open as one may argue that the current physics-based models do not actually carry enough physics in them.
Although we find that static stress changes coupled to rate-and-state frictional laws offer a valid description of the underlying physical processes of earthquake triggering, further formal testing is needed to assess whether increasingly complex models can benefit from the incorporation of additional deformation sources playing out at different time scales.
Questions that should be pursued in future research include: does post-seismic deformation have an important role in keeping the probabilities of strong aftershocks high even after several years from the first large magnitude event? Can we model in near real-time the control of high pore pressure on the spatiotemporal seismicity patterns during aftershock sequences? Are there more triggering mechanisms that we are not aware of and that should be accounted for?

6.2 Lessons Learnt from the HF-induced Seismicity Environment

The first two pseudo-prospective experiments of this thesis explored CRS and ETAS performance in regions characterised by medium shear stressing rates and were developed to forecast daily aftershock rates for total horizons of one month (Ridgecrest) or one year (the AVN sequence). In the framework of tectonic seismicity, interplate environments such as large subduction zones represent instead faster deforming regions with comparatively short-lived aftershock sequences (*Toda & Stein*, 2018). Within different boundary conditions, fluid-induced seismicity environments represent great opportunities to study earthquake triggering under very high shear stressing rates (although originating from human-controlled external forcing rather than tectonic loading), with the advantage of being localised within much more limited areal extents that facilitate enhanced catalogue development and model testing.
However, testing forecasts at a smaller spatial and temporal scale requires extremely good data. This was the case at the Preston New Road hydraulic frac-

turing site in UK, where we used an extraordinarily rich microseismicity dataset to test several versions of the ETAS model as a probabilistic tool to reproduce the temporal (hourly) evolution of hydraulic fracturing induced seismicity (Chapter 5). We found that (1) the standard tectonic ETAS failed to reproduce the high seismicity rates observed during fluid injection as it did not account for the additional forcing of pressurised fluids, (2) the best-performing model was a modified ETAS parameterised on well-specific seismicity and featuring a non-stationary background rate driven by sleeve-specific injection rates, and (3) a modified out-of-sample ETAS outperformed all the in-sample standard ETAS versions.

Not only are these findings encouraging as they show that injection-rate driven ETAS models provide informative time-dependent forecasts for operators and decision-makers during HF operations, but also let us draw a parallel with some results from the CRS experiments on tectonic seismicity.

First, they raise the question of whether Coulomb models of purely tectonic seismicity should also account for triggering sources that are usually overlooked (*e.g.* aseismic slip, high pore pressure effects), with the clear disadvantage that in natural seismicity we do not have any direct control over such forcing. For instance, future work is required to establish the viability of improving model performance by using detailed diffusivity models to inform CRS and ETAS aftershock forecasts with non-stationary background rates in regions where natural circulation of deep fluids is recognised to play a key role, such as the Central Apennines (*e.g. Lombardi et al.*, 2010; *Malagnini et al.*, 2012) or the central US (*e.g. Norbeck & Rubinstein*, 2018).

Second, the findings from the HF environment reinforce the concept that better-performing forecasts rely on high-quality data products allowing for a time dependent model parameterisation: as the next-hour ETAS forecast at PNR needs to be calibrated in real time using the evolving seismicity response to the time-dependent injection rates, the Coulomb-based models should consider updating receiver populations and secondary triggering using unfolding aftershocks.

This poses a grand challenge for future fully prospective model testing: do we have the capability to calibrate our real-time models fast enough on such a large quantity of high-quality data?

6.3 Final Remarks

In conclusion, although much work is needed to understand the physical mechanisms governing aftershock sequences and how to implement them in earthquake forecast models, the development of OEF has reached a critical turning point.
The development of artificial intelligence techniques such as machine learning algorithms promises to deliver an unprecedented availability of data to clarify the many aspects of model performance that are still unresolved and to reduce those crucial data gaps that are currently documented in earthquake physics (*Ben-Zion*, 2019). Our ability to produce increasingly informative short-term earthquake forecasts will then likely improve with the continuous testing of both physics-based and statistical models on a growing number of case studies around the world.
However, in this fast-evolving research discipline, seeking the sole availability of large quantities of input data to improve their predictive skills may result in a vain effort if we do not have a clear understanding of what kind of information is needed for this to happen and to be delivered in near real time. In this context, the path of this thesis can be used to determine the required type of data products that can critically enhance the predictive power of operational earthquake forecast models.

Appendix A

Supplemental Material to Chapter 3

In Text A.1 we explain the reasons for testing of an alternative coefficient of friction and comment on the results.
Figure A1 presents the focal mechanisms of the CRS learning phase overlaid by the DISS database, used for the development of the spatially variable distribution of receiver faults, shown in Figure A2. In Figure A3 we show the ETAS learning phase seismicity used for the MLE inversion procedure. Figure A4 displays the frequency-magnitude distribution for the CRS learning phase catalogue (1990-2016) as well as the Gutenberg-Richter (GR) fit and its magnitude of completeness (M_c). Figure A5 presents the M_c vs. time during the AVN sequence (testing phase).
In Figure A6 we display the stress change maps following the Mw=6.5 Norcia mainshock for preliminary and more complex fault representations as causative and receiver planes. In Figure A7 we report the location of the M $\geq$ 4.0 events within the pattern of calculated Coulomb stress changes after the Mw=6.0 Amatrice and Mw=6.5 Norcia mainshocks. From Figure A8 to Figure A15 we show the complete set of first-day, period-specific and 1-year forecast maps for the 8 forecast models. Figure A16 to A19 provide the full extent of performance evaluation metrics for the whole set of forecast models illustrating the incremental N-test for a 1-year evaluation phase (A16) and period-specific maps of spatially resolved ΔLL (A17-A19). In Figure A20 we show how the implementation of a lower coefficient of effective friction affects the performance of CRS-7.

A.1 Testing a Lower Effective Coefficient of Friction

For all the CRS forecast models presented in this work we set a constant average value of the effective coefficient of friction (μ' = 0.4). However, high pore pressure due to fluid migration is known to play a primary role on the spatiotemporal seismicity patterns of Central Apennines (*e.g. Malagnini et al.*, 2012; *Chiarabba et al.*, 2018). Circulation of fluids at depth causes a reduction of the effective coefficient of friction and an overall effect of strength reduction on faults patches. To account for this fault behaviour, we produce an additional model (CRS-7b) to test whether or not the best performing CRS version further benefits from the implementation of a lower coefficient of friction (μ' = 0.2). Results from the S-test joint log-likelihood scores (Figure A20a) show how the inclusion of a lower μ' exerts a marginal influence on the spatial consistency of the model, with a slight performance degradation ($\Delta jLL_S < 130$). Similarly to the jLL_S scores, the information gains of CRS-7b on CRS-7 (IG_{CRS-7}) present slightly negative values for all the testing periods ($IG_{CRS-7} \approx$ -0.1; Figure A20, b-e). These results are in agreement with those by *Cattania et al.* (2014), who showed how CRS forecasts performance is only modestly influenced by the variation of the effective coefficient of friction.

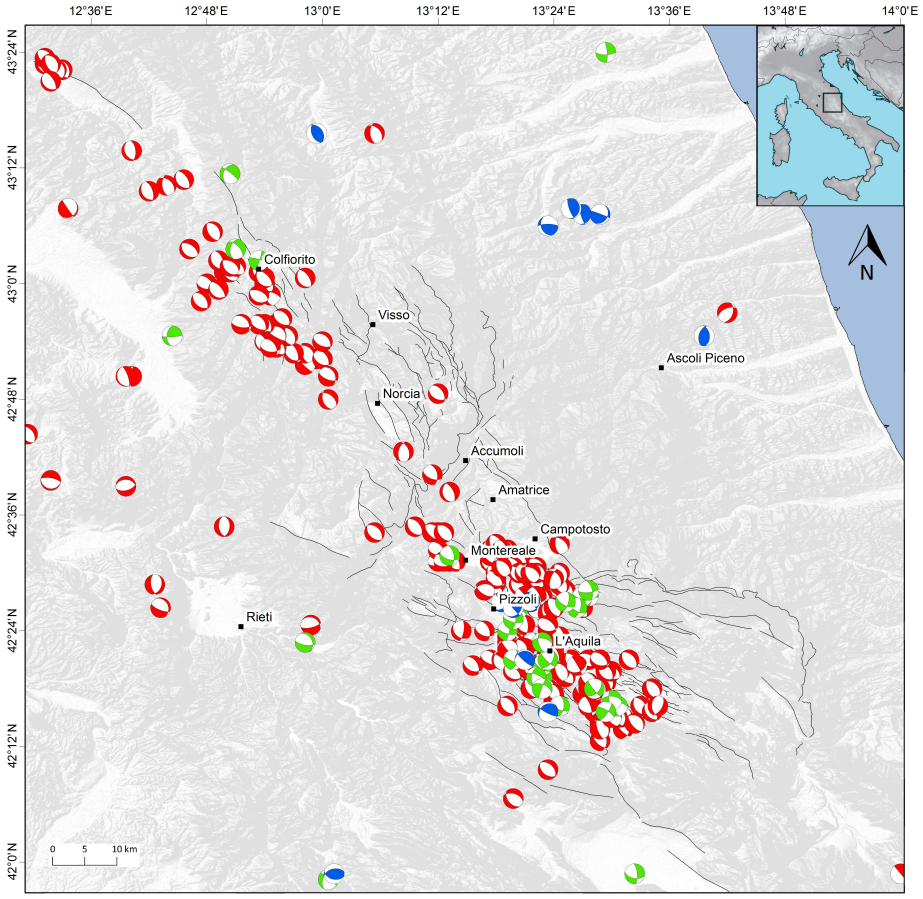

Figure A1: Focal mechanisms (M3+) for the learning phase catalogue (January 1990 - August 2016) within the testing region (longitude 12.5°-14°, latitude 41.95°-43.45°, depth 0-12 km). The colours indicate the prevalent rupture style: red for normal faults, blue for reverse and green for strike-slip ruptures. Black lines show the mapped active faults in the region (*EMERGEO Working Group*, 2016).

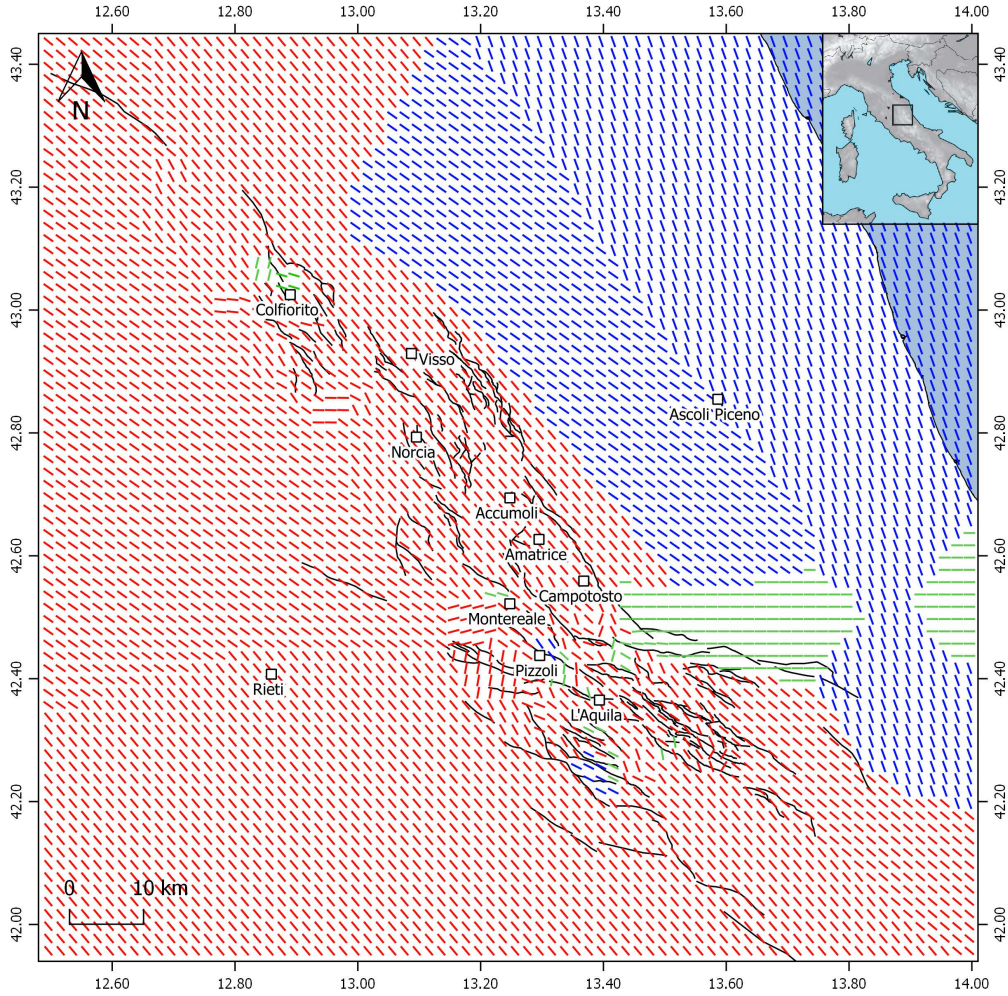

Figure A2: Map of discrete receiver planes in the testing region. Each line represents a hypothetical fault trace, with colours highlighting the faulting style: red for normal sense of slip, blue for thrusts, and green for strike-slip faults. Black lines show the mapped active faults in the region (*EMERGEO Working Group*, 2016). Geometries have been assigned according to prevailing deformation styles of regional faults and focal mechanism catalogues (Figure A1) available during the learning phase following the approach of *Segou et al.* (2013). We note that we implement a 3D grid, which is here collapsed in 2D for representation reasons. The spatial association reveals predominantly normal, SW dipping faults, with antithetic normal structures in the south-western sector and strike-slip faults south and east of Campotosto, while the north-eastern part features reverse faulting of the outer compression front of the Apennines (*DISS Working Group*, 2018). The role of inherited compressional cross-structures contributing to the complex segmentation patterns and controlling fault reactivation is critical (*Buttinelli et al.*, 2018, *Chiarabba et al.*, 2018).

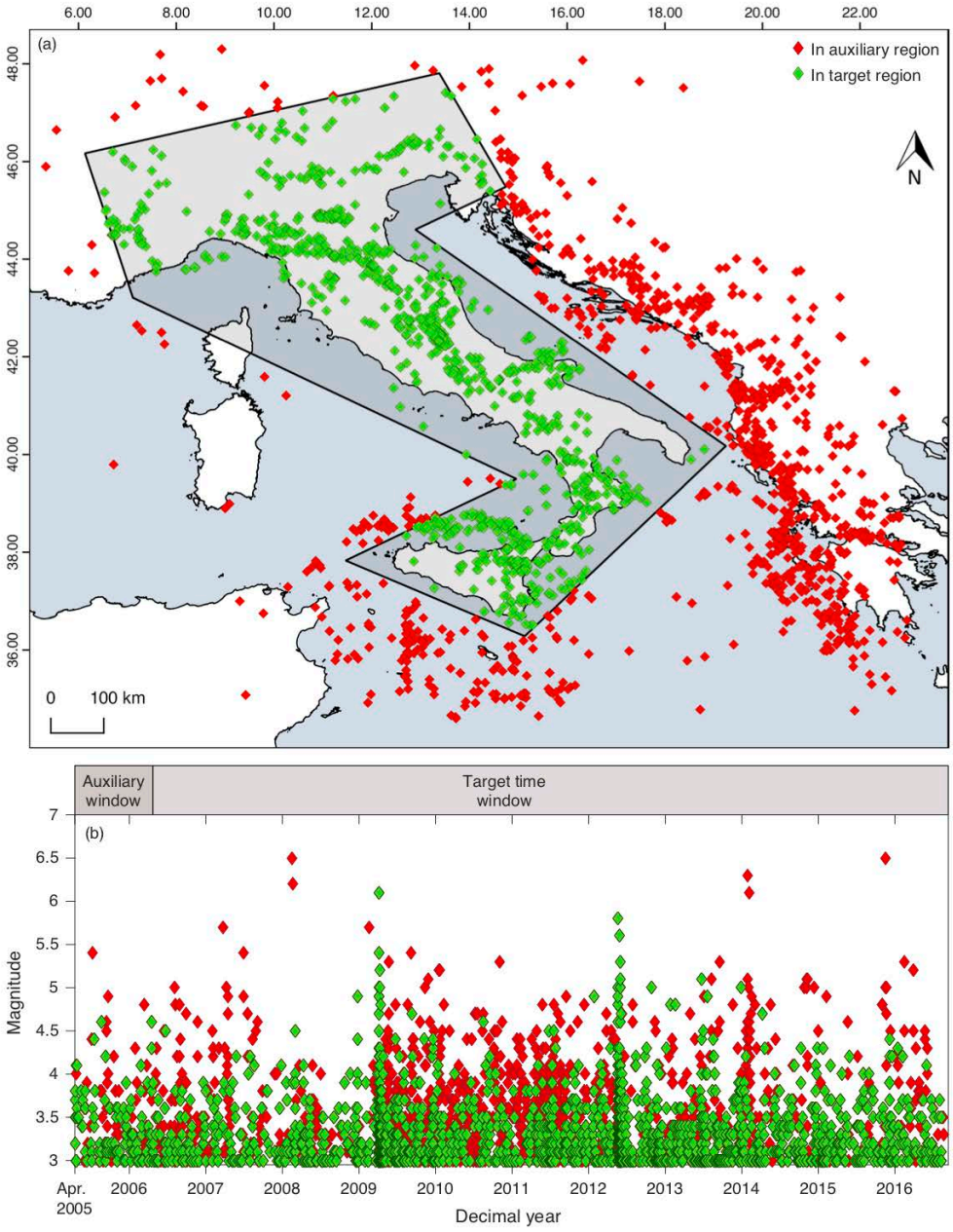

Figure A3: Seismicity (M3+) from April 2005 to August 2016 used for the MLE estimation of the ETAS parameters. (a) Seismicity map and target region polygon; in order to avoid temporal and spatial boundary effects, we consider a wider auxiliary region (red events) to include triggering effects from outside the target polygon. (b) Time-magnitude plot; we take the first year of seismicity as auxiliary time window for parameters inversion.

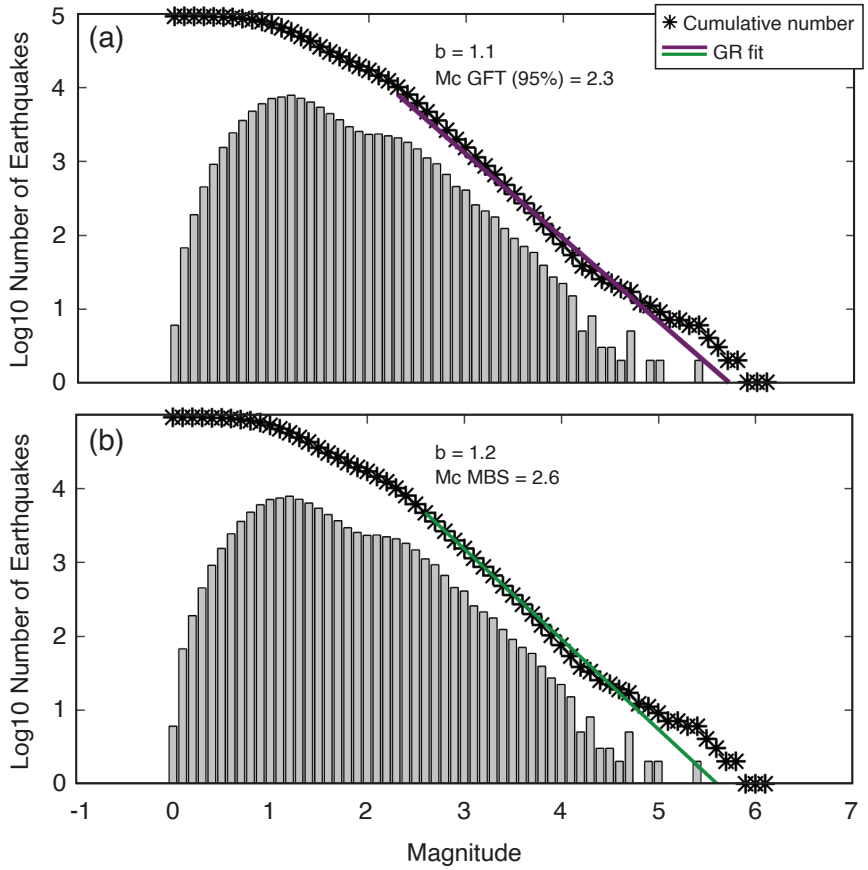

Figure A4: Frequency-magnitude distribution for the CRS learning phase catalogue (January 1st 1990 - August 23rd 2016) within the testing region. We estimate b-value and M_c, using (a) b-value stability (MBS; *Cao & Gao*, 2002; *Woessner & Wiemer*, 2005) and (b) goodness of fit (GFT; *Wiemer & Wyss*, 2000) method. We use a conservative M_c=3.0 to account for early aftershock incompleteness within the AVN sequence immediately after the primary mainshocks (See Figure A5).

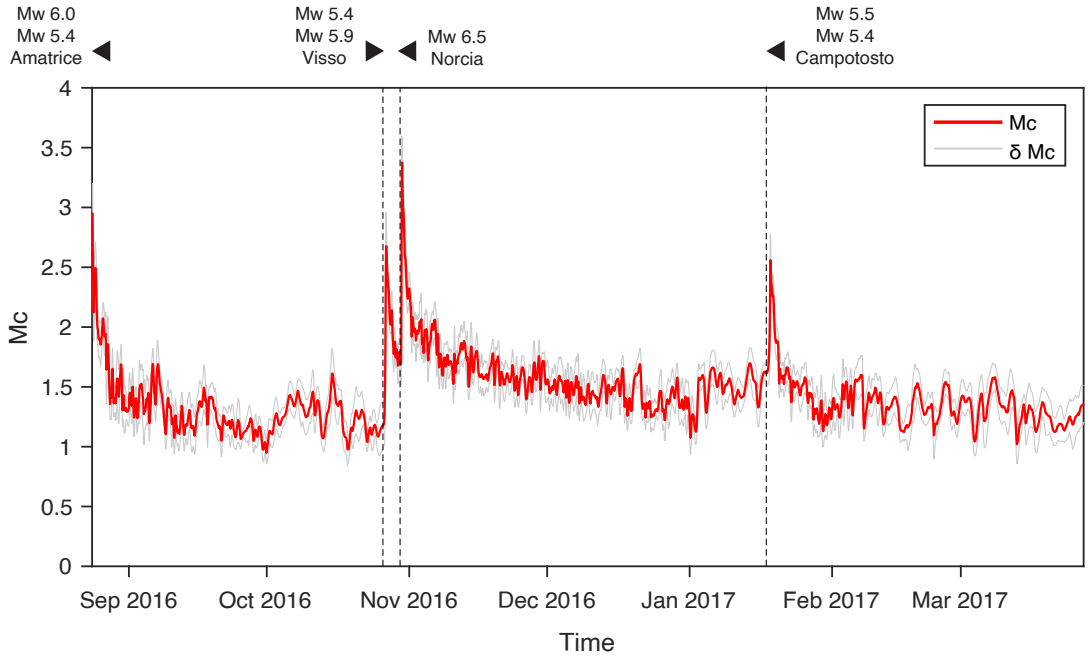

Figure A5: Magnitude of completeness over time for the AVN sequence. We present the temporal variation $M_c(t)$ of the completeness magnitude of the real-time catalogue for approximately 7 months of the Amatrice-Visso-Norcia sequence. We use the Maximum Curvature technique (*Wyss et al.* 1999; *Wiemer & Wyss*, 2000), with a sampling window of 200 earthquakes (minimum number of events = 50) and 200 bootstrap samples. The vertical dashed lines mark the occurrence of the primary events within the sequence. $M_c(t)$ calculation routines are implemented within the open source software ZMAP (*Wiemer*, 2001).

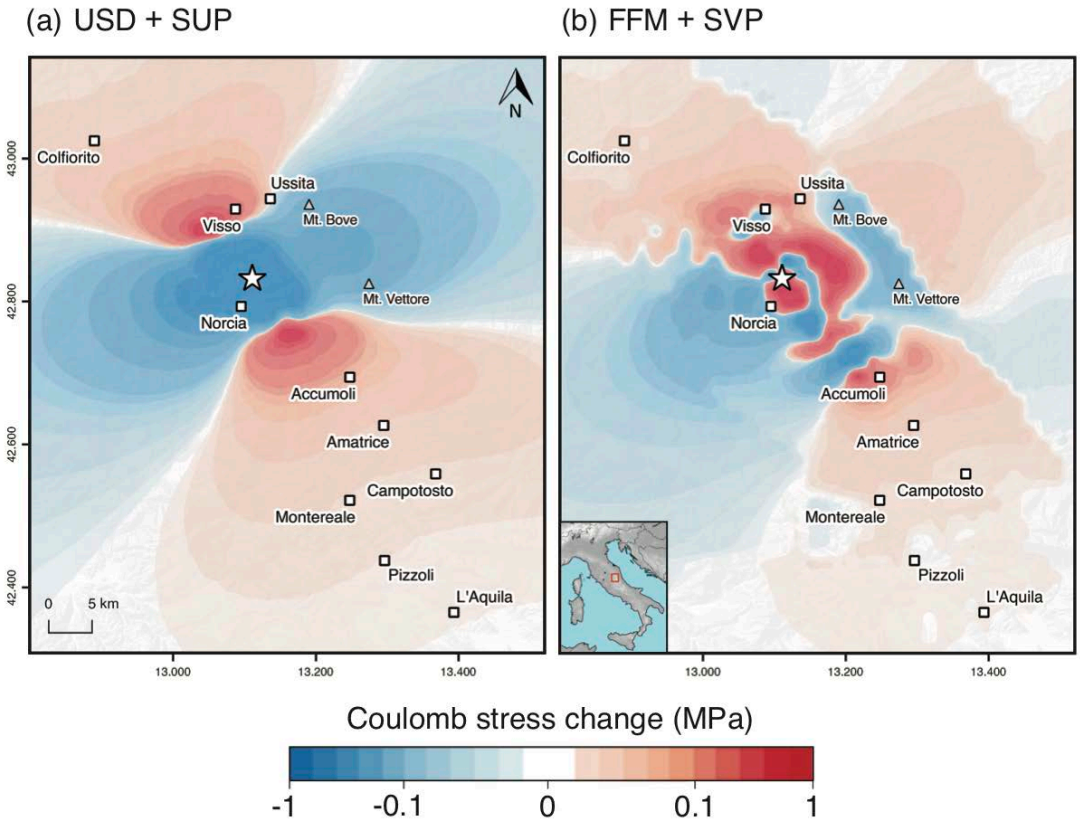

Figure A6: The effect of receiver planes and variable slip distribution on Coulomb stress change calculations. We show the coseismic Coulomb stress changes following the Mw=6.5 Norcia event at a reference depth of 9 km. The coefficient of effective friction (μ') is set to 0.4. USD = uniform slip distribution; SUP = spatially uniform receiver planes; FFM = finite-fault slip model; SVP = spatially variable planes.

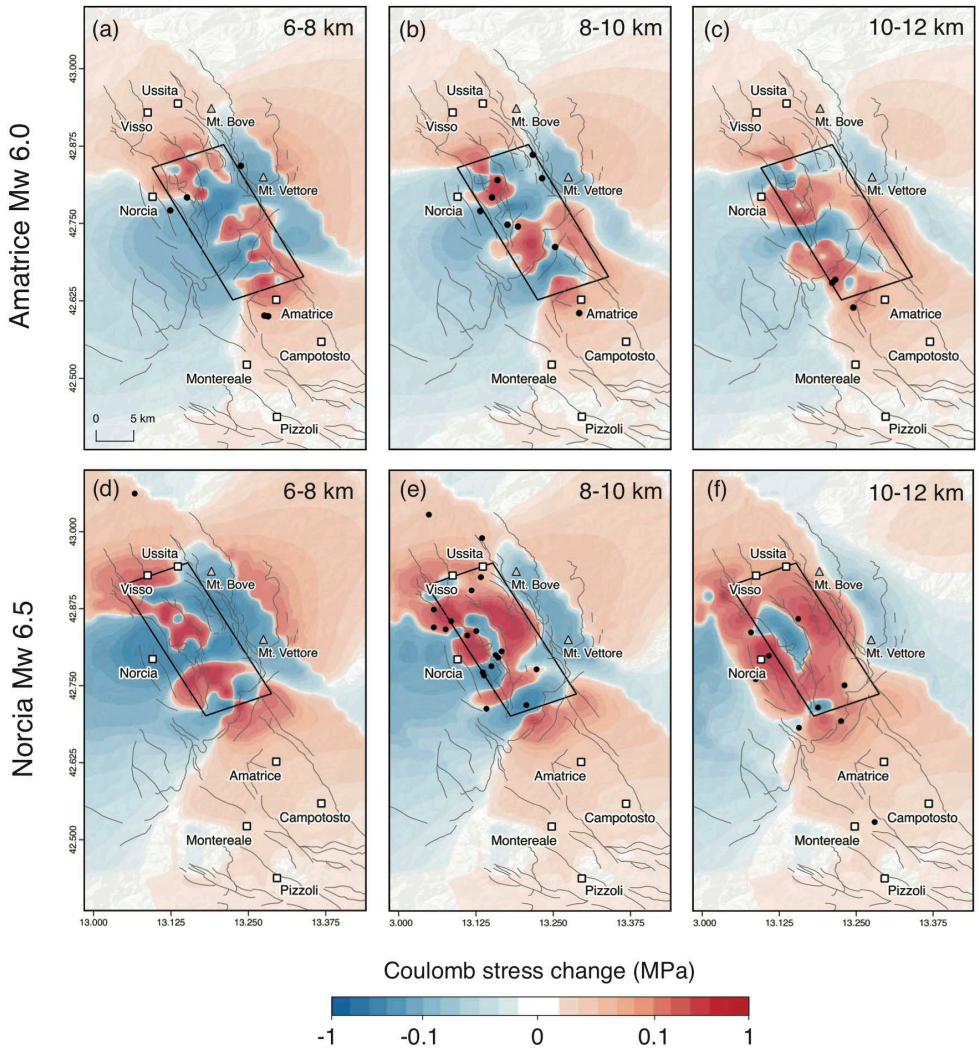

Figure A7: Coulomb stress changes used in CRS models. We calculate the coseismic stress changes for a depth between 0-12 km, which represents the average seismogenic thickness for moderate-magnitude earthquakes in the Central Apennines (*Chiarabba et al.*, 2005; *Chiarabba & De Gori*, 2016). Here, the Coulomb stress changes at selected depths (7, 9, 11 km) are overlaid by the epicentral locations of the M4+ triggered events (black circles) occurred between the Mw=6.0 Amatrice and Mw=5.4 Visso I earthquakes (a-c) and between the Mw=6.5 Norcia mainshock and the first of the Campotosto events (d-f). Black solid lines enclose the surface projections of the mainshock faults. We note that ∼70% of the M4+ aftershocks after the Amatrice and Norcia events occur in the already stress-increased near fault zones.

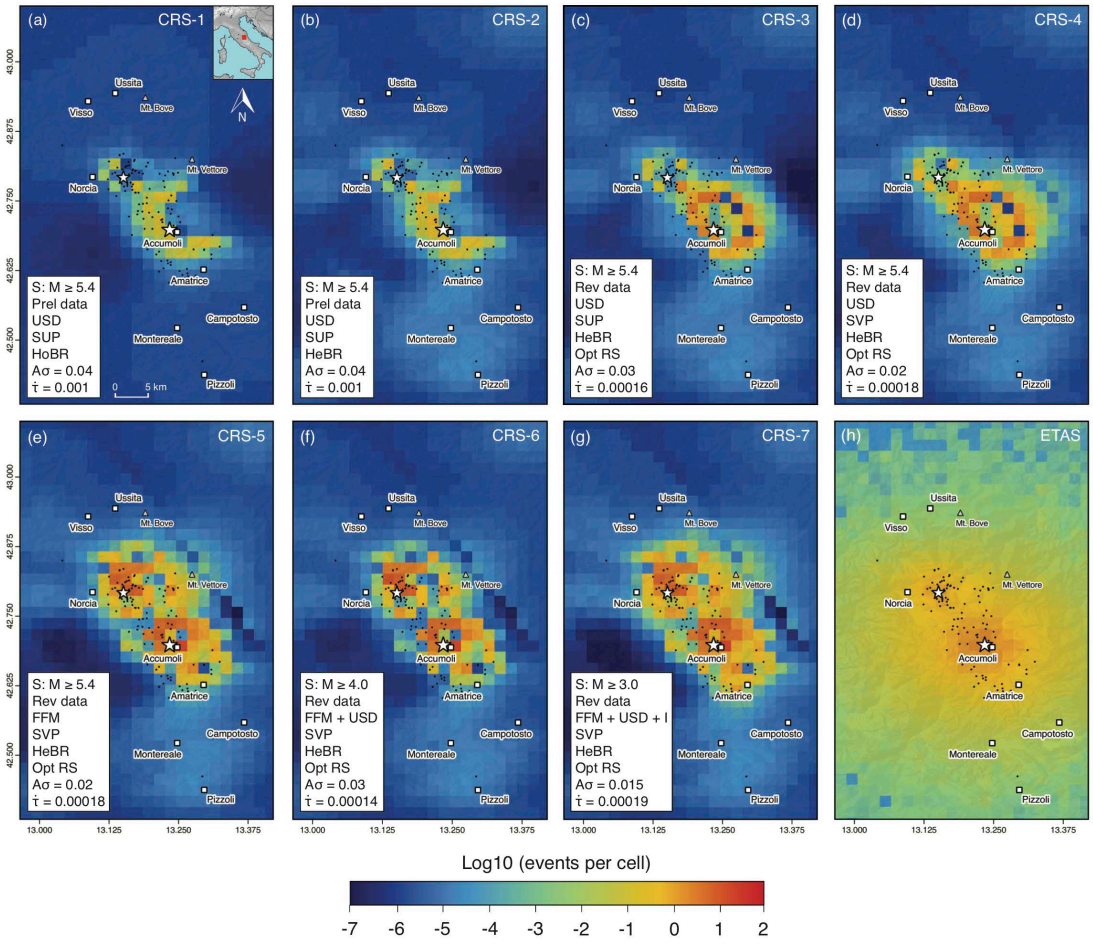

Figure A8: 24-hour forecast maps following the Mw=6.0 Amatrice earthquake (24 August 2016, 01:36:32 UTC). Black circles indicate the observed events, while white stars indicate the M5+ sources within the same time window. S = sources; Prel = preliminary; Rev = revised; USD = uniform slip distribution; FFM = finite-fault rupture model; I = isotropic stress field; SUP = spatially uniform receiver planes; SVP = spatially variable planes; HoBR = spatially homogeneous background rate; HeBR = spatially heterogeneous background rate; Opt RS = optimised rate-state parameters. σ values are in MPa, $\dot{\tau}$ values are in MPa/year.

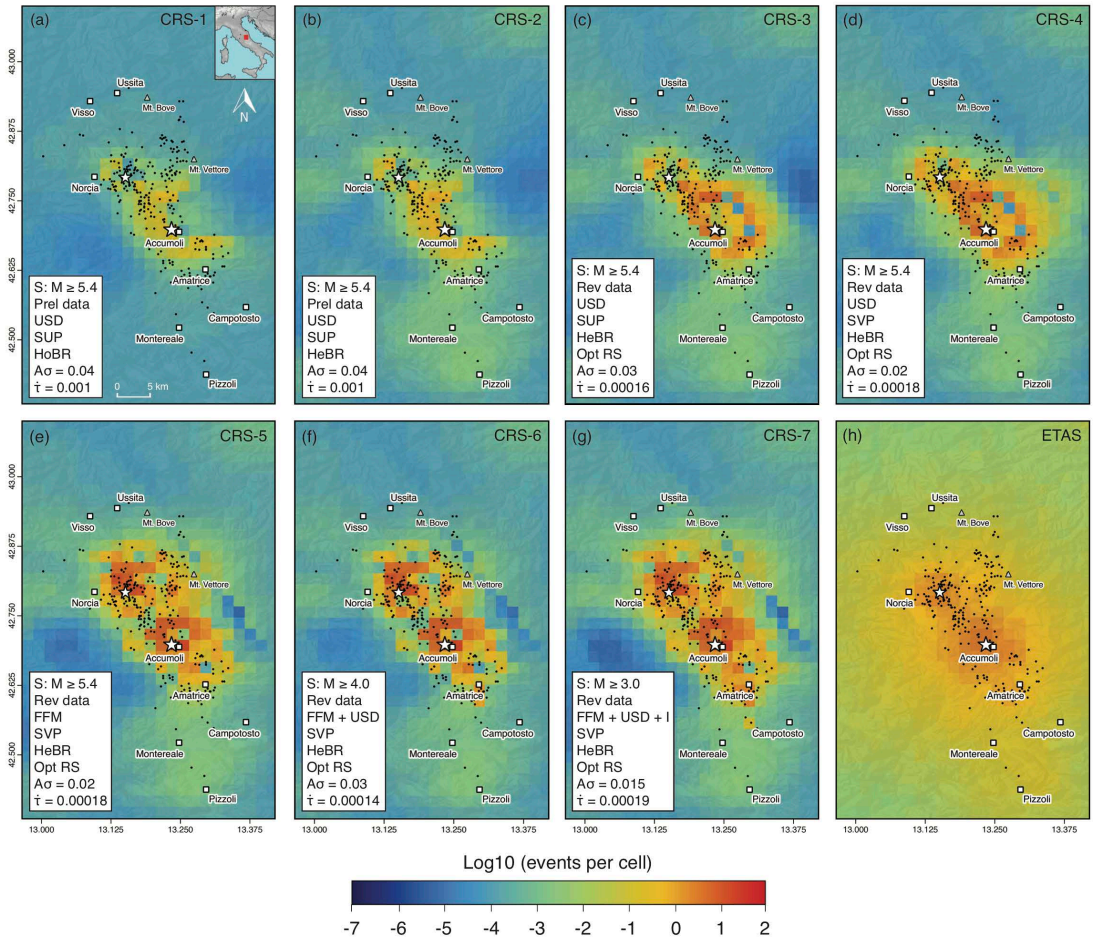

Figure A9: Same as Fig. A8, for the cumulative time window between the Mw=6.0 Amatrice earthquake (24 August 2016, 01:36:32 UTC) and the Mw=5.4 Visso I event (26 October 2016, 17:10:36 UTC).

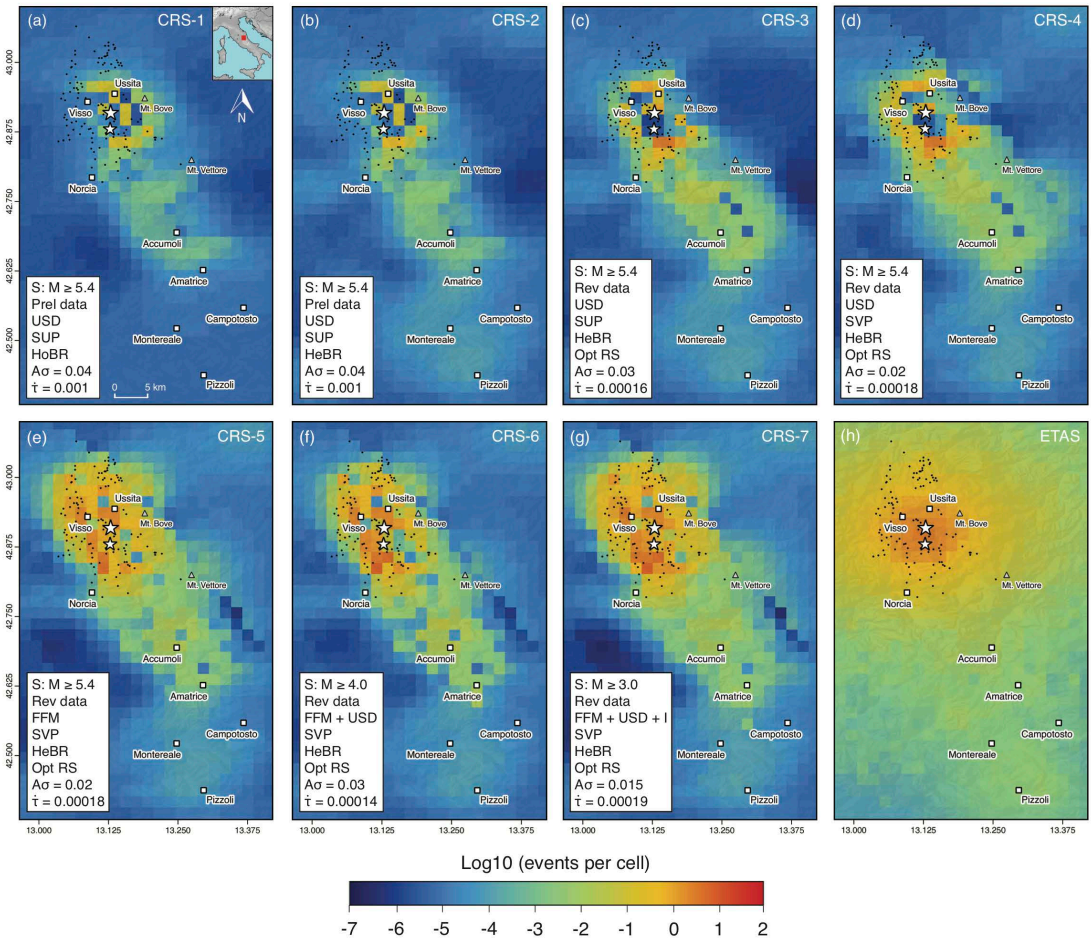

Figure A10: Same as Fig. A8, for the time period between the Mw=5.9 Visso II earthquake (26 October 2016, 19:18:06 UTC) and the Mw=6.5 Norcia mainshock (30 October 2016, 06:40:17 UTC).

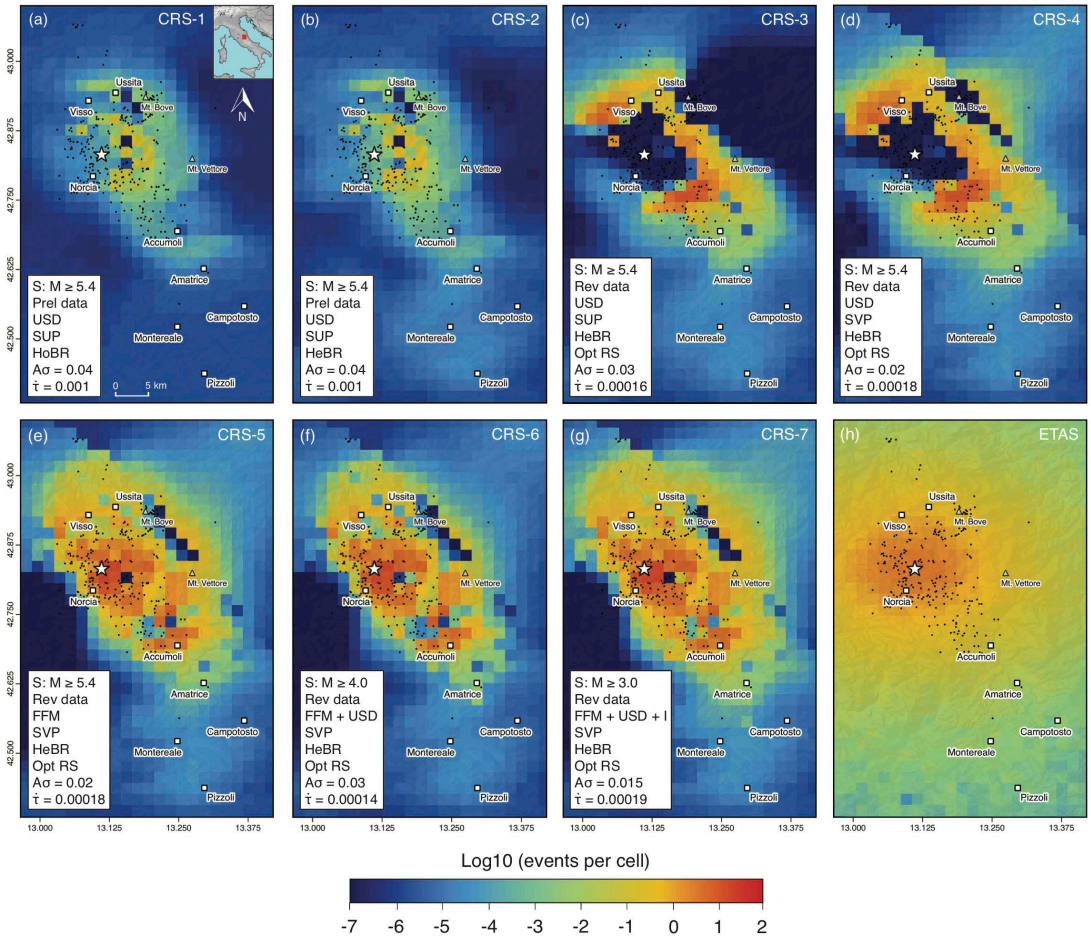

Figure A11: Same as Fig. A8, for the first 24 hours following the Mw=6.5 Norcia earthquake (30 October 2016, 06:40:17 UTC).

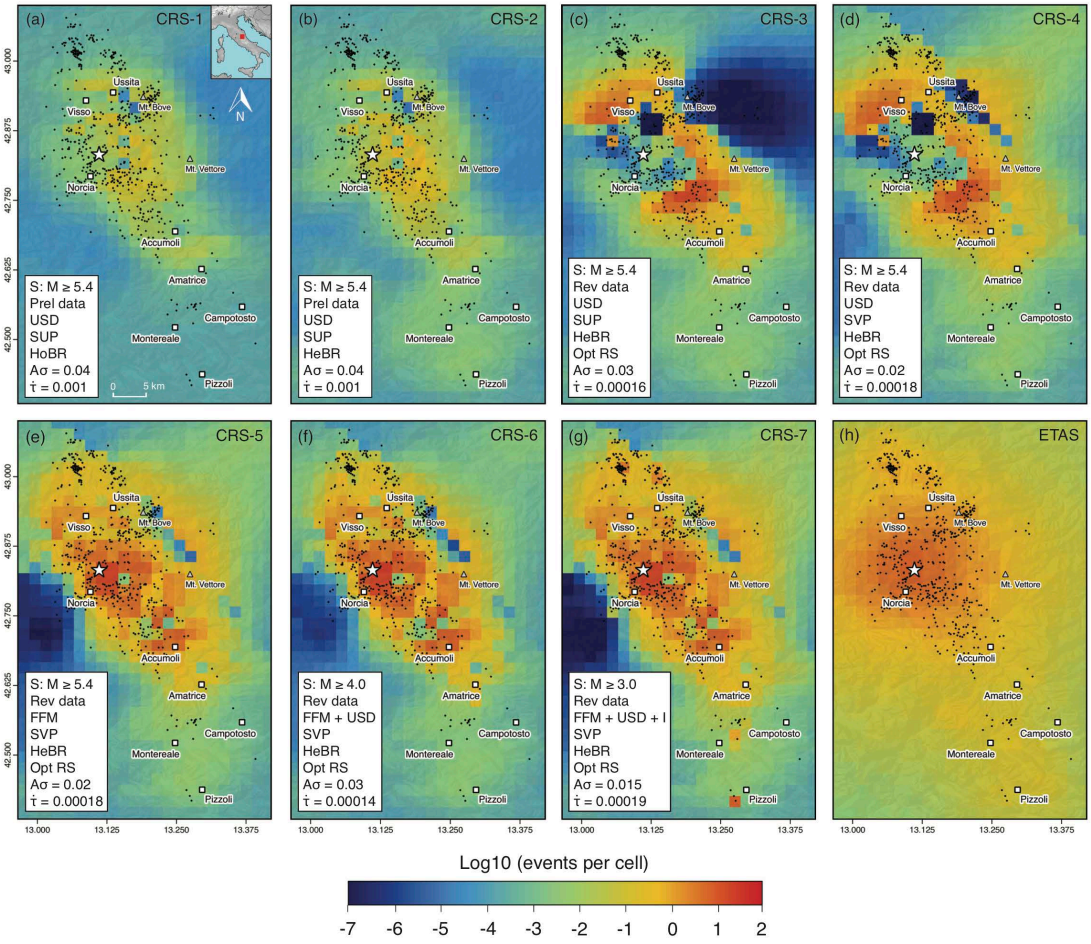

Figure A12: Same as Fig. A8, for the time period between the Mw=6.5 Norcia mainshock (30 October 2016, 06:40:17 UTC) and the first Campotosto event (18 January 2017, 09:25:40 UTC).

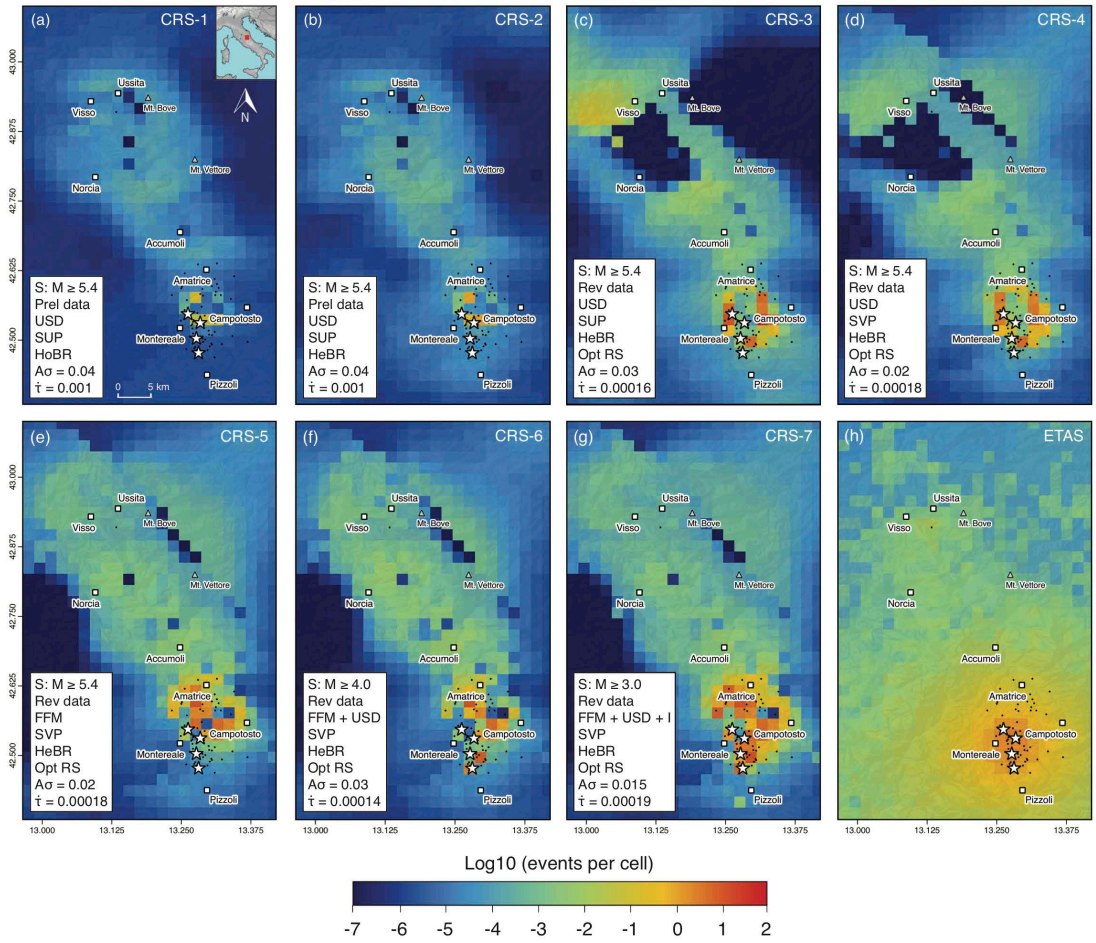

Figure A13: Same as Fig. A8, for the first 24 hours following the Mw=5.1 Campotosto I earthquake (18 January 2017, 09:25:40 UTC).

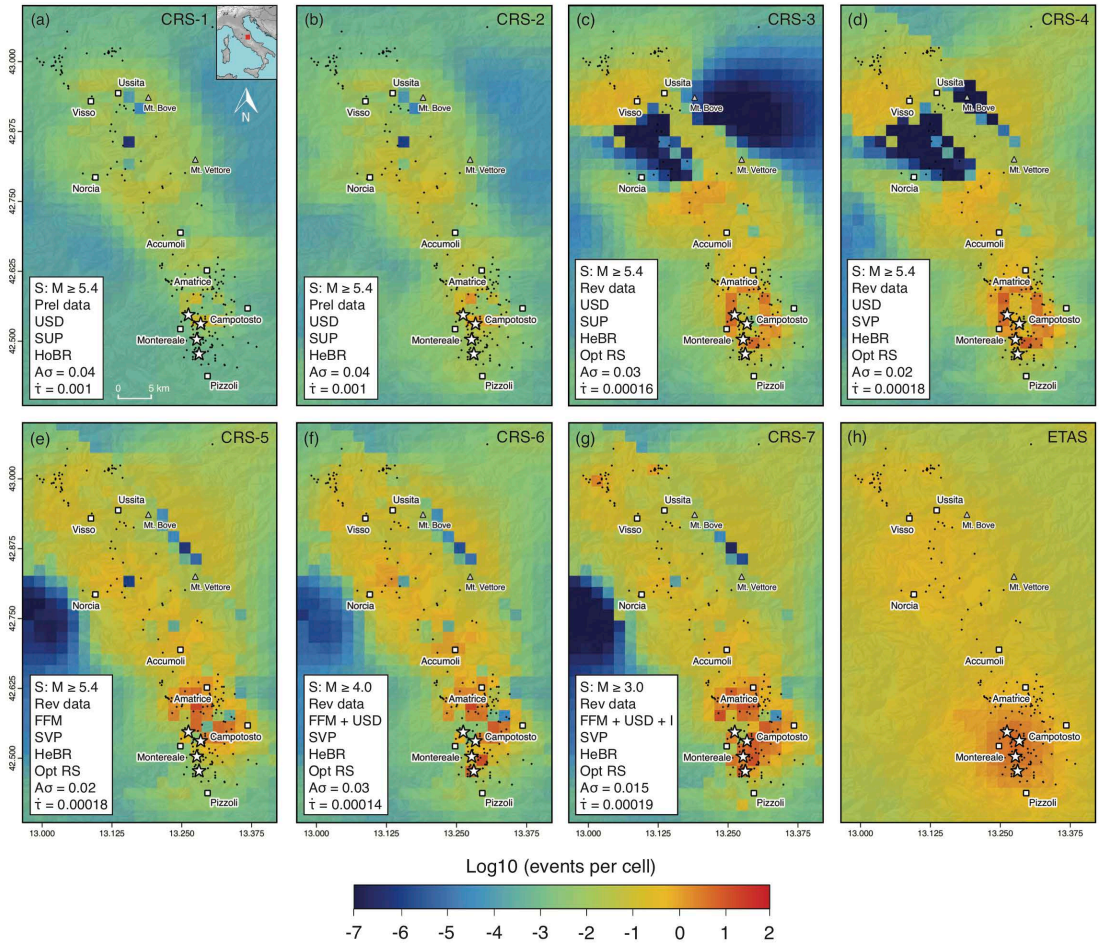

Figure A14: Same as Fig. A8, for the time period between the Mw=5.1 Campotosto I event and the end of the one year testing period (24 August 2017).

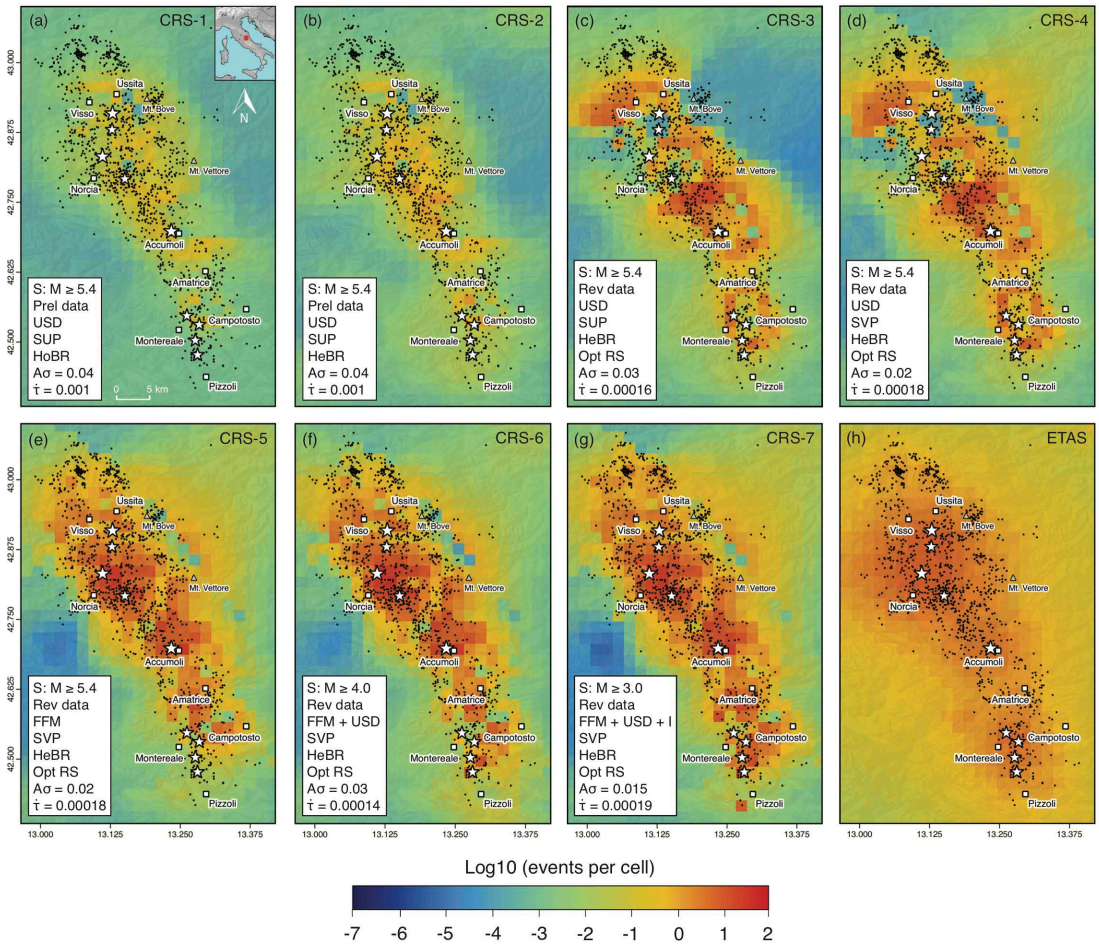

Figure A15: Same as Fig. A8, for the 1-year testing period.

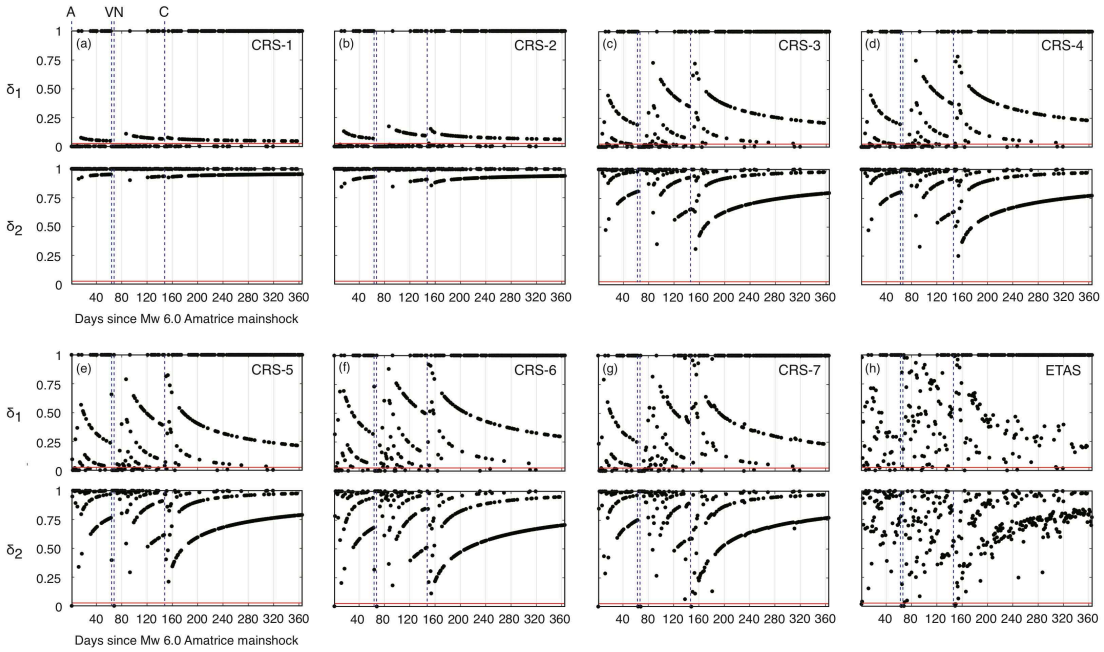

Figure A16: Incremental modified N-test over time. We show the δ_1 (top) and δ_2 (bottom) quantiles for CRS and ETAS models in the whole testing region, within 1-day intervals and for one year following the Mw=6.0 Amatrice mainshock. The red horizontal lines indicate the threshold for model rejection ($\delta_{1,2} < 0.025$). Vertical dashed lines mark the occurrence of the major events (A = Amatrice, V = Visso, N = Norcia, C = Campotosto events). R_N = modified N-test rejection ratio.

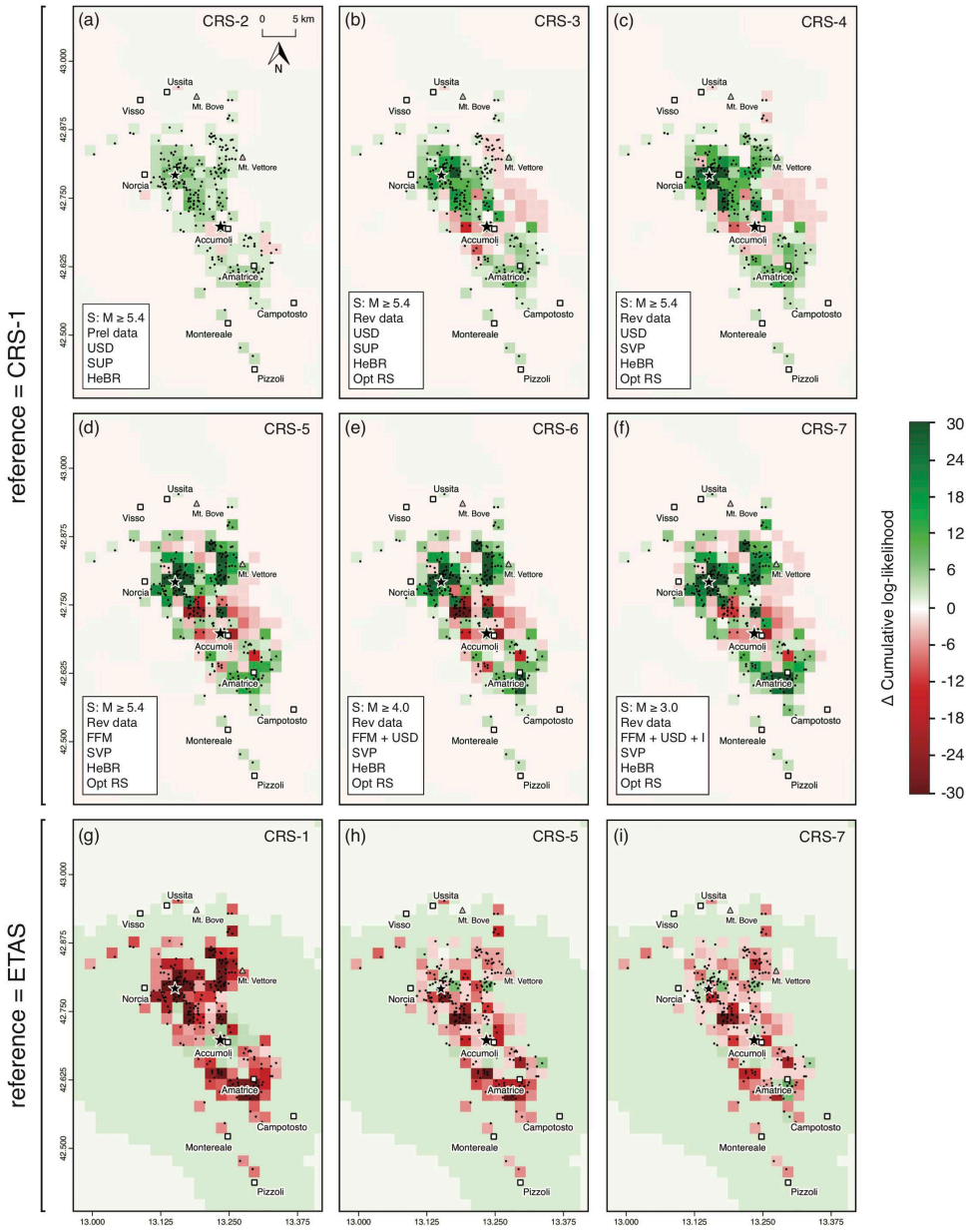

Figure A17: Spatial performance of forecast models between the Mw=6.0 Amatrice mainshock and the Visso events. We portray in each spatial bin the cumulative log-likelihood differences (ΔLL) between pairs of models for the selected evaluation period overlaid by the observed seismicity (circles) and the mainshocks (stars). Log-likelihood values at each spatial bin are obtained summing over all time steps. Positive values in green indicate an improved performance with respect to the reference model. For illustration purposes the ΔLL values are saturated at ± 30.

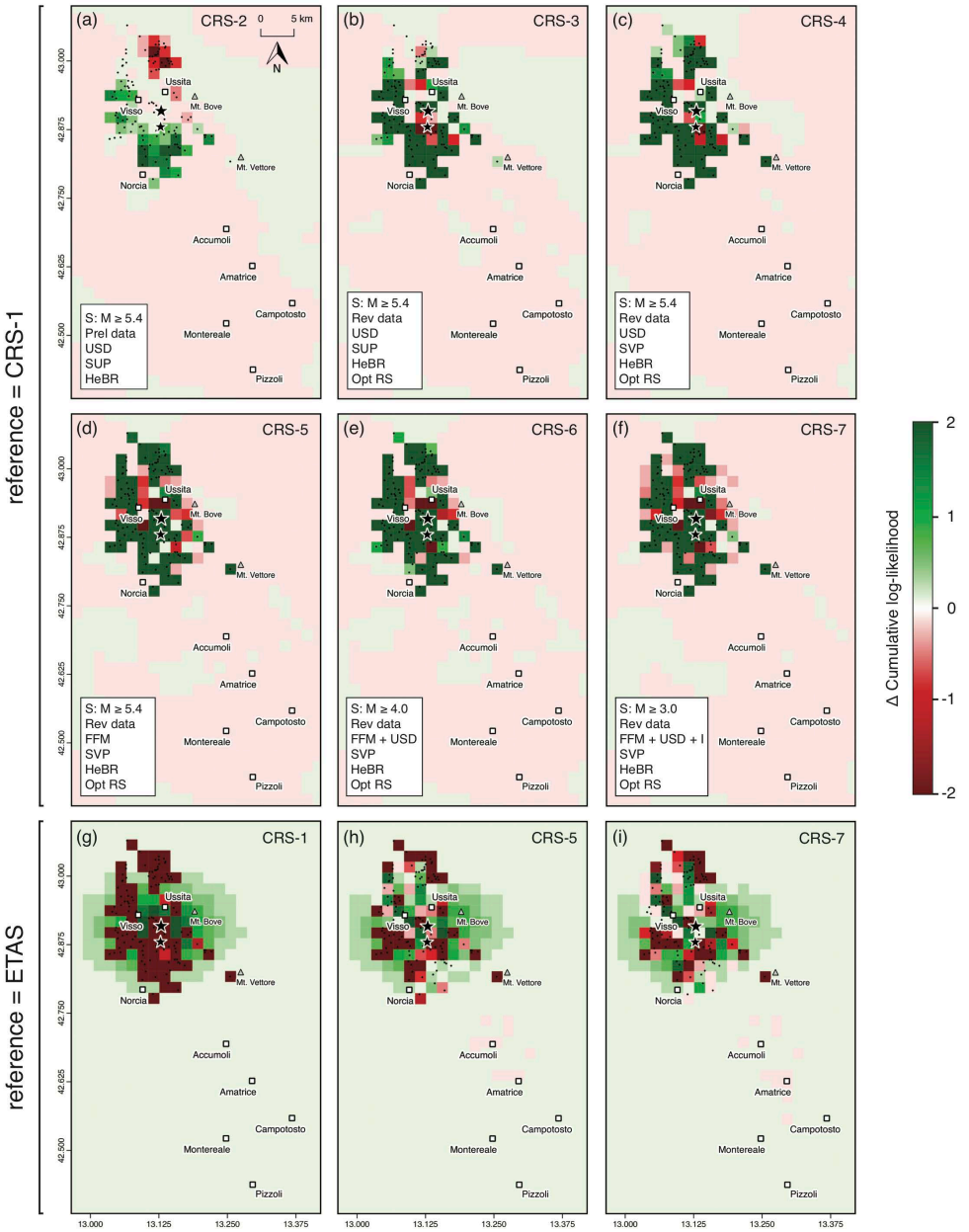

Figure A18: Same as Fig. A17, for the period between the Mw=5.4 Visso I event and the Mw=6.5 Norcia mainshock. Values are saturated at ±2 for visualisation purposes.

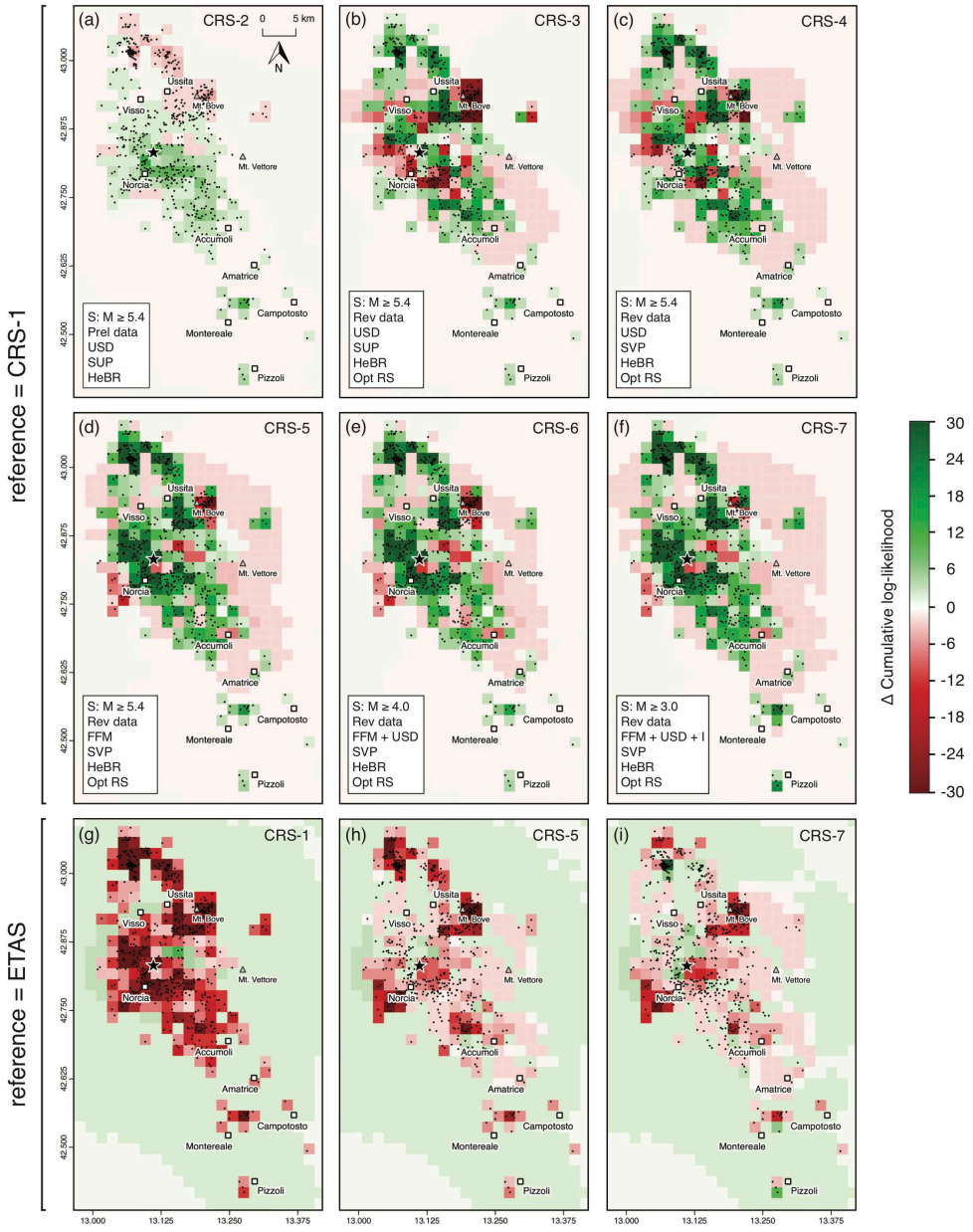

Figure A19: Same as Fig. A17, for the period between the Mw=6.5 Norcia mainshock and the Campotosto events.

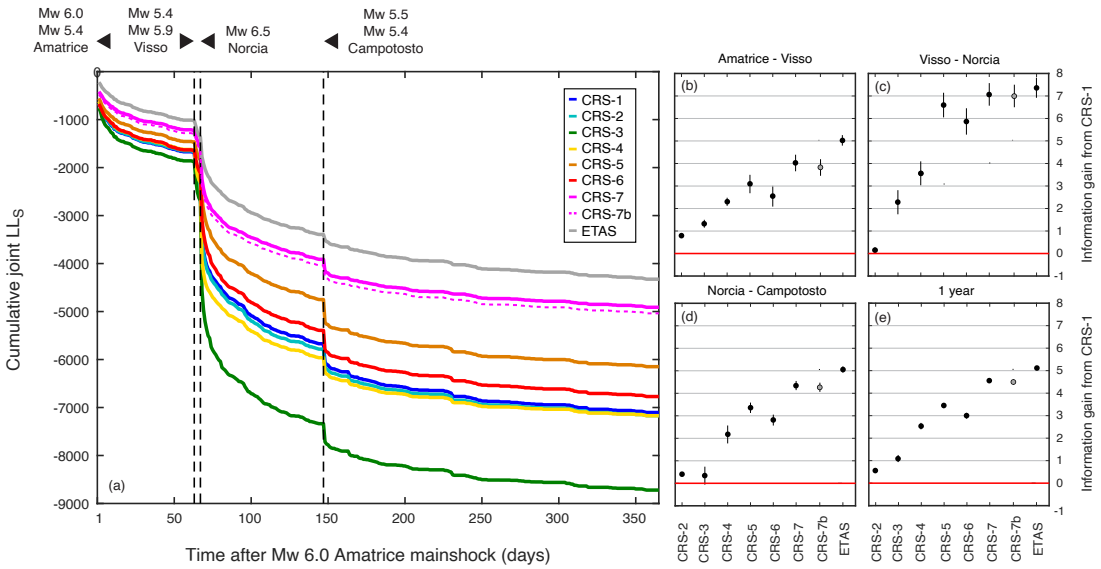

Figure A20: Cumulative joint log-likelihood (jLL_S) versus time (a) and T-test (b-e) including the additional CRS-7b model.

Appendix B

Supplemental Material to Chapter 4

In this appendix we first show details on the surface deformations as represented by InSAR data following the Mw 6.4 Searles Valley and Mw 7.1 Ridgecrest earthquakes (Figure B1). We then provide additional figures for (1) the effect of implementing different values of the coefficient of effective friction on the time performance of the most enhanced CRS model (Figure B2), (2) the complete set of short-term (24 hours following mainshocks) and medium-term (1 month) physics-based and statistical forecasts (Figure B3-B5); (3) the effect of expressing r_0 as either background rate or reference rate on the 1-month forecast of the most enhanced CRS7-*new_FMs* for the whole testing region (Figure B6); (4) spatially resolved ΔLL maps between the two most evolved CRS models and ETAS (Figure B7), and (5) maps of coseismic stress changes when stresses imparted by the combined Mw 6.4 and Mw 7.1 mainshocks are resolved on the average pre-Ridgecrest normal focal mechanism (Figure B8).

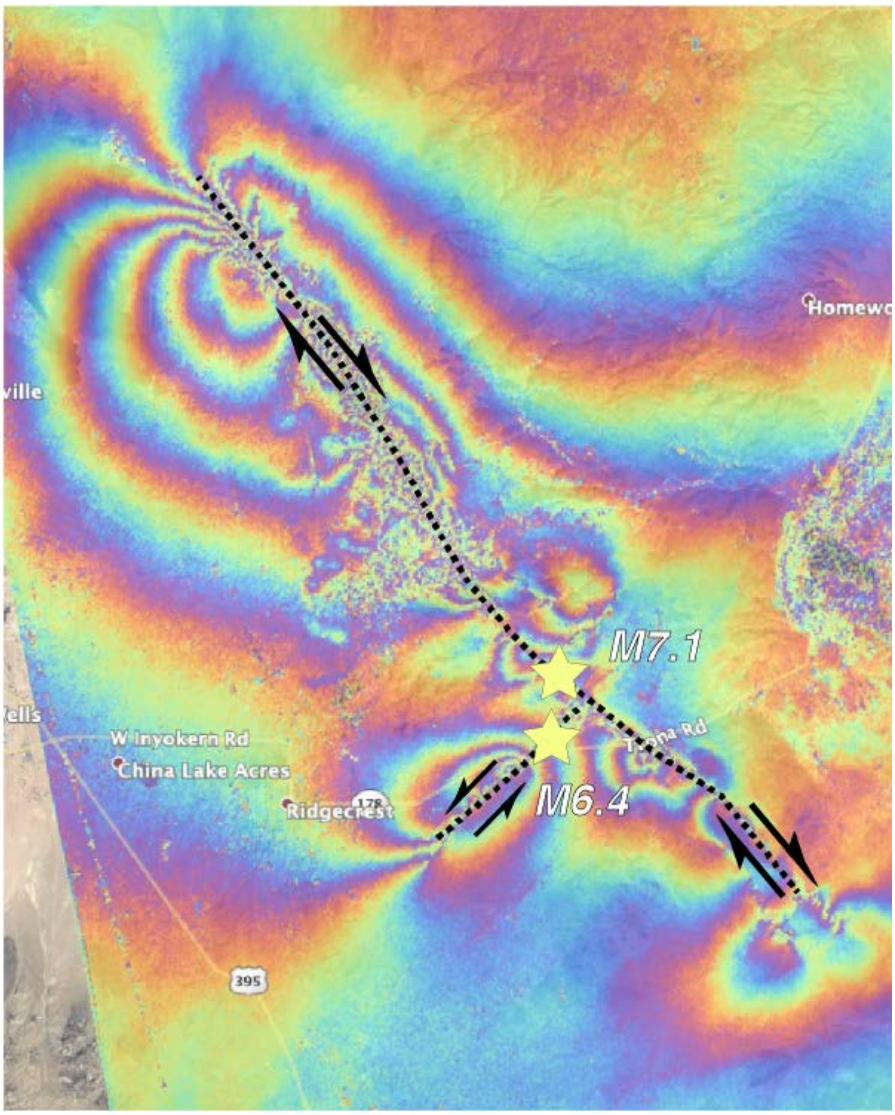

Figure B1: Interferometric Synthetic Aperture Radar (InSAR) map showing surface deformation resulting from the 4 July 2019 Mw 6.4 left-lateral, and 6 July 2019 Mw 7.1 right-lateral earthquakes near Ridgecrest, CA. Source: the Advanced Rapid Imaging and Analysis (ARIA) team at NASA's Jet Propulsion Laboratory and Caltech (`https://www.jpl.nasa.gov/spaceimages/details.php?id=PIA23150`).

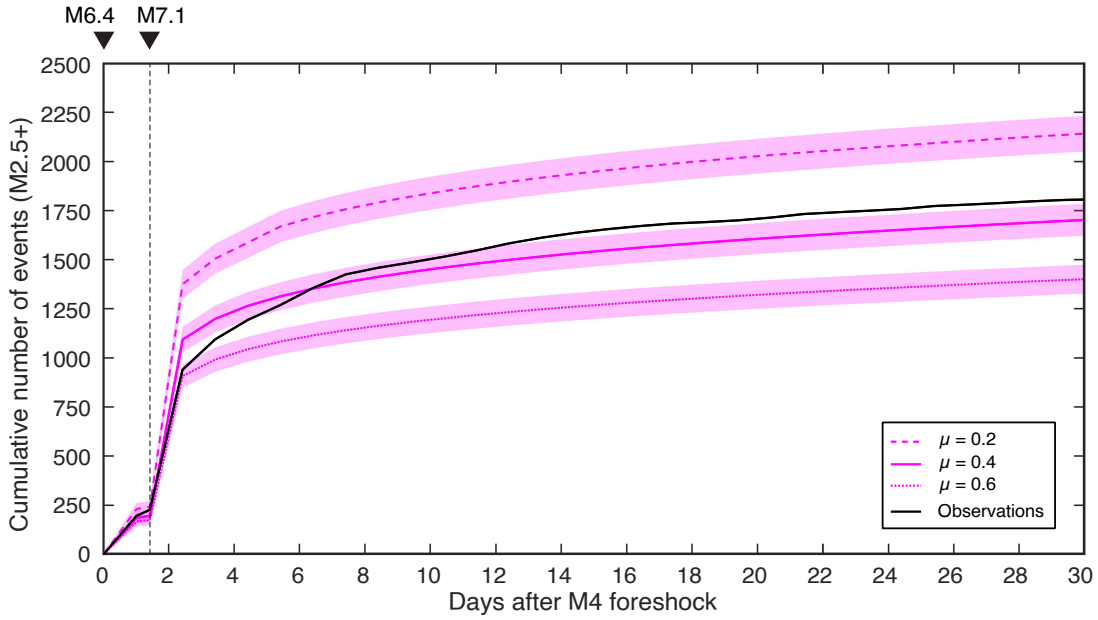

Figure B2: Temporal evolution of the CRS7-*new_FMs* predicted seismicity using different friction coefficients. The black solid line represents the M2.5+ observations, and the shaded areas indicate Poissonian uncertainties.

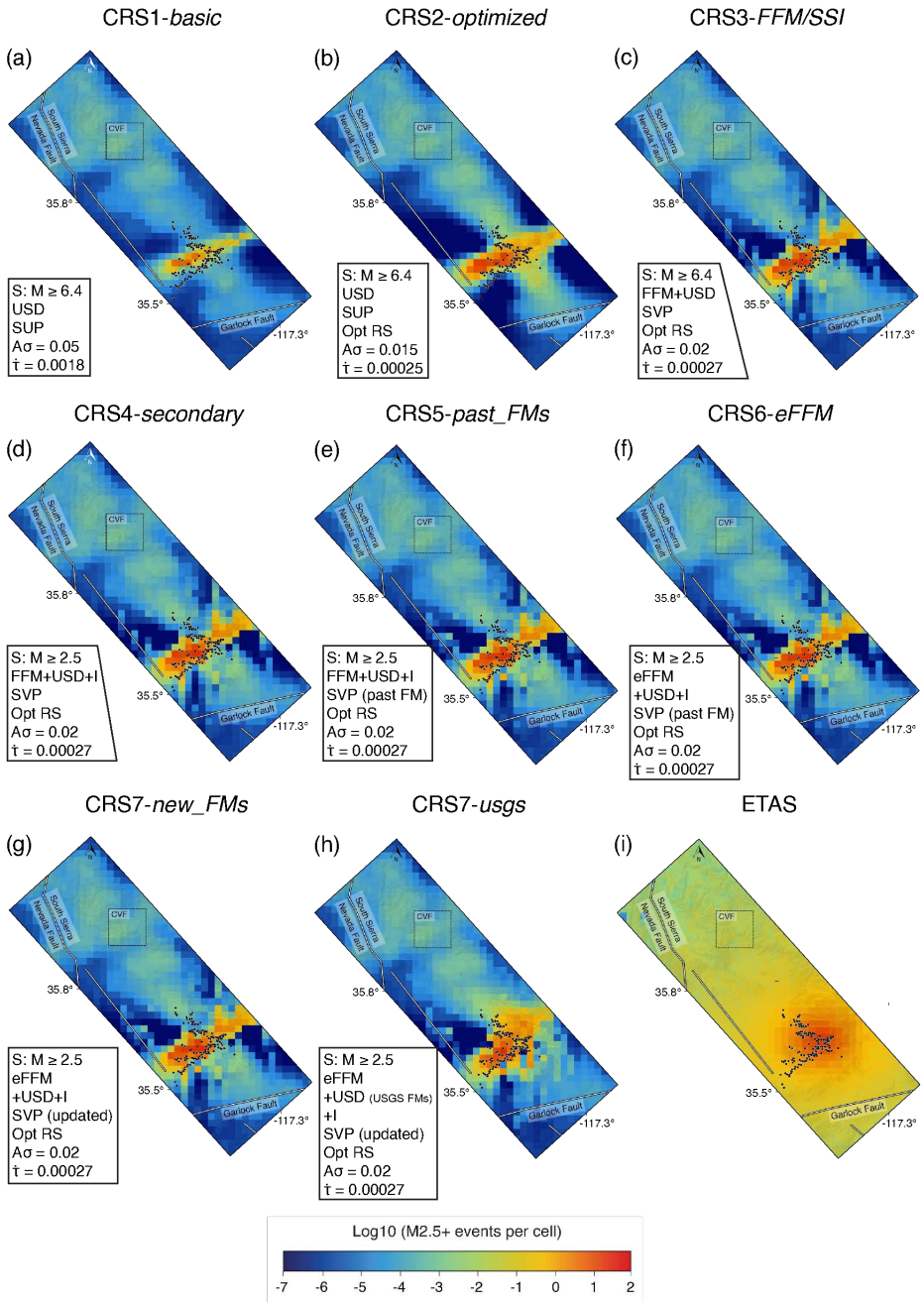

Figure B3: Maps of expected seismicity rates for CRS and ETAS models for the first 24 hours following the Mw 6.4 Searles Valley earthquake in the area of main aftershock productivity. Observed events (M2.5+) are represented as circles. The dashed-line square indicates the area of the Coso volcanic field (CVF). $A\sigma$ values are in MPa, $\dot{\tau}$ values are in MPa/year.

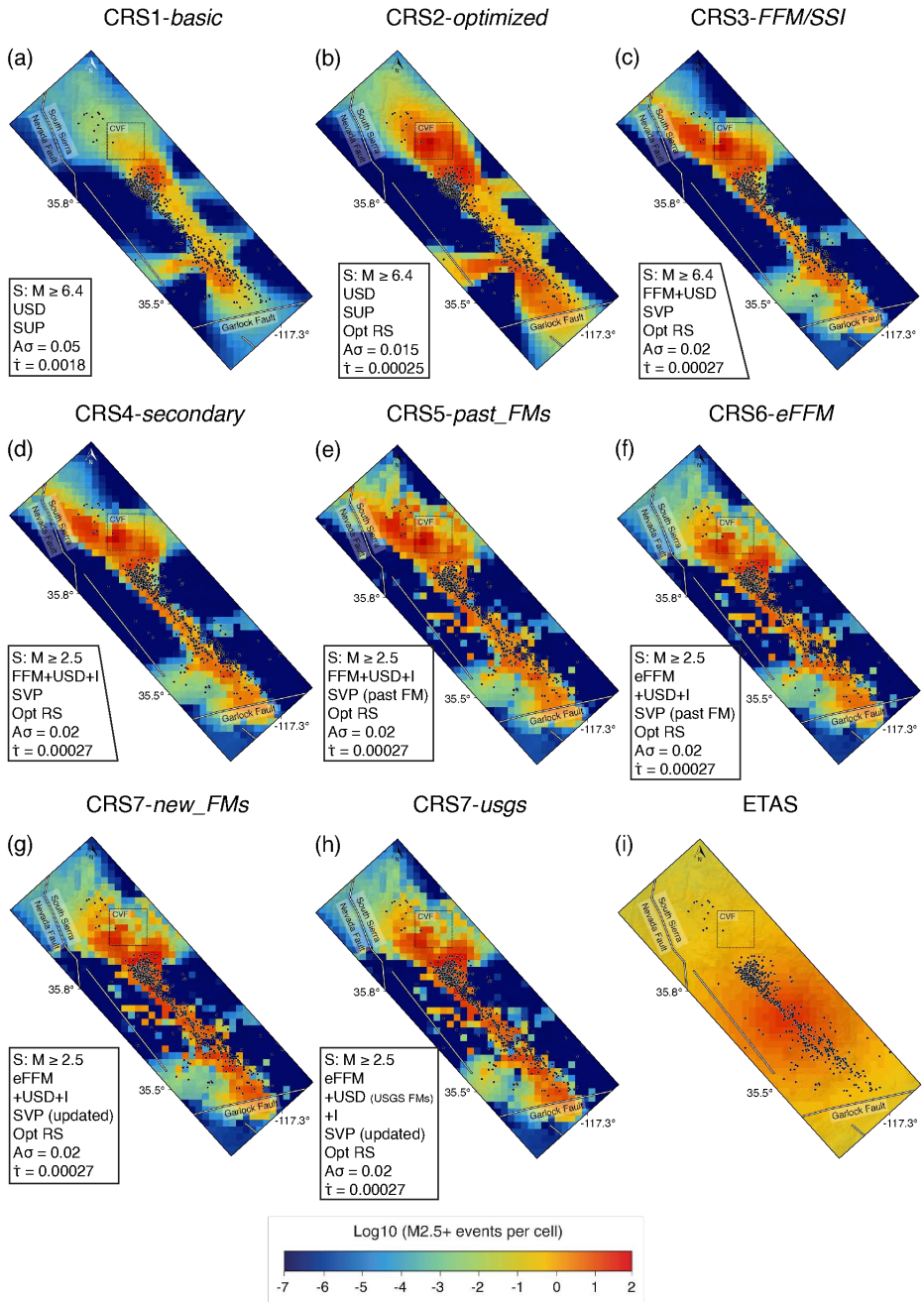

Figure B4: Same as Fig. B3, for the 24-hour time window following the Mw 7.1 Ridgecrest mainshock.

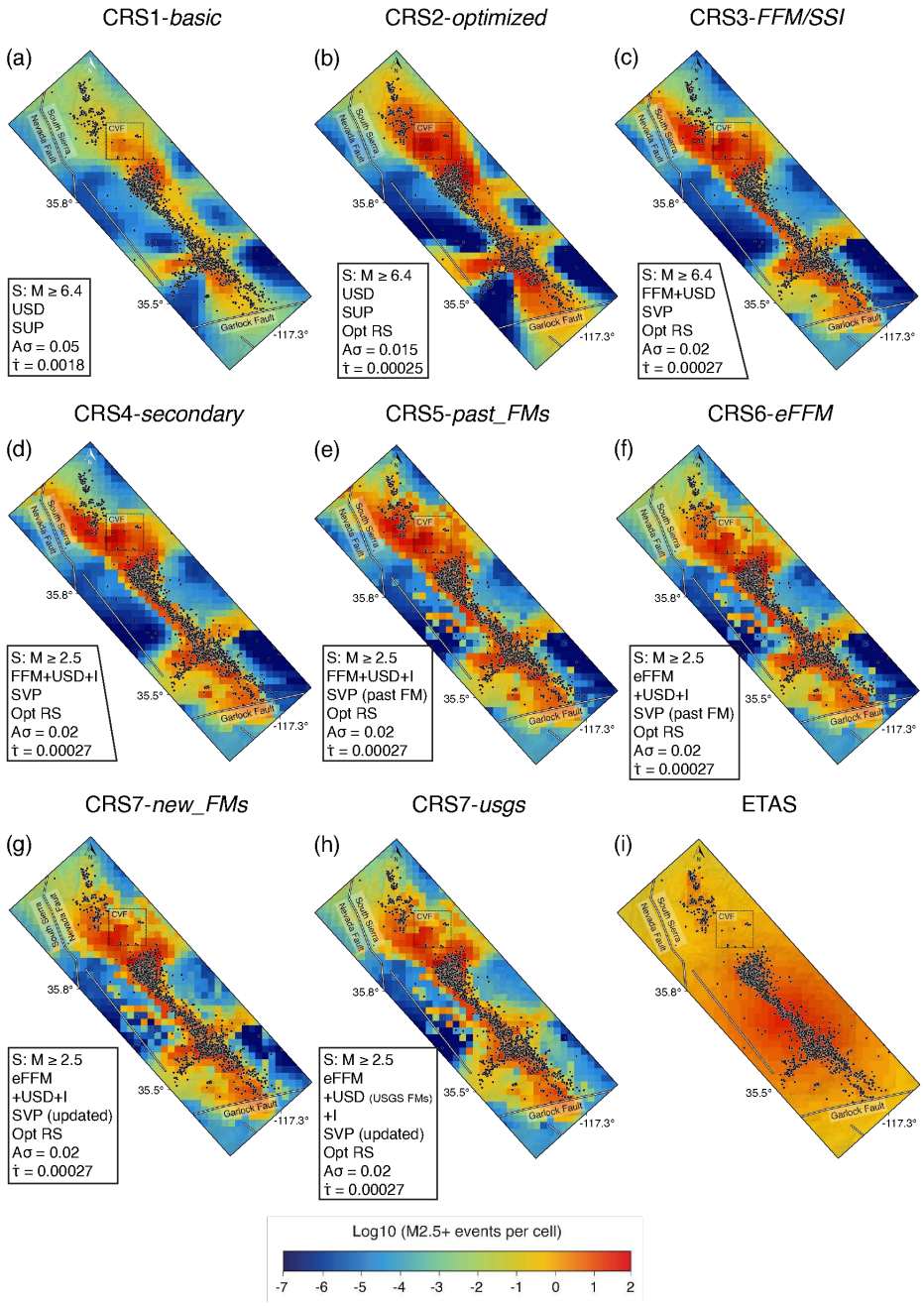

Figure B5: Same as Fig. B3, for the cumulative 1-month time window following the Mw 6.4 Searles Valley earthquake

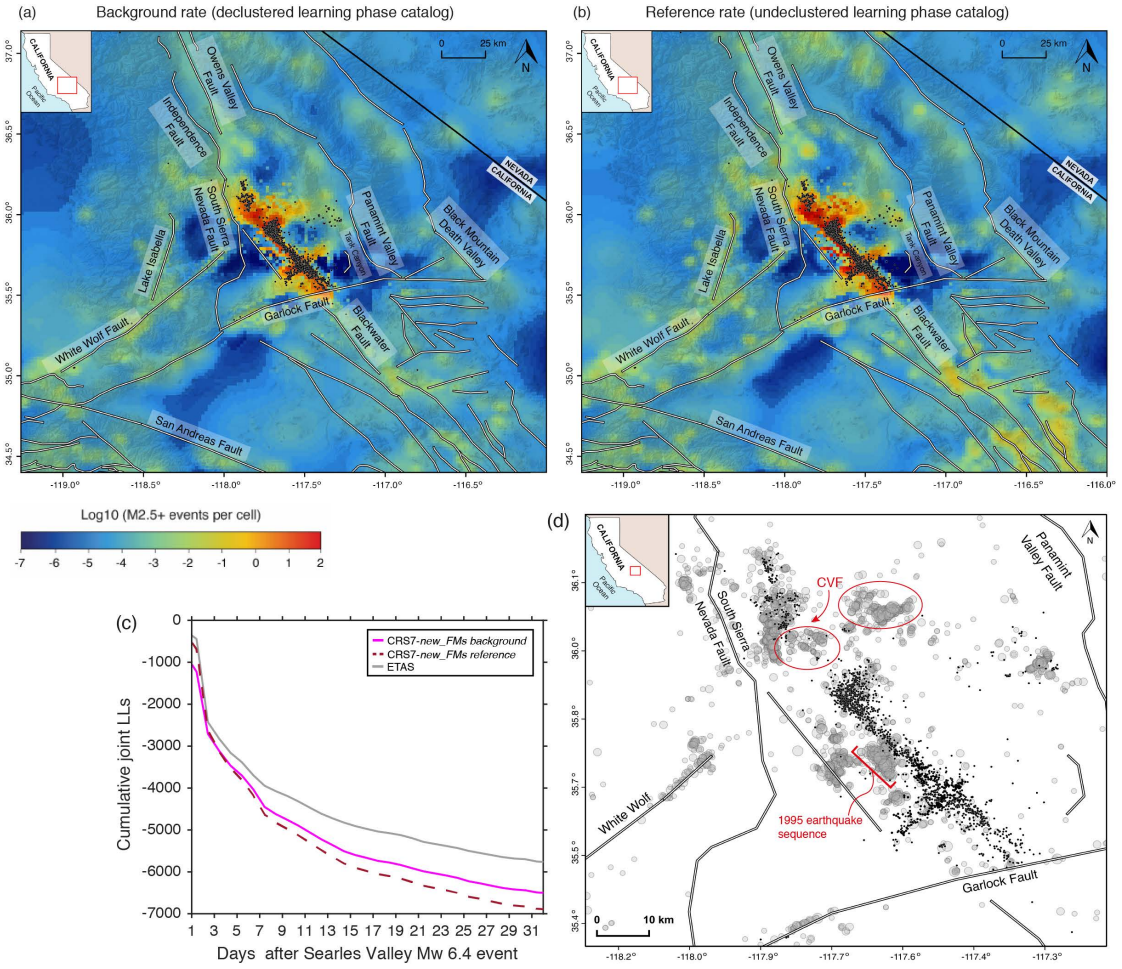

Figure B6: 1-month cumulative forecast map of model CRS7-*new_FMs* for the entire testing region when (a) r_0 is defined as a declustered background rat, or (b) r_0 is defined as a undeclustered reference rate. White lines indicate the individual UCERF3 fault traces. (c) Cumulative S-test joint log-likelihood for the two presented versions of CRS7-*new_FMs* and for the ETAS model. (d) map of M2.5+ reference seismicity for the period 1981-2019 (grey circles) and the first month of M2.5+ seismicity following the Mw 6.4 Searles Valley event; red circles highlight areas of past high clustering that were not interested by triggered seismicity above M2.5 during the first month of the 2019 sequence. CVF = Coso Volcanic Field.

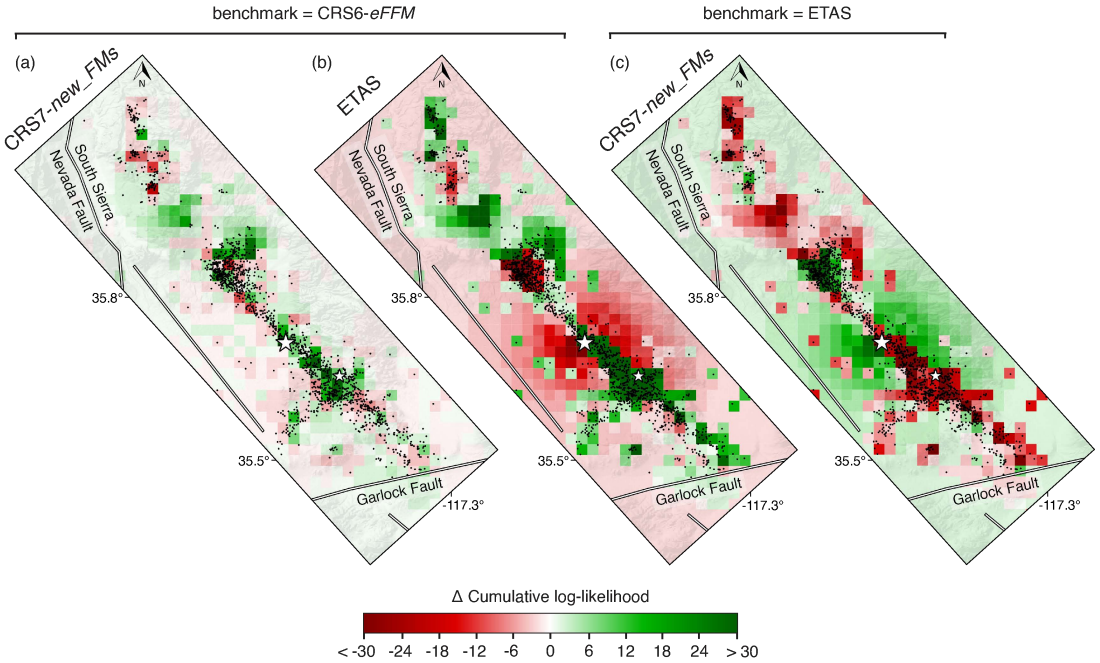

Figure B7: Maps of cumulative log-likelihood differences between pairs of models for the 1-month evaluation period. (a) Log-likelihood differences between CRS7-*new_FMs* and the benchmark CRS6-*eFFM*; (b) log-likelihood differences between ETAS and the benchmark CRS6-*eFFM*; (c) log-likelihood differences between CRS7-*new_FMs* and the benchmark ETAS. Positive (green) values indicate a better performance than the benchmark model indicated above the corresponding horizontal square bracket. Black points indicate the locations of M2.5+ aftershocks between 4 July 2019 and 4 August 2019 and white stars represent the two mainshocks. The ΔLL values are saturated at ± 30 for illustration purposes.

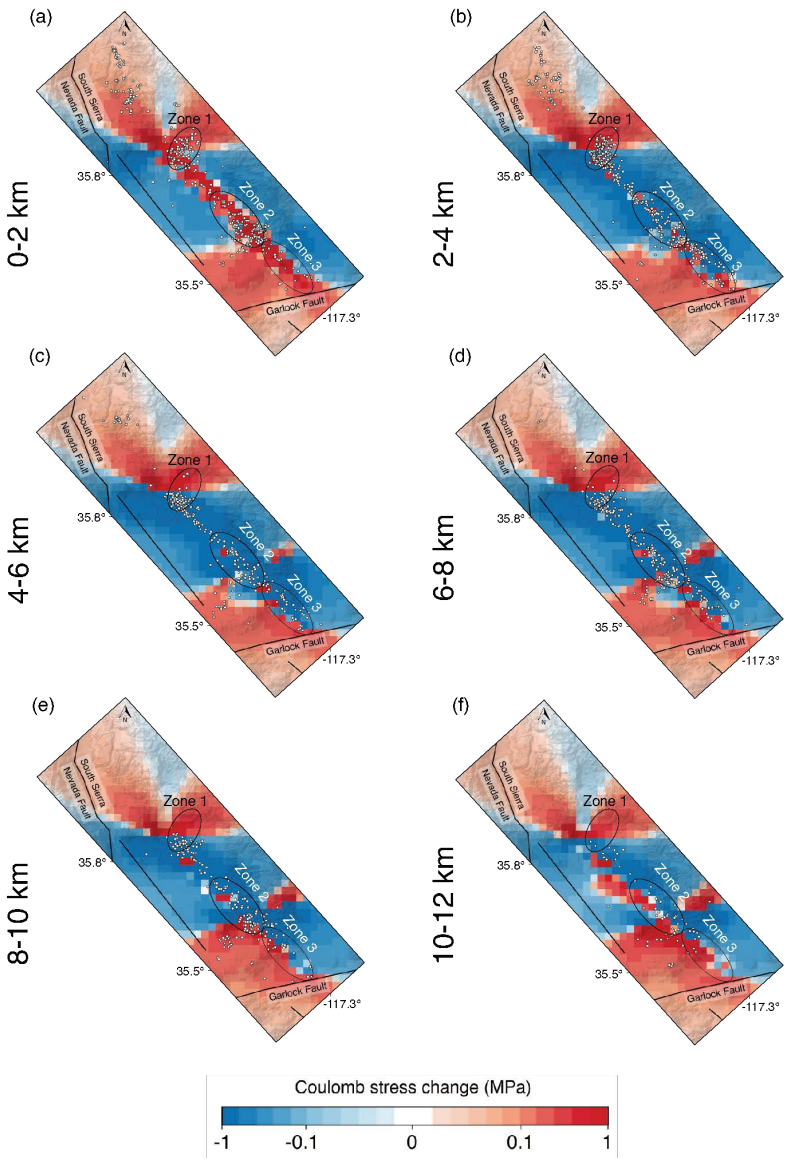

Figure B8: Maps of coseismic stress changes from the combined effect of the Mw 6.4 and Mw 7.1 mainshocks, resolved on the average pre-Ridgecrest normal fault mechanism. We show stress change values calculated at depths of: (a) 0-2 km; (b) 2-4 km; (c) 4-6 km; (d) 6-8 km; (e) 8-10 km; (f) 10-12 km; White dots indicate the M2.5+ aftershocks observed at each depth layer during the first month of the Ridgecrest sequence. Black circles define the regions of interest (Zone 1-2-3) as described in Figure 4.8a. Despite a shallow positive stress change (0-2 km), the vast majority of Zone 2 aftershocks and the z $\geq$ 4 km seismicity of Zone 1 occur within the stress shadow of the average pre-Ridgecrest normal fault. Zone 3 presents a more mixed ΔCFF pattern.

Appendix C

Supplemental Material to Chapter 5

In this appendix we first show the frequency-magnitude distributions of the catalogue used in this study to calibrate the ETAS models at PNR-1z (Figure C1a) and PNR-2 (Figure C1b). We then present examples of histograms from the ETAS simulations at PNR-1z and PNR-2 (Figures C2 and C3). Finally, we report a summary of the tested ETAS versions (Table C1), and the values of the ETAS parameters (Table C2), including the bulk and sleeve-specific values of c_f (Tables C3 and C4 for PNR-1z and PNR-2, respectively).

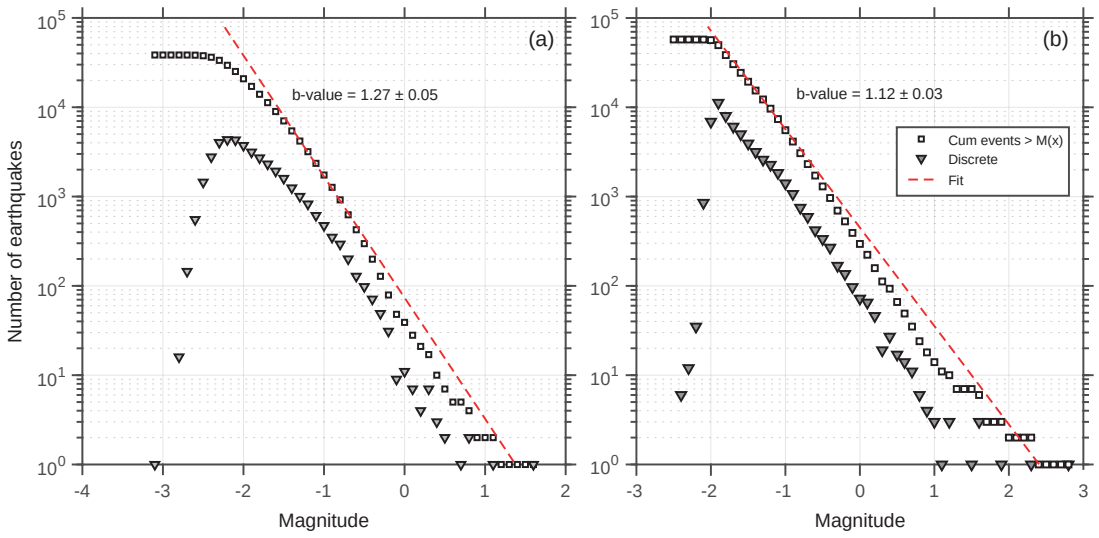

Figure C1: Frequency-magnitude distributions (FMD). (a) FMD for the PNR-1z catalogue used in this work; (b) FMD for the full PNR-2 catalogue used in this work. The red dashed line represents the Gutenberg-Richter fit to the distribution. We use the FMD to estimate the magnitude of completeness (M_c) of the catalogue using the b-value stability method (*Cao & Gao*, 2002).
The surface network detected only larger events using local magnitudes (M_L). At PNR-2, we convert these to M_w using the conversion relationship developed by QCon for Cuadrilla's hydraulic fracturing plan (*Cuadrilla Resources Inc.*, 2019). However, the same relationship does not hold for PNR-1z (*Baptie et al.*, 2020); in that case, we replace the downhole $M_w \geq 0.0$ values with the corresponding M_L from the surface catalogue.

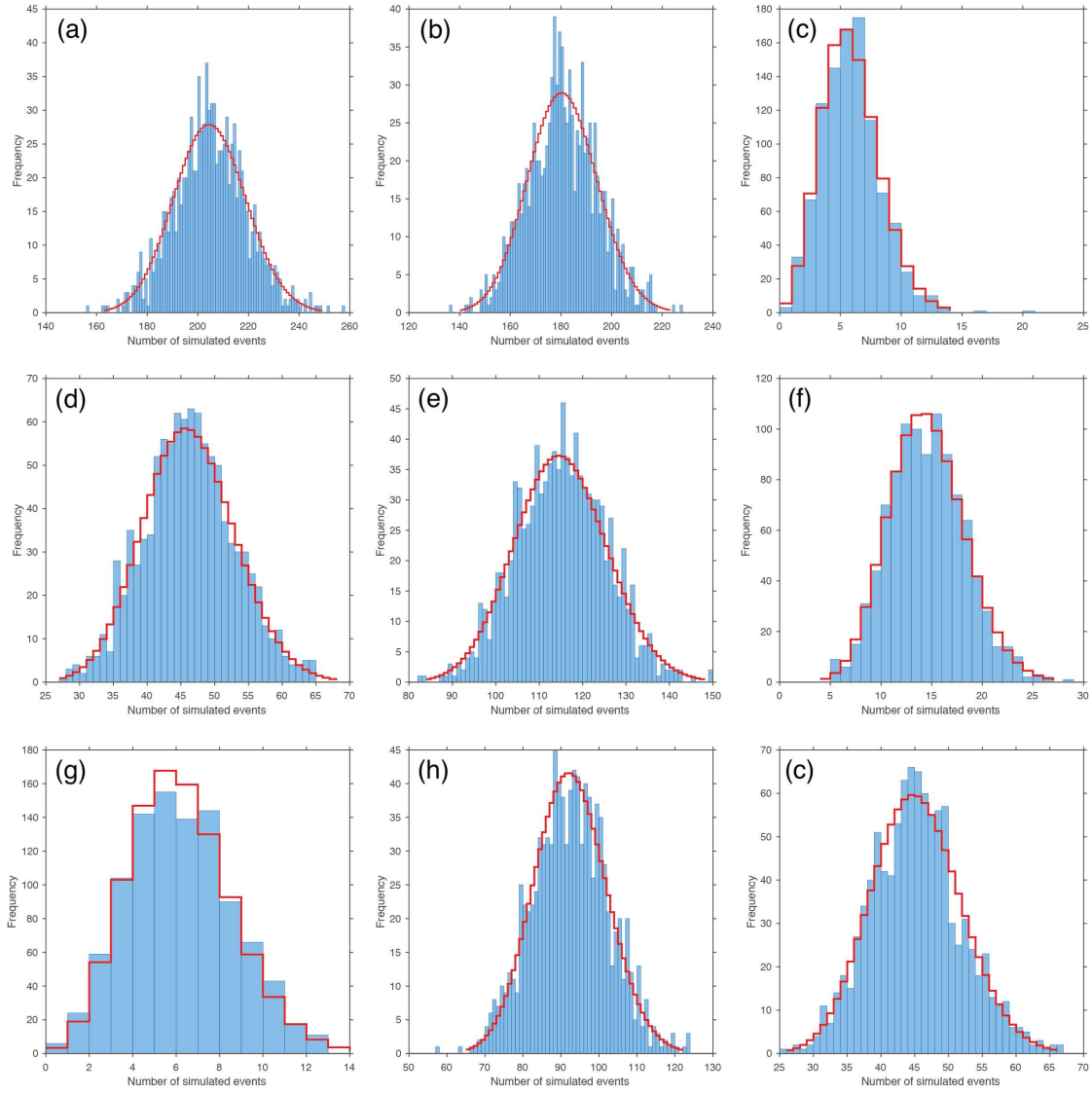

Figure C2: Simulation histograms at PNR-1z. Panels (a-i) show histograms from 9 randomly selected forecast windows, each consisting of 1000 simulations of the number of simulated events over the forecast period. The red lines represent the fits of the negative binomial distributions to the histograms.

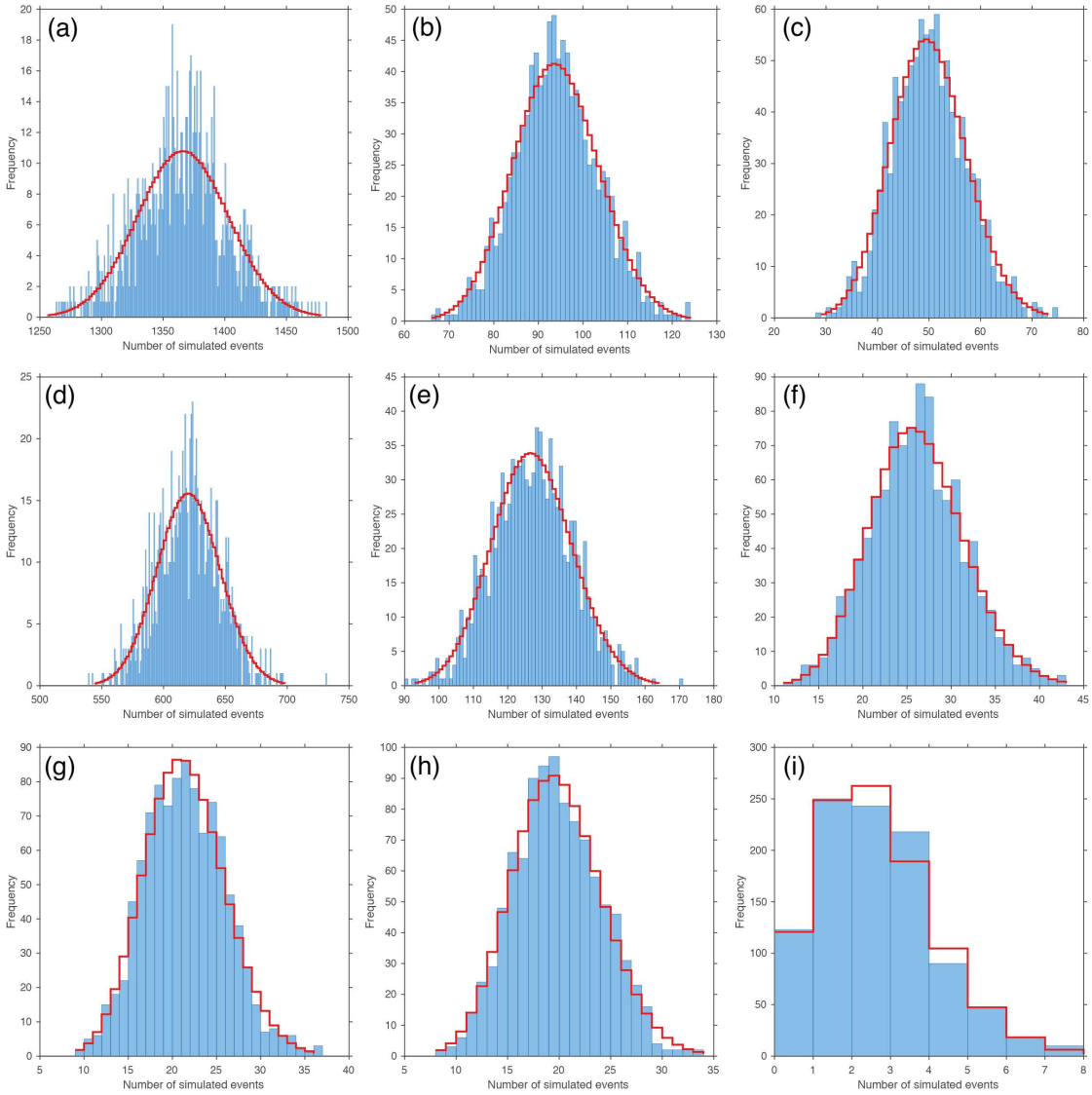

Figure C3: Simulation histograms at PNR-2. Panels (a-i) show histograms from 9 randomly selected forecast windows, each consisting of 1000 simulations of the number of simulated events over the forecast period. The red lines represent the fits of the negative binomial distributions to the empirical histograms.

Table C1: Summary of the developed ETAS models for PNR-1z and PNR-2.

Model	**ETAS parameters**	**ETAS type**	**C_F**	**Update Frequency**	**Target magnitudes**	**Target catalogue**
ETAS1-global	Global ETAS	Standard	No	1 hour	M ≥ -1.5	PNR-1z and 2
ETAS1-unoptimised	Optimised on PNR-1z	Standard	No	1 hour	M ≥ -1.5	PNR-2 only
ETAS1-optimised	Optimised on specific catalog	Standard	No	1 hour	M ≥ -1.5	PNR-1z and 2
ETAS2-unoptimised	Optimised on PNR-1z	Modified	PNR-1z bulk	1 hour	M ≥ -1.5	PNR-2 only
ETAS2-bulk	Optimised on specific catalog	Modified	Bulk	1 hour	M ≥ -1.5	PNR-1z and 2
ETAS2-specific	Optimised on specific catalog	Modified	Sleeve-specific	1 hour	M ≥ -1.5	PNR-1z and 2

Table C2: ETAS parameters. When estimating the ETAS parameters, we constrain the branching ratio (*i.e.* the fraction of triggered events) to be less than 1.

Parameter	ETAS1-optimised		ETAS2-bulk		Global parameters
	PNR-1z	PNR-2	PNR-1z	PNR-2	
μ (eq/day)	0.26	0.31	N/A	N/A	0.26
c_f	N/A	N/A	0.29	1.70	N/A
c (days)	0.58	0.47	0.51	0.5	0.03
p	1.51	1.49	1.52	1.50	1.21
K	0.77	0.79	0.85	0.65	0.04
α	0.66	0.59	0.45	0.56	2.3

Table C3: Constants of proportionality (c_f) between injection rate and seismicity rate at PNR-1z sleeves.

	Sleeve number													
	#1	#2	#3	#12	#13	#14	#18 mini	#22a	#22b	#22c	#30a	#30b	#31	
c_f	1.35	0.24	0.42	0.17	0.13	0.04	0.25	0.54	0.75	0.001	1.96	2.48	0.22	
	Sleeve number													**Bulk**
	#32	#39 mini	#41 mini	#37 mini	#40 mini	#38 mini	#35 mini	#37a	#37b	#38	#39	#40	#41	
c_f	0.28	0.08	1.42	0.001	0.43	0.52	0.21	0.66	0.39	1.28	1.23	1.40	0.30	0.29

Table C4: Constants of proportionality (c_f) between injection rate and seismicity rate at PNR-2 sleeves.

	Sleeve number								**Bulk**
	#1a	#1b	#2	#3	#4	#5	#6a	#6b	#7	
c_f	0.71	4.38	3.39	2.93	1.66	0.73	0.65	1.42	0.70	1.70

Bibliography

[1] Amato, A., Azzara, R., Chiarabba, C., Cimini, G. B., Cocco, M., Di Bona, M., Margheriti, L., Mazza, S., Mele, F., Selvaggi, G., Basili, A., & Boschi, E. (1998). The 1997 Umbria-Marche, Italy, earthquake sequence: A first look at the main shocks and aftershocks. *Geophys. Res. Lett.*, 25(15), 2861–2864. https://doi.org/10.1029/98GL51842.

[2] Asanuma, H., Eto, T., Adachi, M., Saeki, K., Aoyama, K., Ozeki, H., & Haring, M. (2014). Seismostatistical Characterization of Earthquakes from Geothermal Reservoirs. *Proceedings of the Thirty-Ninth Workshop on Geothermal Reservoir engineering Stanford University, Stanford, California*, February 2014, SGP-TR-202.

[3] Atkinson, G. M., Eton, D. W., & Igonin, N. (2020). Developments in understanding seismicity triggered by hydraulic fracturing. *Nat. Rev. Earth Environ.*, 1, 264-277. https://doi.org/10.1038/s43017-020-0049-7.

[4] Bachmann, C., Wiemer, S., Woessner, J., & Hainzl, S. (2011). Statistical analysis of the induced Basel 2006 earthquake sequence: Introducing a probability-based monitoring approach for Enhanced Geothermal Systems. *Geophys. J. Int.*, 186, 793-807.

[5] Baptie, B. & Luckett, R. (2019). Seismicity Induced by Hydraulic Fracturing Operations at Preston New Road, Lancashire, 2018. *Proceedings of the Society of Earthquake and Civil Engineering Dynamics Conference*, September 2019, Greenwich, London.

[6] Baptie, B., Luckett, R., Butcher, A., & Werner, M. J. (2020). Robust relationships for magnitude conversion of PNR seismicity catalogues. *British Geological Survey Open Report.*

[7] Barnhart, W. D., Hayes, G. P., & Gold, R. D. (2019). The July 2019 Ridgecrest, California, earthquake sequence: Kinematics of slip and stressing in cross-fault ruptures. *Geophys. Res. Lett.*, 46. https://doi.org/10.1029/2019GL084741.

[8] Basili R., Valensise, G., Vannoli, P., Burrato, P., Fracassi, U., Mariano, S., Tiberti, M. M., & Boschi, E. (2008). The Database of Individual Seismogenic Sources (DISS), version 3: summarizing 20 years of research on Italy's earthquake geology. *Tectonophysics*, 453, 20-43. https://doi.org/10.1016/j.tecto.2007.04.014.

[9] Ben-Zion, Y. (2019). A Critical Data Gap in Earthquake Physics. *Seismol. Res. Lett*, 90(5), 1721-1722. https://doi.org/10.1785/0220190167.

[10] Bommer, J. J., Oates, S., Cepeda, J. M., Lindholm, C., Bird, J., Torres, R., Marroquin, G., & Rivas, J. (2006). Control of hazard due to seismicity induced by a hot fractured rock geothermal project. *Eng. Geol.* 83, 287-306. https://doi.org/10.1016/j.enggeo.2005.11.002.

[11] Boschi, E., Guidoboni, E., Ferrari, G., Mariotti, D., Valensise, G., & Gasperini P. (eds.) (2000). Catalogue of Strong Italian Earthquakes from 461 B.C. to 1997. *Ann. Geophys.*, 43(4), 609-868. https://doi.org/10.4401/ag-3668.

[12] Bourne, S., & Oates, S. (2017). Development of statistical geomechanical models for forecasting seismicity induced by gas production from the Groningen field. *Netherlands Journal of geosciences*, 96(5), S175-S182. http://doi.org.10.1017/njg.2017.35.

[13] Brodsky, E., & van der Elst, N. J. (2014). The Uses of Dynamic Earthquake Triggering. *Annu. Rev. Earth Planet. Sci.*, 42, 317-339. https://doi.org/10.1146/annurev-earth-060313-054648.

[14] Bryant, W. A., compiler, 2017, Fault number 72, Little Lake fault zone, in Quaternary fault and fold database of the United States: U.S. Geological Survey website, https://earthquakes.usgs.gov/hazards/qfaults.

[15] Buttinelli, M., Pezzo, G., Valoroso, L., De Gori, P., & Chiarabba, C. (2018). Tectonics inversions, fault segmentation, and triggering mechanisms in the central Apennines normal fault system: Insights from high-resolution velocity models. *Tectonics*, 37. https://doi.org/10.1029/2018TC005053.

[16] Cao, A. M., & Gao, S. S. (2002). Temporal variation of seismic b-values beneath northeastern Japan island arc. *Geoph. Res. Lett.*, 29(9). https://doi.org/10.1029/2001GL013775.

[17] Catalli, F., Cocco, M., Console, R., & Chiaraluce, L. (2008). Modeling seismicity changes during the 1997 Umbria-Marche sequence (central Italy) through a rate- and state-dependent model. *J. Geophys. Res.*, 113, B11301. https://doi.org/10.1029/2007JB005356.

[18] Cattania, C., Hainzl, S., Wang, L., Roth, F., & Enescu, B. (2014). Propagation of Coulomb stress uncertainties in physics-based aftershock models. *J. Geophys. Res. Solid Earth*, 119, 7846-7864. https://doi.org/10.1002/2014JB011183.

[19] Cattania, C., & Khalid, F. (2016). A parallel code to calculate rate-state seismicity evolution induced by time dependent, heterogeneous Coulomb stress changes. *Computers & Geosciences*, 94, 48-55. http://dx.doi.org/10.1016/j.cageo.2016.06.007.

[20] Cattania, C., Werner, M. J., Marzocchi, W., Hainzl, S., Rhoades, D., Gerstenberger, M., Liukis, M., Savran, W., Christophersen, A., Helmstetter, H., Jimenez, A., Steacy, S., & Jordan, T. H. (2018). The Forecasting Skill of physics-based models during the 2010-2012 Canterbury, New Zealand, Earthquake Sequence. *Seismol. Res. Lett*, 89(4), 1238-1250. https://doi.org/10.1785/0220180033.

[21] Chen, K. H., Burgmann, R., & Nadeau, R. M. M. (2013). Do earthquakes talk to each other? Triggering and interaction of repeating sequences at Parkfield. *J. Geophys. Res. Solid Earth*, 118, 165-182. https://doi.org/10.1029/2012JB009486.

[22] Chen, K., Avouac, J., Aati, S., Milliner, C., Zheng, F., & Shi, C. (2020). Cascading and pulse-like ruptures during the 2019 Ridgecrest earthquakes in the Eastern California Shear Zone. *Nat. Comm.* 11, 22. https://doi.org/10.1038/s41467-019-13750-w.

[23] Chiarabba, C., Jovane, L., & Di Stefano, R. (2005). A new view of Italian seismicity using 20 years of instrumental recordings. *Tectonophysics*, 395, 251-268. https://doi.org/10.1016/j.tecto.2004.09.013.

[24] Chiarabba, C., Amato, A., Anselmi, M., Baccheschi, P., Bianchi, I., Cattaneo, M., Cecere, G., Chiaraluce, L., Ciaccio, M. G., De Gori, P., et al. (2009). The 2009 L'Aquila (central Italy) M_w 6.3 earthquake: Main shock and aftershocks. *Geophys. Res. Lett.*, 36, L18308. https://doi.org/10.1029/2009GL039627.

[25] Chiarabba, C., & De Gori, P. (2016). The seismogenic thickness in Italy: constraints on potential magnitude and seismic hazard. *Terra Nova*, 28, 402-408. https://doi.org/10.1111/ter.12233.

[26] Chiarabba, C., De Gori, P., Cattaneo, M., Spallarossa, D., & Segou, M. (2018). Faults geometry and the role of fluids in the 2016-2017 Central Italy seismic sequence. *Geoph. Res. Lett*, 45, 6963-6971. https://doi.org/10.1029/2018GL077485.

[27] Chiaraluce, L., Ellsworth, W. L., Chiarabba, C., & Cocco, M. (2003). Imaging the complexity of an active normal fault system: The 1997 Colfiorito (central Italy) case study. *J. Geophys. Res.* 108, no. B6. https://doi.org/10.1029/2002JB002166.

[28] Chiaraluce, L., Amato, A., Cocco, M., Chiarabba, C., Selvaggi, G., Di Bona, M., Piccinini, D., Deschamps, A., Margheriti, L., Courboulex, F., et al. (2004). Complex normal faulting in the Apennines thrust-and-fold belt: The 1997 seismic sequence in Central Italy. *Bull. Seismol. Soc. Am.* 94(1),99-116.

[29] Chiaraluce, L., Valoroso, L., Piccinini, D., Di Stefano, R., & De Gori, P. (2011). The anatomy of the L'Aquila normal fault system (Central Italy) imaged by high resolution foreshock and aftershock locations. *J. Geophys. Res.*, 116, no. B12. https://doi.org/10.1029/2011JB008352.

[30] Chiaraluce, L., Di Stefano, R., Tinti, E., Scognamiglio, L., Michele, M., Casarotti, E., Cattaneo, M., De Gori, P., Chiarabba, C., Monachesi, G., et al. (2017). The 2016 Central Italy seismic sequence: A first look at the mainshocks, aftershocks, and source models. *Seismol. Res. Lett.*, 88(3). https://doi.org/10.1785/0220160221.

[31] Clarke, H., Verdon, J. P., Kettlety, T., Baird, A. F., & Kendall, J. M. (2019). Real-Time Imaging, Forecasting, and Management of Human-Induced Seismicity at Preston New Road, Lancashire, England. *Seismol. Res. Lett.*, 90(5), 1902-1915.

[32] Clemens, R. A., Schoenberg, F. P., & Schorlemmer, D. (2011). Residual analysis methods for space-time point processes with applications to earthquake forecast models in California. *Ann. Appl. Stat.*, 5(4), 2549-2571. https://doi.org/10.1214/11-AOAS487.

[33] Cocco, M., & Rice, J. R. (2002). Pore pressure and poroelasticity effects in Coulomb stress analysis of earthquake interactions. *J. Geophys. Res.*, 107, no., B2. https://doi.org/10.1029/2000JB000138.

[34] Cocco, M., Hainzl, S., Catalli, F., Enescu, B., Lombardi, A. M., & Woessner, J. (2010). Sensitivity study of forecasted aftershock seismicity based on Coulomb stress calculation and rate-and state-depedent frictional response. *J. Geophys. Res.*, 115, B05307. https://doi.org/10.1029/2009JB006838.

[35] Console, R., Murru, M., & Catalli, F. (2006). Physical and stochastic models of earthquake clustering. *Tectonophysics*, 417, 141-153. https://doi.org/10.1016/j.tecto.2005.05.052.

[36] Cremen, G. & Werner, M. J. (2020). A Novel Approach to Assessing Nuisance Risk from Seismicity Induced by UK Shale Gas Development, with Implications for Future Policy Design. *Nat. Hazards Earth Syst. Sci. Discuss.*. https://doi.org/10.5194/nhess-2020-95.

[37] Cuadrilla Resources Inc. (2019). Hydraulic Fracture Plan PNR 2. *Cuadrilla Resources Inc. Report CORP-HSE-RPT-003.*

[38] Das, S., & Scholz, C. H. (1981). Off-fault aftershock clusters caused by shear stress increase? *Bull. Seism. Soc. Am.*, 71(5), 1669-1675.

[39] Dawson, T. E. (2013). Appendix A - Updates to the California Reference Fault Parameter Database - Uniform California Earthquake Rupture Forecast, Version 3 Fault Models 3.1 and 3.2, U.S. Geological Survey Open-File Report, 2013-1165, 18 p., http://pubs.usgs.gov/of/2013/1165/pdf/ofr2013-1165_appendixC.pdf.

[40] Diao, F., Xiong, X., Wang, R., Zheng, Y., Walter, T.R., Weng, H., & Li, J. (2014). Overlapping post-seismic deformation processes: Afterslip and viscoelastic relaxation following the 2011 Mw 9.0 Tohoku (Japan) earthquake. *Geophys. J. Int.* 196(1), 218-229. https://doi.org/10.1093/gji/ggt376.

[41] Dieterich, J. H. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.*, 99(B2), 2601-2618.

[42] Dieterich, J. H., Cayol, V., & Okubo, P. (2000). The use of earthquake rate changes as stress meter at Kilauea volcano. *Nature.*, 408, 457-460.

[43] DISS Working Group (2018). Database of Individual Seismogenic Sources (DISS), Version 3.2.1: A compilation of potential sources for earthquakes larger than M_w 5.5 in Italy and surrounding areas. `http://diss.rm.ingv.it/diss/`.

[44] Ellsworth, W. L. (2013). Injection-induced earthquakes. *Science*, 341(6142), 1225942. `https://doi.org/10.1126/science.1225942`.

[45] Ellsworth, W. L., Giardini, D., Townend, J., Ge, S., & Shimamoto, T. (2019). Triggering of the Pohang, Korea, Earthquake (M_w 5.5) by Enhanced Geothermal System Stimulation. *Seismol. Res. Lett.*, 90(5), 1844-1858.

[46] EMERGEO Working Group (2016). The 24 August 2016 Amatrice Earthquake: Coseismic Effects. `https://doi.org/10.5281/zenodo.61568`.

[47] Felzer, K. R. (2002). Triggering of the 1999 Mw 7.1 Hector Mine earthquake by aftershocks of the 1992 Mw 7.3 Landers earthquake. *J. Geophys. Res.*, 107(B9), 2190. `https://doi.org/10.1029/2001JB000911`.

[48] Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., et al. (2014). Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) - the time-independent model. *Bull. Seismol. Soc. Am.*, 104, 1122-1180. `https://doi.org/10.1785/0120130164`.

[49] Field, E. H., Milner, K. R., Hardebeck, J. L. , Page, M. T., van der Elst, N., Jordan, T. H., Michael, A. J., Shaw, B. E., & Werner, M. J. (2017). A spatiotemporal clustering model for the third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): toward an operational earthquake forecast, *Bull. Seismol. Soc. Am.* 107(3), 1049-1081. `https://doi.org/10.1785/0120160173`.

[50] Freed, A.M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annu. Rev. Earth Planet. Sci.* 33, 335-367. `https://doi.org/10.1146/annurev.earth.33.092203.122505`.

[51] Frohlich, C., & Apperson, K. (1992). Earthquake focal mechanisms, moment tensors, and the consistency of seismic activity near plate boundaries. *Tectonics* 11(2), 279-296.

[52] Gerstenberger, M. C., Wiemer, S., Jones, L. M., & Reasenberg, P. A. (2005). Real-time forecasts of tomorrow's earthquakes in California. *Nature*, 435, 328-331. https://doi.org/0.1038/nature03622.

[53] Gerstenberger, M. C., & Rhoades, D. A. (2010). New Zealand earthquake forecast testing centre. *Pure Appl. Geophys*, 167, 8-9. 877-892. https://doi.org/10.1007/s00024-010-0082-4.

[54] Gomberg, J., Beeler, N. M., Blanpied, M. L., & Bodin, P. (1998). Earthquake triggering by transient and static deformations. *J. Geophys. Res.* 103, B10, 411-426. https://doi.org/10.1029/98JB01125.

[55] Gomberg, J., Reasenberg, P., Bodin, P., & Harris, R. A. (2001). Earthquake triggering by seismic waves following the Landers and Hector Mine earthquakes. *Nature* 411, 462-466. https://doi.org/10.1038/35078053.

[56] Guy, M., Patton, J., Fee, J., Hearne, M., Martinez, E., Ketchum, D., Worden, C., Quitoriano, V., Hunter, E., Smoczyk, G. et al. (2015). National Earthquake Information Center systems overview and integration (USGS numbered series No. 2015-1120), Open-File Report. U.S. Geological Survey, Reston, VA.

[57] Hainzl, S., & Ogata, Y. (2005). Detecting fluid signals in seismicity data through statistical earthquake modeling. *J. Geophys. Res. Solid Earth*, 110(B5).

[58] Hainzl, S., Christophersen, A., & Enescu, B. (2008). Impact of earthquake rupture extension on parameters estimations of point-process models. *Bull. Seismol. Soc. Am.* 98(4), 2066-2072. https://doi.org/10.1785/0120070256.

[59] Hainzl, S., Enescu, B., Cocco, M., Woessner, J., Catalli, F., Wang, R., & Roth, F. (2009). Aftershock modeling based on uncertain stress calculations. *J. Geophys. Res.*, 114, B05309. https://doi.org/10.1029/2008JB006011.

[60] Hainzl, S., Zoller, G., & Wang, R. (2010). Impact of the receiver fault distribution on aftershock activity. *J. Geophys. Res.*, 115, B05315. https://doi.org/10.1029/2008JB006224.

[61] Hainzl, S., Zakharova, O., & Marsan, D. (2013). Impact of aseismic transients on the estimation of aftershock productivity parameters. *Bull. Seismol. Soc. Am.* 103, 1723-1732.

[62] Hanks, T., & Kanamori, H. (1979). A moment magnitude scale. *J. Geophys. Res.*, 84, 2348-2350. https://doi.org/10.1029/JB084iB05p02348.

[63] Hara, S., Fukahata, Y. & Iio, Y. (2019). P-wave first-motion polarity determination of waveform data in western Japan using deep learning. *Earth Planets Space* 71, 127. https://doi.org/10.1186/s40623-019-1111-x.

[64] Hardebeck, J. L., Nazareth, J. J., & Hauksson, E. (1998). The static stress change triggering model: Constraints from two southern California earthquake sequences. *J. Geophys. Res.*, 103, 24427-24437.

[65] Harris, R. A., & Simpson, R. W. (1992). Changes in static stress on southern California faults after the 1992 Landers earthquake. *Nature*, 360(6401), 251-254. https://doi.org/10.1038/360251a0.

[66] Harte, D. (2015). Log-likelihood of earthquake models: evaluation of models and forecasts. *Geophys. J. Int.*, 201, 711-723, https://doi.org/10.1093/gji/ggu442.

[67] Hauksson, E., Hutton, K., Kanamori, H., Jones, L., Mori, J., Hough, S. E., & Roquemore, G. (1995). Preliminary report on the 1995 Ridgecrest earthquake sequence in Eastern California. *Seismol. Res. Lett.*, 66(6), 54-60.

[68] Hauksson, E., Yang, W., & Shearer, P. M. (2012). Waveform relocated earthquake catalog for southern California (1981 to 2011). *Bull. Seismol. Soc. Am.* 102(5), 2,239-2,244.

[69] Heimisson, E. R., & Segall, P. (2018). Constitutive law for earthquake production based on rate-and- state friction: Dieterich 1994 revisited. *J. Geophys. Res. Solid Earth* 123, 4141-4156. https://doi.org/10.1029/2018JB015656.

[70] Heimisson, E. R. (2019). Constitutive law for earthquake production based on rate-and-state friction: Theory and Application of Interacting Sources. *J. Geophys. Res. Solid Earth* 124, 1802- 1821. https://doi.org/10.1029/2018JB016823.

[71] Helmstetter, A., & Sornette, D. (2003). Predictability in the Epidemic-Type Aftershock Sequence model of interacting triggered seismicity. *J. Geophys. Res.*, 108(B10), 2482. https://doi.org/10.1029/2003JB002485.

[72] Helmstetter, A., Kagan, Y., & Jackson, D. D. (2005). Importance of small earthquakes for stress transfers and earthquake triggering. *J. Geophys. Res.*, 110, B05S08. https://doi:10.1029/2004JB00328.

[73] Helmstetter, A., Kagan, Y., & Jackson, D. D. (2006). Comparison of short-term and time-independent earthquake forecast models for Southern California. *Bull. Seismol. Soc. Am.*, 96(1), 90-106. https://doi.org/10.1785/0120050067.

[74] Helmstetter, A., Kagan, Y., & Jackson, D. D. (2007). High-resolution Time-independent Grid-based Forecast for M $\geq$ 5 Earthquakes in California. *Seismol. Res. Lett.*, 78(1), 78-86. https://doi.org/10.1785/gssrl.78.1.78.

[75] Ishibe, T., Shimazaki, K., Kenji, S., & Tsuruoka, H. (2011). Change in seismicity beneath the Tokyo metropolitan area due to the 2011 off the Pacific coast of Tohoku earthquake. *Earth Planets Space* 63, 40. https://doi.org/10.5047/eps.2011.06.001.

[76] Jackson, D. D. (2018). Testing the Classic 1988 Forecast. *Seismol. Res. Lett*, 89(4), 1288-1297. https://doi.org/10.1785/0220180039.

[77] Jia, K., Zhou, S., Zhuang, J., Jiang, C., Guo, Y., Gao, Z., Gao, S., Ogata, Y., & Song, X. (2020). Nonstationary Background Seismicity Rate and Evolution of Stress Changes in the Changning Salt Mining and Shale-Gas Hydraulic Fracturing Region, Sichuan Basin, China. *Seismol. Res. Lett.* 91, 2170-2181. https://doi.org/10.1785/0220200092.

[78] Jordan, T. H. (2006). Earthquake Predictability, Brick by Brick. *Seismol. Res. Lett.* 77(1), 3-6. https://doi.org/10.1785/gssrl.77.1.3.

[79] Jordan, T. H., Chen, Y., & Main, I. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Ann. Geophys.* 54(4). https://doi.org/10.4401/ag-5350.

[80] Jordan, T. H., Marzocchi, W., Michael, A. J., & Gerstenberger, M. C. (2014). Operational Earthquake Forecasting Can Enhance Earthquake Preparedness. *Seismol. Res. Lett.*, 85(5), 955-959. https://doi.org/10.1785/0220140143.

[81] Kagan, Y. Y. (2003). Accuracy of modern global earthquake catalogs. *Phys. Earth Planet. In.* 135, 173-209.

[82] Kagan, Y.Y. (2010). Statistical distribution of earthquake numbers: consequence of branching process. *Geophys. J. Int.* 180, 1313-1328. https://doi.org/10.1111/j.1365-246X.2009.04487.x.

[83] Kao, H., Visser, R., Smith, B., & Venables, S. (2018). Performance assessment of the induced seismicity traffic light protocol for northeastern British Columbia and western Alberta. *Lead. Edge* 37, 117-126.

[84] Keranen, K. M., Savage, H. M., Abers, G. A., & Cochran, E. S. (2013). Potentially induced earthquakes in Oklahoma, USA: Links between wastewater injection and the 2011 M_w 5.7 earthquake sequence. *Geology*, 41(6), 699-702. https://doi.org/10.1130/G34045.1.

[85] Kettlety, T., Verdon, J. P., Butcher, A., Hampson, M., & Craddock, L. (2020). High-resolution imaging of the ML 2.9 August 2019 earthquake in Lancashire, United Kingdom, induced by hydraulic fracturing during Preston New Road PNR-2 operations. *Seismol. Res. Lett.*, 92(1), 151-169. https://doi.org/10.1785/0220200187.

[86] King, G. C. P., Stein, R. S., & Lin, J. (1994). Static stress changes and the triggering of earthquakes. *Bull. Seismol. Soc. Am.*, 84(3), 935-953.

[87] Kiraly-Proag, E., Zechar, J. D., Gischig, V., Wiemer, S., Karvounis, D., & Doetsch, J. (2016). Validating induced seismicity forecast models-Induced Seismicity Test Bench. *J. Geophys. Res. Solid Earth* 121, 6009-6029. https://doi.org/10.1002/2016JB013236.

[88] Lee, K., Ellsworth, W. L., Giardini, D., Townend, J., Ge, S., Shimamoto, T., Yeo, I., Kang, T., Rhie, J., Sheen, D., Chang, C., Woo, J., & Langenbruch, C. (2019). Managing injection-induced seismic risks. *Science*, 364 (6442), 730-32.

[89] Lei, X., Huang, D., Su, J., Jiang, G., Wang, X., Wang, H., Guo, X., & Fu, H. (2017). Fault reactivation and earthquakes with magnitudes of up to M_w 4.7 induced by shake-gas hydraulic fracturing in Sichuan Basin, China. *Sci. Rep.*, 7 (7971). https://doi.org/10.1038/s41598-017-08557-y.

[90] Lei, X., Wang, Z., & Su, J. (2019). The December 2018 ML 5.7 and January 2019 ML 5.3 earthquakes in South Sichuan Basin induced by shale gas hydraulic fracturing. *Seismol. Res. Lett.* 90(3), 1099-1110. https://doi.org/10.1785/0220190029.

[91] Linker, M. F., & Dieterich, J. H. (1992). Effects of variable normal stress on rock friction: Observations and constitutive equations. *J. Geophys. Res.*, 97(92), 4923-4940. https://doi.org/10.1029/92JB00017.

[92] Lippiello, E., Giacco, F., de Arcangelis, L., Marzocchi, W. & Godano, C. (2014). Parameter Estimation in the ETAS Model: Approximations and Novel Methods. *Bull. Seism. Soc. Am.* 104, 985-994. https://doi.org/10.1785/0120130148.

[93] Llenos, A. L., & Michael, A. J. (2013). Modeling earthquake rate change in Oklahoma and Arkansas: possible signatures of induced seismicity. *Bull. Seismol. Soc. Am.*, 103(5), 2850-2861. https://doi.org/10.1785/0120130017.

[94] Lombardi, A. M., Cocco, M., & Marzocchi, W. (2010). On the Increase of Background Seismicity Rate during the 1997-1998 Umbria-Marche, Central Italy, Sequence: Apparent Variation or Fluid-Driven Triggering? *Bull. Seismol. Soc. Am.* 100(3), 1138-1152. https://doi.org/10.1785/0120090077.

[95] Lombardi, A. M. (2015). Estimation of the parameters of ETAS models by Simulated Annealing. *Sci. Rep.* 5, 8417. https://doi.org/10.1038/srep08417.

[96] Luttrell, K., & Smith-Konter, B. (2017). Limits on crustal differential stress in southern California from topography and earthquake focal mechanisms. *Geophys. J. Int.* 211, 472-482. https://doi.org/10.1093/gji/ggx301.

[97] Maeda, K. (2006). Estimation of the fault constitutive parameter $A\sigma$ and stress accumulation rate from seismicity response to a large earthquake. *In Proceedings of the 4th International Workshop on Statistical Seismology*, pp. 156-159.

[98] Malagnini, L., Lucente, F. P., De Gori, P., Akinci, A., & Munafo, I. (2012). Control of pore fluid pressure diffusion on fault failure mode: Insights from the 2009 L'Aquila seismic sequence. *J. Geophys. Res.*, 117, B05302. https://doi.org/10.1029/2011JB008911.

[99] Mallman, E. P., & Parsons, T. (2008). A global search for stress shadows. *J. Geophys. Res.*, 113, 1-16. https://doi.org/10.1029/2007JB005336.

[100] **Mancini, S.**, Segou, M., Werner, M. J., & Cattania, C. (2019). Improving physics-based aftershock forecasts during the 2016–2017 Central Italy Earthquake Cascade. *J. of Geophys. Res. Solid Earth*, 124, 8626-8643. https://doi.org/10.1029/2019JB017874.

[101] **Mancini, S.**, Segou, M., Werner, M. J., & Parsons, T. (2020). The Predictive Skills of Elastic Coulomb Rate-and-State Aftershock Forecasts during the 2019 Ridgecrest, California, Earthquake Sequence. *Bull. Seismol. Soc. Am.*, 110(4), 1736-1751. https://doi.org/10.1785/0120200028.

[102] Marsan, D. (2005). The role of small earthquakes in redistributing crustal elastic stress. *Geophys. J. Int.*, 163, 141-151. https://doi.org/10.1111/j.1365-246X.2005.02700.x.

[103] Martinez-Garzon, P., Zaliapin, I., Ben-Zion, Y., Kwiatek, G., & Bohnhoff, M. (2018). Comparative study of earthquake clustering in relation to hydraulic activities at geothermal fields in California. *J. Geophys. Res.* 123, 4041-4062.

[104] Marzocchi, W., & Lombardi, A. M. (2009). Real-time forecasting following a damaging earthquake. *Geophys. Res. Lett.*, 36, L21302. https://doi.org/10.1029/2009GL040233.

[105] Marzocchi, W., Murru, M., Lombardi, A. M., Falcone, G., & Console, R. (2012). Daily earthquake forecasts during the May-June 2012 Emilia earthquake sequence (northern Italy). *Ann. Geophys.*, 55(4), 561-567. https://doi.org/10.4401/ag-6161.

[106] Marzocchi, W., Zechar, J. D., & Jordan, T. H. (2012). Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting. *Bull. Seismol. Soc. Am.*, 102(6), 2574-2584. https://doi.org/10.1785/0120110327.

[107] Marzocchi W., Lombardi, A. M., & Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in Italy. *Seismol. Res. Lett.* 85(5), 961-969.

[108] Marzocchi, W., Taroni, M., & Falcone, G. (2017). Earthquake forecasting during the complex Amatrice-Norcia seismic sequence. *Science Advances* 3, e1701239.

[109] McCloskey, J., Nalbant, S. S., Steacy, S., Nostro, C., Scotti, O., & Baumont, D. (2003). Structural constraints on the spatial distribution of aftershocks. *Geophys. Res. Lett.* 30(12), 1610. https://doi.org/10.1029/2003GL017225.

[110] Mena, B., Wiemer, S., & Bachmann, C. (2013). Building robust models to forecast the induced seismicity related to geothermal reservoir enhancement. *Bull. Seismol. Soc. Am.*, 103(1), 383-393. https://doi.org/10.1785/0120120102.

[111] Michael, A. J., & Werner, M. J. (2018). Preface to the Focus Section on the Collaboratory for the Study of Earthquake Predictability (CSEP): New Results and Future Directions. *Seismol. Res. Lett.*, 89(4), 1226-1228. https://doi.org/10.1785/0220180161.

[112] Molchan, G. M. (1991). Structure of optimal strategies in earthquake prediction. *Tectonophys.*, 193, 267-276. https://doi.org/10.1016/0040-1951(91)90336-Q.

[113] Molchan, G. M., & Kagan, Y. Y. (1992). Earthquake prediction and its optimization. *J. Geophys. Res.*, 97, 4823-4838. https://doi.org/10.1029/91JB03095.

[114] Mousavi, M., Zhu, W., Sheng, Y., & Beroza, G. (2019). CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection. *Sci. Rep.* 9, 10267. https://doi.org/10.1038/s41598-019-45748-1.

[115] Norbeck, J. H., & Rubinstein, J. L. (2018). Hydromechanical earthquake nucleation model forecasts onset, peak, and falling rates of induced seismicity in Oklahoma and Kansas. *Geophys. Res. Lett.*, 45, 2963-2975. https://doi.org/10.1002/2017GL076562.

[116] Okada, Y. (1992). Internal deformation due to shear and tensile faults in a half-space. *Bull. Seismol. Soc. Am.* 82, 1018-1040.

[117] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.*, 83(401), 9-27.

[118] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Stat. Math.*, 50(2), 379-402. https://doi.org/10.1023/A:1003403601725.

[119] Ogata, Y., & Zhuang, J. (2006). Space-time ETAS models and an improved extension. *Tectonophysics* 413(1-2), 13-23.

[120] Omi, T., Ogata, Y., Shiomi, K., Enescu, B., Sawazaki, K., & Aihara, K. (2016). Automatic Aftershock Forecasting: A Test Using Real-Time Seismicity Data in Japan. *Bull. Seismol. Soc. Am.*, 106(6), 2450-2458. https://doi.org/10.1785/0120160100.

[121] Omi T., Ogata, Y., Shiomi, K., Enescu, B., Sawazaki, K., & Aihara, K. (2019). Implementation of a real-time system for automatic aftershock forecasting in Japan. *Seismol. Res. Lett.* 90(1), 242-250. https://doi.org//10.1785/0220180213.

[122] Parsons, T., Ogata, Y., Zhuang, Z., & Geist, E. L. (2012). Evaluation of static stress change forecasting with prospective and blind tests. *Geophys. J. Int.*, 188, 1425-1440. https://doi.org/10.1111/j.1365-246X.2011.05343.x.

[123] Parsons, T., & Segou, M. (2014). Stress, Distance, Magnitude, and Clustering Influences on the Success or Failure of an Aftershock Forecast: The 2013 M6.6 Lushan Earthquake and Other Examples. *Seismol. Res. Lett.* 85(1), 44-51. https://doi.org/10.1785/0220130100.

[124] Parsons, T., Segou, M., Sevilgen, V., Milner, K., Field, E., Toda, S., & Stein, R. S. (2014). Stress based aftershock forecasts made within 24h postmain shock: Expected north San Francisco Bay area seismicity changes after the 2014 M=6.0 West Napa earthquake. *Geophys. Res. Lett.*, 41, 8792-8799. https://doi.org/10.1002/2014GL062379.

[125] Parsons, T., Stein, R. S., Simpson, R. W., & Reasenberg, P. A. (1999). Stress sensitivity of fault seismicity: a comparison between limited-offset oblique and major strike-slip faults. *J. Geophys. Res.* 104, 20183-20202.

[126] Perfettini, H. (2004). Postseismic relaxation driven by brittle creep: A possible mechanism to reconcile geodetic measurements and the decay rate of aftershocks, application to the Chi-Chi earthquake, Taiwan. *J. Geophys. Res.* 109, B02304. https://doi.org//10.1029/2003JB002488.

[127] Rice, J. R. (1992). Fault stress states, pore pressure distributions, and the weakness of the San Andreas Fault, in Fault mechanics and transport proper-

ties of rocks; a festschrift in honour of W. F. Brace, edited by B. Evans & T. Wong, Academic Press, San Diego, CA, USA., 475-503.

[128] Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., & Imoto, M. (2011). Efficient testing of earthquake forecasting models. *Acta Geophysica*, 59(4), 728-747. https://doi.org/10.2478/s11600-011-0013-5.

[129] Richards-Dinger, K. B., & Shearer, P. M. (2000). Earthquake locations in southern California obtained using source-specific station terms. *J. Geophys. Res.*, 105, 10, 939-10, 960.

[130] Richter, C.F. (1958). Elementary Seismology (1st First Edition). San Francisco, CA: W.H. Freeman & Co.

[131] Rong, Y., Bird, P., & Jackson, D. D. (2016). Earthquake potential and magnitude limits inferred from a geodetic strain-rate model for southern Europe. *Geophys. J. Int.*, 205, 509-522. https://doi.org/10.1093/gji/ggw018.

[132] Ross, Z., Rollins, C., Cochran, E. S., Hauksson, E., Avouac, J.-P, & Ben-Zion, Y. (2017). Aftershocks driven by afterslip and fluid pressure sweeping through a fault-fracture mesh. *Geophys. Res. Lett.*, 44, 8260-8267. https://doi.org/10.1002/2017GL074634.

[133] Ross, Z., Meier, M. A., & Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. *J. Geophys. Res. Solid Earth* 122, 5120-5129. https://doi.org/10.1029/2017JB015251.

[134] Ross, Z., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, Z., Zhan, Z., Simons, M., Fielding, E. J., Yun, S., Hauksson, E. et al. (2019). Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence. *Science* 366, 346-351.

[135] Rovida A., Locati M., Camassi R., Lolli B., & Gasperini P. (Editors) (2016). CPTI15, the 2015 Version of the Parametric Catalogue of Italian Earthquakes, Istituto Nazionale di Geofisica e Vulcanologia. https://doi.org/10.6092/INGV.IT-CPTI15.

[136] Savran, W. H., Werner, M. J., Marzocchi, W., Rhoades, D. A., Jackson, D. D., Milner, K., Field, E., & Michael, A. (2020). Pseudoprospective Evaluation

of UCERF3-ETAS Forecasts during the 2019 Ridgecrest Sequence. *Bull. Seismol. Soc. Am.* 110(4), 1799-1817. https://doi.org/10.1785/0120200026.

[137] Schneider, M., Clements, R., Rhoades, D., & Schorlemmer, D. (2014). Likelihood- and residual-based evaluation of medium-term earthquake forecast models for California. *Geophys. J. Int.*, 198, 1307-1318. https://doi.org/doi:10.1093/gji/ggu178.

[138] Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., & Rhoades, D. A. (2007). Earthquake likelihood model testing. *Seismol. Res. Lett.*, 78(1), 17-29. https://doi.org/doi:10.1785/gssrl.78.1.17.

[139] Schorlemmer, D., Mele, F., & Marzocchi, W. (2010). A completeness analysis of the National Seismic Network of Italy. *J. Geophys. Res.*, 115, B04308. https://doi.org/10.1029/2008JB006097.

[140] Schultz, R., Skoumal, R. J., Brudzinski, M. R., Eaton, D., Baptie, B., & Ellsworth, W. (2020). Hydraulic fracturing-induced seismicity. *Rev. Geophys.*, 58, e2019RG000695. https://doi.org/10.1029/2019RG000695.

[141] Scognamiglio, L., Tinti, E., & Quintiliani, M. (2016). The first month of the 2016 Amatrice seismic sequence: Fast determination of time domain moment tensors and finite fault model analysis of the ML 5.4 aftershock. *Ann. Geophys.*, 59. https://doi.org/10.4401/ag-7246.

[142] Segou, M., Parsons, T., & Ellsworth, W. (2013). Comparative evaluation of physics-based and statistical forecasts in Northern California. *J. Geophys. Res. Solid Earth*, 118, 6219-6240. https://doi.org/10.1002/2013JB010313.

[143] Segou, M. (2016). Physics-based and statistical earthquake forecasting in a continental rift zone: the case study of Corinth Gulf (Greece). *Geophys. J. Int.*, 204, 591-605. https://doi.org/10.1093/gji/ggv467.

[144] Segou, M., & Parsons, T. (2016). Prospective Earthquake Forecasts at the Himalayan Front after the 25 April 2015 M=7.8 Gorkha Mainshock. *Seismol. Res. Lett.*, 87(4). https://doi.org/10.1785/0220150195.

[145] Segou, M., & Parsons, T. (2020). A New Technique to Calculate Earthquake Stress Transfer and to Probe the Physics of Aftershocks. *Bull. Seismol. Soc. Am.* 110(2), 863-873. https://doi.org/10.1785/0120190033.

[146] Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating ETAS: The effects of truncation, missing data, and model assumptions. *J. Geophys. Res. Solid Earth*, 121, 449-469. https://doi.org/10.1002/2016JB012809.

[147] Serpelloni, E., Anzidei, M., Baldi, P., Casula, G., & Galvani, A. (2005). Crustal velocity and strain-rate fields in Italy and surrounding regions: New results from the analysis of permanent and non-permanent GPS networks. *Geophys. J. Int.*, 161(3), 861-880.

[148] Shapiro, S. A., Dinske, C., & Kummerow, J. (2007). Probability of a given-magnitude earthquake induced by a fluid injection. *Geophys. Res. Lett.*, 34, L22, 314. https://doi.org/10.1029/2007GL031615.

[149] Shapiro, S. A., Dinske, C., & Langenbruch, C. (2010). Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations. *Lead. Edge*, 29(3), 304-309. https://doi.org/10.1190/1.3353727.

[150] Steacy, S., Nalbant, S. S., McCloskey, J., Nostro, C., Scotti, O., & Baumont, D. (2005). Onto what planes should Coulomb stress perturbations be resolved? *J. Geophys. Res.*, 110, B05S15. https://doi.org/10.1029/2004JB003356.

[151] Steacy, S., Gerstenberger, M., Williams, C., Rhoades, D., & Christophersen, A. (2014). A new hybrid Coulomb/statistical model for forecasting aftershock rates. *Geophys. J. Int.* 196, 918-923. https://doi.org/10.1093/gji/ggt404.

[152] Stein, R. S., King, G. C. P., & Lin, J. (1992). Change in failure stress on the southern San Andreas fault system caused by the 1992 magnitude = 7.4 Landers earthquake. *Science* 258, 1328-1332.

[153] Stein, R. S., King, G. C. P., & Lin, J. (1994). Stress triggering of the 1994 M=6.7 Northridge, California, earthquake by its predecessors. *Science*, 265, 1432-1435.

[154] Stein, R. S. (1999). The role of stress transfer in earthquake occurrence. *Nature*, 402, 605-609. https://doi.org/10.1038/45144.

[155] Stein, S., & Liu M. (2009). Long aftershock sequences within continents and implications for earthquake hazard assessment. *Nature Letters*, 462(5), 87-89. https://doi.org/10.1038/nature08502.

[156] Strader, A. E. (2014). The Effects of Coulomb Stress Change on Southern California Earthquake Forecasting. UCLA. Retrieved from https://escholarship.org/uc/item/45g4p4b6.

[157] Stucchi, M., Pinho, R., & Cocco, M. (2016). After the L'Aquila Trial. *Seismol. Res. Lett.* 87(3), 591-596. https://doi.org/10.1785/0220150261.

[158] Tinti, E., Scognamiglio, L., Michelini, A., & Cocco, M. (2016). Slip heterogeneity and directivity of the ML 6.0, 2016, Amatrice earthquake estimated with rapid finite-fault inversion. *Geoph. Res. Lett.*, 43. https://doi.org/10.1002/2016GL071263.

[159] Toda, S., Stein, R. S., Reasenberg, P. A., & Dieterich, J. H. (1998). Stress transferred by the 1995 Mw=6.9 Kobe, Japan, shock: Effect in aftershocks and future earthquake probabilities. *J. Geophys. Res.*, 103(B10), 543-565.

[160] Toda, S., Stein, R. S., Richards-Dinger, K., & Bozkurt, B. S. (2005). Forecasting the evolution of seismicity in southern California: Animations built on earthquake stress transfer. *J. Geophys. Res.*, 110, B05S16. https://doi.org/10.1029/2004JB003415.

[161] Toda, S., & Enescu, B. (2011). Rate/state Coulomb stress transfer model for the CSEP Japan seismicity forecast. *Earth Planets Space*, 63, 171-185. https://doi.org/10.5047/eps.2011.01.004.

[162] Toda, S., Stein, R. S., Sevilgen, V., & Lin, J. (2011). Coulomb 3.3 Graphic-rich deformation and stress-change software for earthquake, tectonic, and volcano research and teaching - user guide: U.S. Geological Survey Open-File Report 2011-1060, 63 p., http://pubs.usgs.gov/of/2011/1060/.

[163] Toda, S., & Stein, R. S. (2018). Why aftershock duration matters for probabilistic seismic hazard assessment. *Bull. Seismol. Soc. Am.* 108(3A), 1414-1426. https://doi.org/10.1785/0120170270.

[164] Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophys. Mag.*, 30, 521-605.

[165] Veen, A. & Schoenberg, F. P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *J. Am. Stat. Assoc.* 103, 614-624.

[166] Verdon, J., & Budge, J. (2018). Examining the Capability of Statistical Models to Mitigate Induced Seismicity during Hydraulic Fracturing of Shale Gas Reservoirs. *Bull. Seismol. Soc. Am.*, 108(2), 690-701. https://doi.org/10.1785/0120170207.

[167] Waldhauser, F., & Ellsworth, W. L. (2000). A Double-Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California. *Bull. Seismol. Soc. Am.*, 90(6), 1353-1368.

[168] Waldhauser, F. (2009). Near-Real-Time Double-Difference Event Location Using Long-Term Seismic Archives, with Application to Northern California. *Bull. Seismol. Soc. Am.* 99(5), 2736-2748. https://doi.org/10.1785/0120080294.

[169] Wang, K., Hu, Y., & He, J. (2012). Deformation cycles of subduction earthquakes in a viscoelastic Earth. *Nature* 484(7394), 327-32. https://doi.org/10.1038/nature11032.

[170] Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull. Seismol. Soc. Am.*, 84, 974-1002.

[171] Werner, M. J., Helmstetter, A., Jackson, D. D., & Kagan, Y. Y. (2011). High-resolution long-term and short-term earthquake forecasts for California. *Bull. Seismol. Soc. Am.*, 101(4), 1630-1648. https://doi.org/doi:10.1785/0120090340.

[172] Wiemer, S. (2001). A software package to analyze seismicity: ZMAP. *Seismol. Res. Lett.*, 72 (3), pp.373-382. https://doi.org/10.1785/gssrl.72.3.373.

[173] Wiemer, S., & Wyss, M. (2000). Minimum magnitude of complete reporting in earthquake catalogs: examples from Alaska, the western United States, and Japan. *Bull. Seismol. Soc. Am.*, 90, 859. https://doi.org/10.1785/0119990114.

[174] Woessner, J. & Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bull. Seism. Soc. Am.*, 95(2), 684-698.

[175] Woessner, J., Hainzl, S., Marzocchi, W., Werner, M. J., Lombardi, A. M., Catalli, F., Enescu, B., Cocco, M., Gerstenberger, M.C., & Wiemer, S. (2011). A retrospective comparative forecast test on the 1992 Landers sequence. *J. Geophys. Res.*, 116, B05305. https://doi.org/10.1029/2010JB007846.

[176] Woessner, J., Jonsson, S., Sudhaus, H., & Baumann, C. (2012). Reliability of Coulomb stress changes inferred from correlated uncertainties of finite-fault source models. *J. Geophys. Res.*, 117, B07303. https://doi.org/10.1029/2011JB009121.

[177] Woessner, J., Laurentiu, D., Giardini, D., Crowley, H., Cotton, F., Grunthal, G., et al. (2015). The 2013 European Seismic Hazard Model: key components and results. *Bull Earthquake Eng.*, 13, 3553-3596. https://doi.org/10.1007/s10518-015-9795-1.

[178] Wyss, M., Hasegawa, A., Wiemer, S., & Umino, N. (1999). Quantitative mapping of precursory seismic quiescence before the 1989, M7.1 Osanriku earthquake, Japan. *Annali Di Geofisica*, 42, 851.

[179] Yang, W., Hauksson, E., & Shearer, P. M. (2012). Computing a large refined catalog of focal mechanisms for southern California (1981-2010): Temporal stability of the style of faulting. *Bull. Seismol. Soc. Am.*102(3), 1179-1194.

[180] Zechar, J. D., & Jordan, T. H. (2008). Testing alarm-based earthquake predictions. *Geophys. J. Int.*, 172, 715-724. https://doi.org/10.1111/j.1365-246X.2007.03676.x.

[181] Zechar, J. D. (2010). Evaluating earthquake predictions and earthquake forecasts: a guide for students and new researchers. *Community Online Resource for Statistical Seismicity Analysis*. https://doi.org/10.5078/corssa-77337879.

[182] Zechar, J. D., Gerstenberger, M. C., & Rhoades, D. A. (2010). Likelihood based tests for evaluating space-rate-magnitude earthquake forecasts. *Bull. Seismol. Soc. Am.*, 100, 1184-1195. https://doi.org/10.1785/0120090192.

[183] Zechar, J. D., & Zhuang, J. (2014). A parimutuel gambling perspective to compare probabilistic seismicity forecasts. *Geophys. J. Int.*, 199, 60-68. https://doi.org/10.1093/gji/ggu137.

[184] Zhang, L., Werner, M. J., & Goda, K. (2020). Variability of ETAS parameters in global subduction zones and applications to mainshock-aftershock hazard assessment. *Bull. Seismol. Soc. Am.*, 110(1), 191-212. https://doi.org/10.1785/0120190121.

[185] Zhuang, J., Ogata, Y., & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *J. Am. Stat. Assoc.*, 97(458), 369-380. https://doi.org/10.1198/016214502760046925.

[186] Zhuang, J., Harte, D., Werner, M. J., Hainzl, S., & Zhou, S. (2012). Basic models of Seismicity: Temporal Models. *Community Online Resource for Statistical Seismicity Analysis*. https://doi.org/10.5078/corssa-79905851.

[187] Zhuang, J., & Touati, S. (2015). Stochastic simulation of earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis*. https://doi.org/10.5078/corssa-43806322.