



**British
Geological Survey**
Expert | Impartial | Innovative



University of
BRISTOL

Statistical Modelling of the Preston New Road Seismicity: Towards Probabilistic Forecasting Tools

Earth Hazards Programme

Commissioned Report CR/19/068

BRITISH GEOLOGICAL SURVEY

EARTH HAZARDS PROGRAMME

COMMISSIONED REPORT CR/19/068

Statistical Modelling of the Preston New Road Seismicity: Towards Probabilistic Forecasting Tools

S. Mancini, M. Segou, M.J. Werner, B.J. Baptie

Keywords

Report; induced seismicity;
hydraulic fracturing; magnitude
distribution; probabilistic
forecasting.

Bibliographical reference

MANCINI, S., SEGOU, M.,
WERNER, M.J., BAPTIE, B.J.
2019.
Statistical Modelling of the
Preston New Road Seismicity:
Towards Probabilistic
Forecasting Tools. *British
Geological Survey
Commissioned Report*,
CR/19/068. 41pp.

Copyright in materials derived
from the British Geological
Survey's work is owned by
UK Research and Innovation
(UKRI) and/or the authority that
commissioned the work. You
may not copy or adapt this
publication without first
obtaining permission. Contact the
BGS Intellectual Property Rights
Section, British Geological
Survey, Keyworth,
e-mail ipr@bgs.ac.uk. You may
quote extracts of a reasonable
length without prior permission,
provided a full acknowledgement
is given of the source of the
extract.

Maps and diagrams in this book
use topography based on
Ordnance Survey mapping.

BRITISH GEOLOGICAL SURVEY

The full range of our publications is available from BGS shops at Nottingham, Edinburgh, London and Cardiff (Welsh publications only) see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications, including maps, for consultation.

We publish an annual catalogue of our maps and other publications; this catalogue is available online or from any of the BGS shops.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as basic research projects. It also undertakes programmes of technical aid in geology in developing countries.

The British Geological Survey is a component body of UK Research and Innovation.

British Geological Survey offices

Environmental Science Centre, Keyworth, Nottingham NG12 5GG

Tel 0115 936 3100

BGS Central Enquiries Desk

Tel 0115 936 3143

email enquiries@bgs.ac.uk

BGS Sales

Tel 0115 936 3241

email sales@bgs.ac.uk

The Lyell Centre, Research Avenue South, Edinburgh EH14 4AP

Tel 0131 667 1000

email scotsales@bgs.ac.uk

Natural History Museum, Cromwell Road, London SW7 5BD

Tel 020 7589 4090

Tel 020 7942 5344/45

email bgs_london@bgs.ac.uk

Cardiff University, Main Building, Park Place, Cardiff CF10 3AT

Tel 029 2167 4280

Maclean Building, Crowmarsh Gifford, Wallingford OX10 8BB

Tel 01491 838800

Geological Survey of Northern Ireland, Department of Enterprise, Trade & Investment, Dundonald House, Upper Newtownards Road, Ballymiscaw, Belfast, BT4 3SB

Tel 01232 666595

www.bgs.ac.uk/gsni/

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon SN2 1EU

Tel 01793 411500

Fax 01793 411501

www.nerc.ac.uk

UK Research and Innovation, Polaris House, Swindon SN2 1FL

Tel 01793 444000

www.ukri.org

Website www.bgs.ac.uk

Shop online at www.geologyshop.com

Foreword

This report is the published product of a study by the British Geological Survey (BGS) and the University of Bristol that was commissioned by the Oil and Gas Authority (OGA) to research the feasibility of statistically forecasting the microseismicity observed during and after unconventional shale gas exploration in 2018 by Cuadrilla Resources at its Preston New Road site near Blackpool, UK.

Contents

Foreword	i
Contents	i
Plain Language Summary	iv
Extended Plain Language Synopsis	v
1 Introduction	1
2 Epidemic Type Aftershock Sequence (ETAS) Model Forecasts	1
2.1 Background.....	2
2.2 Technical Description of the ETAS Framework	6
2.3 Statistical Validation of Forecast Models.....	7
3 Summary of Operations	8
4 Data and Exploratory Analysis	8
4.1 Event behaviour in space and time	9
4.2 Frequency Magnitude Distribution.....	11
4.3 Distribution of the INTER-EVENT TIMES	14
5 Model Calibration and Simulation Framework	16
5.1 Magnitude Distribution, b-value and Maximum Magnitude.....	16
5.2 Etas Model Calibration	16
5.3 Residual Analysis (Goodness-of-fit)	17
5.4 Forecasts based on ETAS Simulations	19
6 Performance Evaluation of the Forecasts	19
6.1 Daily ETAS- $1M_{cut}$ Model	20
6.2 hourly ETAS- $1M_{cut}$ and ETAS- $1.5M_{cut}$ Models	20
6.3 Comparison of Frequency Magnitude Distributions	24
7 Conclusions	24
References	26

FIGURES

Figure 4-1. Downhole acquisition geometry showing locations of sleeves in well PNR-1z (red squares) and the geophones in PNR-2 (blue squares). (a) plan view, (b) East-West cross-section and (c) North-South cross section. 8

Figure 4-2. Comparison of the M_w estimates from the downhole catalogue with the M_L estimates for matching events in the surface catalogue. The red lines shows the M_L/M_w conversion relationship given the in Cuadrilla HFP (2019)..... 9

Figure 4-3. (a) Map of all events in the microseismic catalogue. Events are coloured by time and scaled by magnitude. The coloured squares show the locations of the sleeves that were hydraulically fractured. The squares are coloured using the same colour scale as the events. Axes show British National Grid Eastings and Northings. Grey squares show geophone positions. (b) Depth cross-section showing event depths along an east-west profile..... 10

Figure 4-4. Histograms showing the number of events/hour as a function of time during operations. The grey shaded bars show all events, the red bars show events with magnitudes of $-1.5 M_w$ or above. Blue lines show the cumulative volume of injected fluid during hydraulic fracturing operations. (a) shows seismicity in October/November and (b) shows activity in December. No hydraulic fracturing was carried out between 3 November and 4 December..... 11

Figure 4-5. Histograms showing the number of events/hours as a function of time during operations. The grey shaded bars show all events, the red bars show events with magnitudes of $-1.5 M_w$ or above. Blue lines show the cumulative volume of injected fluid during hydraulic fracturing operations..... 12

Figure 4-6. b -value (a) and a -values estimates as a function of completeness magnitude M_c for the entire downhole catalogue. 13

Figure 4-7. Frequency magnitude distributions for incremental (red) and cumulative data (blue). The blue dashed lines show the maximum likelihood estimates of the b -value and activity rate for completeness magnitudes of -1.0 (a) and -1.5 (b)..... 13

Figure 4-8. Frequency magnitude distributions, b -values and activity rates calculated using events that occurred during each stage of operations. B -values and activity rates were calculated using $M_c = -1.0$ 14

Figure 4-9. Inter-event time distributions of the PNR-1z microseismic catalogue above the completeness magnitude $M_c=-1$. Red line indicates Omori law temporal clustering with p -value of unity. (a) Stage 3, combining a mini and main frack. (b) During the pause in operations between 3 November and 7 December. (c) Stage 37. (d) Entire catalogue. Panels (a)-(c) show inter-event times normalised to the average rate during the period..... 15

Figure 5-1. Top: Magnitudes above $M_c=-1$ and occurrence times of the PNR-1z microseismicity between August and December 2018. Bottom: calibrated conditional intensity function (instantaneous probability) of $ETAS_{-1M_{cut}}$ illustrating its variability and the cumulative effect of individual contributions during particularly active periods. 18

Figure 5-2. Residual analysis showing goodness-of-fit of the calibrated model. The black lines shows the observed versus the expected number of events, obtained by integrating the calibrated conditional intensity function with the continuously updated history. The dashed red line indicates a perfect fit. The goodness-of-fit is good overall, but shows underestimation between event numbers 500 and 1000. 19

Figure 6-1. Forecast time series and observed vs. expected number of events for the daily updated $ETAS_{-1M_{cut}}$ model. (a) Observed (circles) and mean forecasted (squares) incremental number of $M \geq -1$ events for the 64 days of operations. The yellow bars represent the 95% confidence bounds of the ETAS forecast. The colour of the circles indicates whether the

observed number of events per daily time bin falls within (green) or outside (red) the ETAS confidence bounds. (b) Observed (black line) and mean expected (blue line) cumulative seismicity with shaded areas representing Poissonian uncertainties. (c) Observed vs. expected number of events; red circles indicate forecast rejection, while green ones indicate that the model adequately forecasts the number of events. 21

Figure 6-2. Forecast time series and observed vs. expected number of events for the hourly updated ETAS- $1M_{cut}$ model. (a) Observed (circles) and mean forecasted (squares) incremental number of $M \geq -1$ events for the 64 days of operations. The yellow bars represent the 95% confidence bounds of the ETAS forecast. The colour of the circles indicate whether the observed number of events per daily time bin falls within (green) or outside (red) the ETAS confidence bounds. (b) Observed (black line) and mean expected (blue line) cumulative seismicity with shaded areas representing Poissonian uncertainties. (c) Observed vs. expected number of events; red circles indicate forecast rejection, while green ones indicate that the model adequately forecasts the number of events. 22

Figure 6-3. Forecast time series and observed vs. expected number of events for the hourly updated ETAS- $1.5M_{cut}$ model. (a) Observed (circles) and mean forecasted (squares) incremental number of $M \geq -1$ events for the 64 days of operations. The yellow bars represent the 95% confidence bounds of the ETAS forecast. The colour of the circles indicate whether the observed number of events per daily time bin falls within (green) or outside (red) the ETAS confidence bounds. (b) Observed (black line) and mean expected (blue line) cumulative seismicity with shaded areas representing Poissonian uncertainties. (c) Observed vs. expected number of events; red circles indicate forecast rejection, while green ones indicate that the model adequately forecasts the number of events. 23

Figure 6-4. Observed vs. simulated frequency-magnitude distributions for the daily and hourly updated ETAS- $1M_{cut}$ and hourly updated ETAS- $1.5M_{cut}$ models. 24

Plain Language Summary

The present authors were commissioned by the Oil and Gas Authority (OGA) to research the feasibility of statistically forecasting the microseismicity observed during and after unconventional shale gas exploration in 2018 by Cuadrilla Resources at its Preston New Road site near Blackpool, UK. The scope of the research was limited by the available time and resources, and we focussed on the following: (i) a statistical analysis of the microseismicity catalogues recorded by the surface and downhole network of seismometers, including an analysis of the observed magnitude distribution and level of recording completeness, (ii) the calibration of a widely used temporal statistical model of tectonic seismicity to the present data, (iii) the generation and evaluation of probabilistic forecasts of the variable event rate and magnitude distribution as simulated by the model, (iv) an assessment of the geological and physical processes that are not (yet) captured by the statistical model. Our findings can be summarised as follows.

The microseismic datasets collected at the well PNR-1z present globally unique opportunities to understand and model subsurface processes in response to hydraulic fracturing, including induced seismicity. The insights are likely to contribute to improved operations and risk mitigation strategies. The two magnitude scales used by the surface and the downhole networks present a challenge because they are difficult to reconcile. This implies that the standard extrapolation of seismicity rates observed at small magnitudes to the rates and chances of larger magnitudes, including those that might be felt or cause damage, is more uncertain than in the case of tectonic seismicity. We observe a b -value of the Gutenberg-Richter frequency magnitude distribution of about 1.5, which appears to be robustly larger than b -values of tectonic seismicity, notwithstanding issues with the magnitude scales. We estimate that the downhole catalogue is approximately complete down to a moment magnitude of -1. We adopted a precautionary maximum magnitude of 6.5 on the basis of the most likely tectonic maximum magnitude according to the most recent national seismic hazard map. A more context-specific assessment of maximum magnitude is difficult and was beyond the scope of this project.

The calibration of the statistical seismicity model highlighted similarities and differences between tectonic and induced seismicity. Several model parameters were similar to those found in tectonic applications, but some were substantially different. Similarities include clear evidence for temporal clustering of seismicity and the applicability of the Gutenberg-Richter distribution, as well as aspects of the seismic source parameters themselves. Differences include that the parameters suggest that the magnitude of an event has little effect on the number of subsequently triggered events, in clear contrast to the magnitude-dependent aftershock generation of tectonic earthquakes. In addition, we observed a more rapid decay of temporal clustering than is customary for tectonic sequences, which may indicate a more dominant role of fluids in the clustering characteristics of the microseismicity. We caution that we examined a relatively narrow range of magnitudes above the completeness magnitude of -1, and further work is needed to understand the robustness of these conclusions, ideally on an independent dataset.

The performance of the forecasts generated by the statistical model highlighted some successes and some discrepancies. In broad terms, the forecasts performed well during periods with moderate seismic response to hydraulic fracturing as well as in between stages and during the prolonged halt of operations in November 2018 during which the seismicity decayed slowly. On the other hand, the model frequently underpredicted the most seismically intense periods. Whether the current forecasts are good enough to be useful depends on the particular usage as well as other available predictive tools, and would require discussions with potential users.

We interpret the mixed performance of the forecasts as a result of features in the induced seismicity that are rare in natural seismicity. Specifically, events induced during fracturing do not appear to show the type of magnitude-dependent triggering and clustering observed in natural seismicity.

Instead, we observe during these times a high rate of apparently unclustered seismicity. This feature is not currently built into the model.

In our view, the present findings warrant further research, ideally on independent but comparable datasets. The issue of the irreconciled magnitude scales requires solutions. The present forecasting model should be tested on an independent, out-of-sample dataset to better mimic real-time situations. The 2019 dataset from PNR-2 would be ideal for this purpose. The apparent lack of magnitude-dependent clustering in the induced microseismicity needs verification with more data, e.g. below the magnitude threshold of -1.5 Mw that we analysed and on other datasets. Next, the model can be modified to accommodate the strong but unclustered seismicity during stages by linking it to injection data. Finally, the physical process of seismic slip nucleation in response to pore pressure changes can be combined with and incorporated in the present statistical model. We believe this further research would contribute to improved probabilistic predictive tools and seismic risk mitigation tools.

Extended Plain Language Synopsis

Cuadrilla carried out hydraulic fracturing of the well PNR-1z between August and December 2018 for the purpose of shale gas exploration. Surface and borehole monitoring networks operated by the British Geological Survey (BGS) and Cuadrilla together detected over 38,000 microseismic events during this period, including the largest event of local magnitude 1.6 ML. Seven events equal to or greater than magnitude 0.5 ML occurred, exceeding the UK regulatory threshold that requires operators to halt and pause operations.

The collected datasets present globally unique opportunities to understand and model microseismicity induced by hydraulic fracturing operations. The surface arrays detected around 170 events with magnitudes given in the local magnitude scale, while the more sensitive downhole array recorded over 38,000 events with magnitudes given in a moment magnitude scale. Seismologists use different magnitude scales depending on the type of network sensors, noise levels and properties of the events. The UK's regulatory framework uses local magnitudes.

The use of two different magnitude scales complicated the analysis of the magnitude distribution as well as the subsequent modelling. A consistent dataset is required for analysis and modelling. Existing conversion relationships between moment and local magnitude do not match the observed data well. When applied, they introduce artefacts in the magnitude dataset. The process of fitting the observed magnitudes with the common Gutenberg-Richter distribution, as well as extrapolating the observed rates at low magnitudes to rates at higher magnitudes, is therefore subject to greater uncertainty than usual. We recommend further work to solve this issue and to quantify the added uncertainty in extrapolations.

Tectonic (natural) seismicity is widely observed to obey the Gutenberg-Richter law, which states that the frequency of events decreases by a constant factor for each unit increase in magnitude. This factor is nearly universally reported to be about 10 (corresponding to the seismological b -value of one). For example, magnitude 5 events are 10 times less frequent than magnitude 4 events, and so on, over the entire range of observed tectonic earthquakes.

The PNR-1z dataset obeys the Gutenberg-Richter law, too, suggesting that the eventual magnitudes of the induced events were controlled by similar processes as those of natural events. Interestingly, and as reported previously elsewhere, the factor by which the frequency of larger magnitudes decreases is greater: instead of a factor of 10, the frequency decreased by a factor of about 32 per unit magnitude. It is conceivable, but unlikely, that this difference is an artefact of the different magnitude scales. This is important because the factor is nearly always used to extrapolate to the chances of larger hypothetical seismic events such as those that might be felt or cause damage. A high factor would decrease the chance of larger events relative to natural seismicity.

Our approach to forecasting recognises that deterministic predictions of individual seismic events are not currently possible, neither for tectonic earthquakes nor for induced events. Instead, statistical forecasts aim to construct a distribution of the number of events with associated chances of occurring over a time period and magnitude range of interest. In general, good forecasts provide a narrow range that matches the observations well. Poor forecasts provide either very broad ranges and are uninformative, or they provide narrow ranges that do not match the data. Whether or not forecasts are good enough to be useful depends on the particular use of the forecasts. For example, even relatively uninformative forecasts may still be useful for taking low-cost actions, such as simply providing information, e.g. to the operators, regulators and the public, about a particular risk view of future seismic hazards. More informative and reliable forecasts may warrant more costly mitigation actions, such as changing or halting operations.

To forecast the PNR-1z microseismicity, we used a statistical model originally developed for tectonic seismicity. The Epidemic Type Aftershock Sequence (ETAS) model offers many advantages, including relative simplicity, flexible calibration, wide usage and a record of outperforming other models in a range of tectonic settings. Several non-UK government agencies employ the model as part of operational forecasting systems, including the US Geological Survey, the Italian National Institute of Geophysics and Volcanology (INGV) and New Zealand's GNS Science. The ETAS model has also been used with some success in the context of induced seismicity in the central US, in China and elsewhere.

The ETAS model describes seismicity as a result of two processes: spontaneous earthquakes generated by slow, steady, plate tectonic loading; and earthquakes triggered by previous earthquakes. The spontaneous earthquakes occur randomly in time according to a constant long-term rate. Each spontaneous earthquake can trigger earthquakes according to relationships observed between earthquakes and their aftershocks: larger magnitudes on average generate exponentially more aftershocks; the aftershock frequency decays approximately inversely with time; every earthquake can trigger its own (secondary, tertiary, etc.) aftershocks; and, crucially, the magnitudes of all events obey the Gutenberg-Richter law and are random, i.e. they do not depend on the history, time, location or ancestor earthquake. That is, any earthquake has a (small) chance of being large.

To assess the suitability of the ETAS model to induced microseismicity at PNR-1z, we calibrated the temporal ETAS model to the microseismic catalogue and then performed a retrospective forecasting experiment. The calibration consists of estimating the magnitude distribution and estimating five additional parameters of the ETAS model.

Estimating the magnitude distribution means determining the b -value (the above-mentioned factor by which the frequency of magnitudes decrease with increasing magnitude) and a maximum magnitude above which seismic events are extremely unlikely to happen. The above-mentioned factor of 32 translates to a b -value of about 1.5. We chose a precautionary maximum magnitude of 6.5 taken from the UK's national seismic hazard map because a more context-specific estimation of the maximum magnitude is challenging and beyond the scope of this project.

The estimation of the other five ETAS parameters, together with an analysis of temporal patterns of the seismicity, highlighted some of the similarities and differences between natural and induced seismicity. We observed temporal clustering of events in both cases, which provides some justification for the decision to apply a model designed for tectonic seismicity to this case of induced seismicity. Here, we observed a faster decay of triggered seismicity than in natural seismicity, which might be interpreted as resulting from a more prominent role of fluids in the generation of seismicity. In addition, the parameters appear to indicate the surprising result that larger magnitudes do not trigger more events than small ones, in stark contrast to tectonic earthquake sequences. Finally, we observed that during some stages a strong and unclustered rate of seismicity appears directly influenced by pore pressure changes due to injection rather than being triggered by previous seismicity.

We caution that further work is needed to understand the robustness of these findings, including testing the effect of the small range of magnitudes over which the catalogue is complete and analysing the parameter uncertainties and trade-offs. On balance, the former might be important, while the latter is unlikely to change the findings.

Despite the surprising parameter estimates, an analysis of the quality of the fit of the model to the data generally revealed a good match, with the exception of an underprediction of activity towards the second half of October. This match provides further justification for applying the flexible ETAS model, especially if future model improvements can be made.

We conducted three retrospective forecast experiments to assess the influence of two factors: the frequency at which forecasts are updated and the magnitude threshold above which events are forecast. In the context of tectonic seismicity, more frequent updating and a lower magnitude threshold have both improved forecast skill. The ETAS model is a stochastic model, which means that future seismicity is governed by probability distributions. Simulations create scenarios from these probability distributions. Here, the three forecast experiments entailed (1) forecasts for events greater than magnitude -1 updated daily or at the beginning of each injection stage (whichever came sooner), (2) forecasts for events greater than magnitude -1 updated hourly or 15 minutes after the beginning of each stage (whichever came sooner), and (3) forecasts for events greater than magnitude -1.5 updated hourly or 15 minutes after the beginning of each stage (whichever came sooner). At the start of each forecast period, we provide the model with the observed seismicity until then, and numerically simulate at least 1,000 scenarios over the next forecast period. The scenarios consist of occurrence times and magnitudes of events. From these, we compute the distribution of the number of events (or means, medians and 95% predictive intervals) and the magnitude distribution. We then compare these distributions to the observed data to check how well the data match the model simulations.

On balance, the performance of the ETAS forecasts showed frequent successes and some discrepancies. The daily forecasts (experiment 1) were least successful: in nearly one in five forecast periods, the observed number of events was greater than the 95% confidence range forecast by the model. In other words, the model underpredicted significantly in about 20% of the forecast windows. The hourly forecasts (experiments 2 and 3) were more successful. In experiment 2, the observed number of events were greater than the modelled 95% range in only 4% of the forecast windows. In experiment 3, about 8% of the time windows saw more events than were forecast by the model. Frequent updates are therefore important for reliable model forecasts. The forecasts slightly deteriorated when we lowered the magnitude threshold to -1.5 in experiment 3. The results of experiment 2, in particular, are encouraging: the modelled 95% range of the number of events in an hourly forecast window captures the observed number approximately 96% of the time. More qualitatively, the model also successfully captured some features of the data, namely the seismicity patterns characteristic of event-to-event triggering and the seismicity rate decays between stages that are similar to Omori's law of aftershock decay.

At the same time, however, the model sometimes underpredicted the number of events in the next forecast period, especially when that period contained stages with high and unclustered seismic rates. In short, the number of observed events fell into the modelled 95% confidence ranges 96% of the time, but during the 4% of windows when the observed number was outside the range, the observed number was larger than expected, and often substantially larger. (In the worst case, the model expected 14 events, the modelled 95% range was 0-40, but 98 events actually occurred.) We believe that these challenges result from the model's current inability to anticipate the variable seismic responses to hydraulic fracturing.

Despite their moderate performance, the ETAS model forecasts could potentially have been helpful for operators, regulators and the public as a legitimate albeit imperfect view of the seismic risk as part of a broader decision making framework. The forecasts may be used to monitor expectations and to influence some decisions, with the explicit caveat that spontaneously strong seismic responses to injection might occasionally be missed by the model before it can update.

In our view, the above findings present a clear case for further research, particularly involving the independent dataset collected during subsequent hydraulic fracturing in 2019 at well PNR-2. Approaches need to be developed to solve the current discrepancies between two magnitude scales. This will allow more robust conclusions regarding b -values, differences with tectonic seismicity (including the maximum magnitude) and the chances of larger events, including felt and damaging events. Next, the present evaluation of the forecast performance of the original ETAS model needs to be tested on an independent dataset that was not used to calibrate the model. This will provide more confidence in the model performance, in addition to further development opportunities. In addition, and as mentioned above, it is possible that the lack of magnitude-dependent triggering we observe is a result of the relatively narrow range of magnitudes we examined. This should be revisited with more data. If more typical (tectonic) triggering is observed, we would expect the forecast skill of the model to improve when lowering the magnitude threshold. Finally, the current ETAS model is flexible enough to accommodate desirable modifications that take into account operational parameters such as injected volume or pressures, and to be combined with physical models of the change in the pore pressure and stress field. The datasets collected at Preston New Road present globally unique opportunities to further develop and improve these and other predictive tools for real-time seismic risk mitigation of shale gas exploration and other fluid injection operations such as geothermal energy exploration and carbon capture and storage.

1 Introduction

Modelling the evolution of seismicity, natural or induced, is a challenge given our limited knowledge of the preparatory processes of seismic events. Recent advances in the development of earthquake forecast models can inform us about the spatio-temporal distribution of aftershocks in natural seismicity (Segou and Parsons, 2016). The ability to make reliable statistical forecasts of seismicity induced by hydraulic fracturing (HF) operations is highly desirable for the mitigation of risks of both disturbing and damaging ground motions. The large down-hole catalogue of micro-seismicity recorded at the Preston New Road (PNR) site during HF operations between October-December 2018 represents a unique opportunity to develop statistical forecast models that accurately describe the temporal distribution of the observed seismicity.

Epidemic Type Aftershock Sequence models (ETAS – Ogata, 1988) encode well-documented statistical observations of seismicity in a flexible, stochastic manner. Magnitudes are drawn randomly and independently from the Gutenberg-Richter law (e.g. Van der Elst et al., 2016). Each event can trigger further events according to its magnitude (large quakes trigger exponentially more events), and triggered events have a (small) chance of being larger than the parent event. This feature of the model is consistent with the observation that the largest magnitudes do not exclusively occur in early post-mainshock phases (natural seismicity) or at the beginning of hydraulic fracturing operations (induced seismicity).

Triggered seismicity is modelled to decay in time according to the Omori law, which characterises the rate of decay as an inverse power law. The model is completed with a rate of spontaneous events, which can generate triggered events that in turn can trigger subsequent generations. In the tectonic context, the spontaneous event rate is assumed to be constant. The ETAS formulation represents a powerful method to compute short-term occurrence probabilities of potentially damaging earthquakes, and the PNR calibrated model can be used in future stages, wells and developments.

Although ETAS was originally developed for natural seismicity, its application has been extended to induced seismicity. Recent efforts to model earthquakes associated with enhanced geothermal systems (e.g. Bachmann et al., 2011; Mena et al., 2013) and waste-water injection for oil and gas production (e.g. Llenos and Michael, 2013), showed the effectiveness of this approach to model seismicity patterns during and after injection phases and forecast robustly the observed seismicity in terms of the number distribution of events and the magnitude distribution.

The aim of this project was to evaluate the use of a time dependent ETAS model for making a statistical forecast of seismicity during hydraulic fracturing at PNR-1z. The primary input parameters for this model are dates, times, locations and magnitudes in the down-hole catalogue of micro-seismicity recorded at the Preston New Road (PNR) site during HF operations between October-December 2018, which consists of over 38,000 individual events with magnitudes as low as -2.5 Mw. We sought to assess if such a model could be used to inform future real-time decision-making and risk mitigation during HF operations at sites across the UK and elsewhere.

2 Epidemic Type Aftershock Sequence (ETAS) Model Forecasts

In this section we describe the previous use of statistical ETAS models with examples from tectonic settings (natural seismicity) as well as cases of induced seismicity (HF operations). We present the mathematical framework for the implementation of the model together with description of the parameterisation procedure and the statistical metrics used for the validation of the forecast models.

2.1 BACKGROUND

Although the prediction of future large earthquakes in natural seismicity settings is currently unattainable, the development of mathematical models that track the evolution of hazardous phenomena, such as earthquake cascades, are now within our reach (Jordan et al., 2011). Such mathematical models follow either empirical/statistical formulations, where the earthquake sequence is represented as a point process, or physical laws about earthquake nucleation. In this study we focus on the implementation of the ETAS model (Ogata, 1988), which corresponds to the most frequently used statistical descriptor of future earthquake probabilities.

The primary input parameters are dates, times, locations and magnitudes reported in standard earthquake catalogues. The model relies on the robust quantification of key factors of the seismicity such as the Gutenberg-Richter b -value, which describes the behaviour of the frequency-magnitude distribution for an earthquake population, the Omori p -value that characterizes how fast seismicity decays over time, as well as an exponent that describes the increase in triggering productivity with magnitude. The ETAS model is a widely used statistical model in which earthquake triggering follows a stochastic branching evolution and each earthquake produces its own offspring population, where triggered aftershocks with even larger magnitude than their parent events can occur at any time. This feature of the model is consistent with the observation that the maximum expected magnitude (M_{\max}) events do not exclusively occur in early post-mainshock phases (natural seismicity) or at the beginning of hydraulic fracturing operations (induced seismicity). To illustrate the importance behind a reliable estimate of the expected earthquake rates it should be noted that a 10-fold increase of the number of expected events with magnitude $M > 1.0$ results in a 100 times higher probability of observing at least one magnitude $M = 4$ event.

The success of forecast models to track the evolution of natural tectonic seismicity is often shown in pseudo-prospective experiments, where scientists are using data products that are usually available in real-time conditions, and/or fully retrospective experiments, where best-available data are employed to maximize the predictive skills of the models under test. Here we provide examples from recent research illustrating outputs of ETAS forecasts from California, Italy and elsewhere.

Werner *et al.* (2011) showed that large earthquakes tend to occur near the locations of small ($M > 2$) events by implementing different types of statistical forecasts for California, including an ETAS model, and testing their performance using the statistical metrics introduced by the Collaboratory for the Study of Earthquake Predictability (CSEP; Michael and Werner, 2018). They found that time-independent (Poissonian) earthquake rates (blue line; Figure 2-1) during earthquake crises fail to represent the short-term variability in the number of expected events. The results illustrate the importance behind the development of forecast models (red line; Figure 2-1) that can approximate the transient increase of observed seismicity rates (circles; Figure 2-1).

The time-independent earthquake rates assume that seismicity is constant with time, which may be an accurate descriptor of long-term (e.g. hundreds of years) average seismicity in a region of interest; often implemented in building codes or provisions for critical structures. Within shorter time intervals the observed long-term earthquake rate fluctuates due to short-lived (transient) perturbations of the natural system (e.g. tectonic earthquakes) or other type of local perturbations (e.g. anthropogenic activity). Within these time intervals the probability of occurrence of a given magnitude event may be many times higher than the long-term average suggesting a higher transient in nature seismic hazard. However, any increased seismic activity will return to the long-term average either slowly or rapidly, depending on the nature of the perturbation. The advantage of the ETAS model is the efficiency to determine within statistical significance these increased (above long-term average) seismicity rates within short-term (e.g. daily) time windows, here referred to as expected seismicity rates.

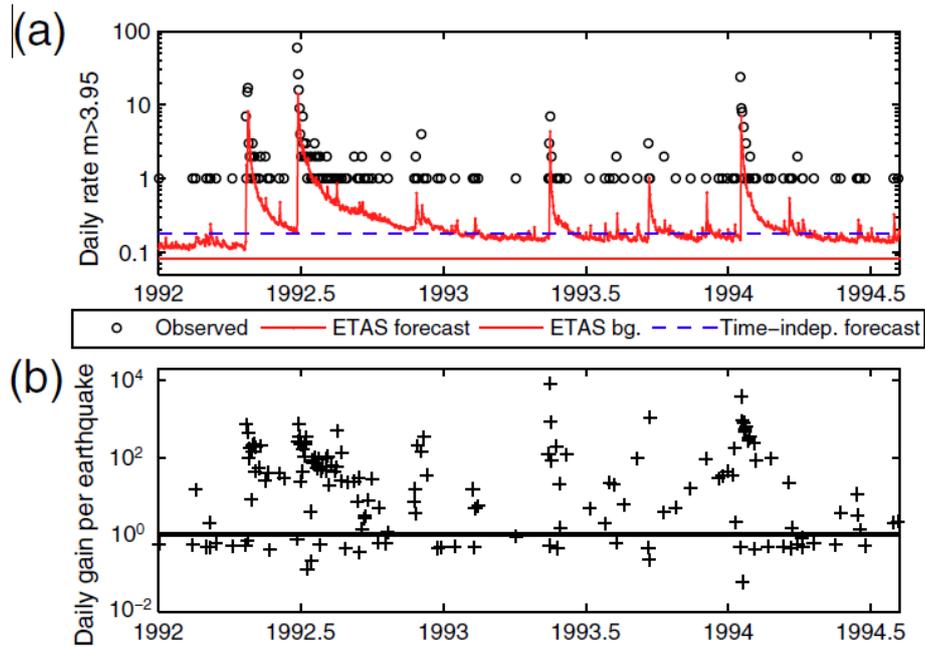


Figure 2-1. (a) Comparison of the number of observed earthquakes per day greater than magnitude 3.95 with the forecasts of the ETAS model during the M6.1 Joshua-Tree, M7.3 Landers and M6.4 Northridge earthquake sequences in California. The dashed blue line represents the time-independent forecast, the solid red line represents the ETAS model forecasts, and the solid horizontal line represents the background rate, describing the seismicity of the region of interest before the experiment starts (learning phase). (b) Probability gains per earthquake using as reference model the long-term model (blue line in (a)). Figure taken from Werner *et al.* (2011).

One advantage in the implementation of ETAS models is the fact that they are catalogue-based and hence do not require any other information about the activated fault source or the regional fault system that are usually not available in real-time. The simplicity of the required input data, all of which is available in real-time, due to the existence of modern permanent seismic networks, makes ETAS the preferred model when it comes to real time implementation.

Temporary networks deployed during evolving earthquake sequences provide scientists with the opportunity for new high-resolution testing of forecasts through the comparison of stress-based and statistical ETAS models. Mancini *et al.* (2019) showed the efficiency of preliminary ETAS models to describe the evolving 2016-18 earthquake cascade in the Central Apennines in short- and long-term time windows (Figure 2-2). The data used corresponds to earthquakes with magnitude $M > 3.0$. The evaluation of the models' performance showed that only stress-based models with a high level of complexity in the representation of the seismic source and the surrounding shallow crust can compete with a standard ETAS model with optimized model parameters.

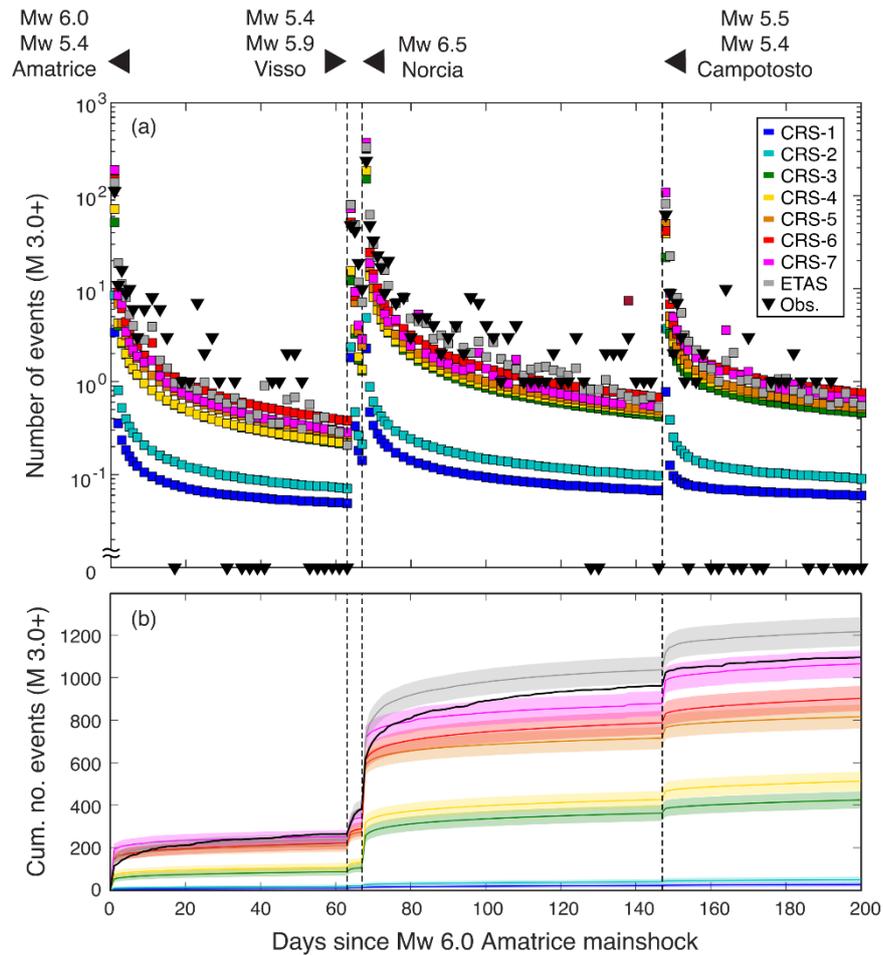


Figure 2-2. Forecast time series for several physical stress based models (CRS) and an ETAS model (light grey). (a) Observed (triangles) and forecasted (squares) incremental number of events with M3+ for 200 days following the Mw=6.0 Amatrice mainshock. We note that for illustration purposes we plot 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. Figure taken from Mancini et al. (2019).

Although ETAS was originally developed for natural seismicity, its application has been extended to studies of induced seismicity. Recent efforts to model earthquakes associated with geothermal exploitation (e.g. Bachmann *et al.*, 2011; Mena *et al.*, 2013) and waste-water injection for oil and gas production (e.g. Llenos and Michael, 2013) showed the effectiveness of this approach to model seismicity patterns during and after injection phases and forecast the observed seismicity in terms of number of events and magnitude distribution.

Llenos and Michael (2013) demonstrated using ETAS models and statistical tests that the earthquake rate change (after injections started) is statistically significant, and both the background rate of independent earthquakes (before their time period of interest) and the aftershock productivity must increase in 2009 (after operations) to explain the observed increase in seismicity in Oklahoma and Arkansas (Figure 2-3). Both regions under study had previously experienced low but steady earthquake rates of approximately 2 events per year. The same authors suggest that for induced earthquakes the inter-event distance was significantly smaller than for natural earthquakes and that this parameter might be a useful criterion to distinguish between before, after, or far from active injection.

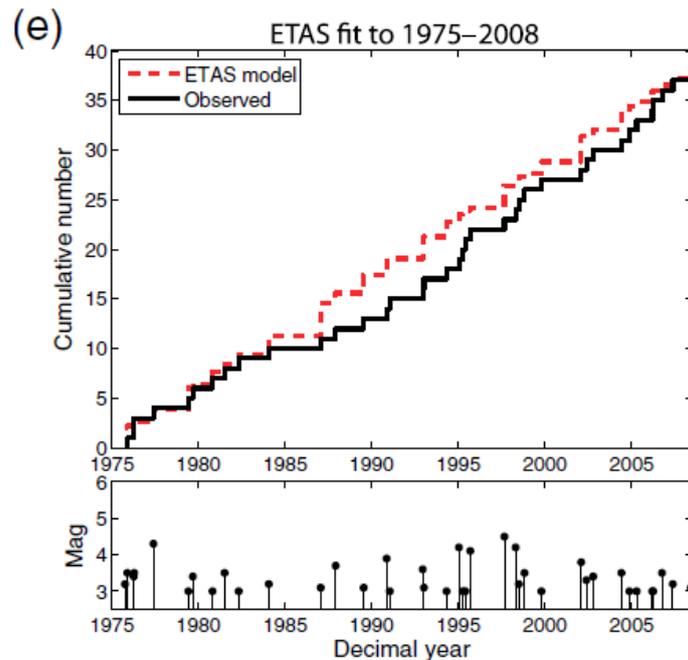


Figure 2-3. Oklahoma ETAS modelling results from Llenos and Michael (2013). The red dashed line represents the ETAS model for the catalogue fit before and during operations. ETAS provides a good fit, visually and quantitatively.

Lei *et al.* (2017) investigated the case of hydraulic fracturing seismicity at the Shangluo shale gas site in the Sichuan Basin, China, an intraplate low seismicity environment (Figure 2-4). In the time period from 1970 to Oct. 2008 there are only 60 earthquakes with magnitude $M > 2.5$ observed. However according to Lei *et al.* (2019) “between the start of shale gas prospecting in 2008 and 2012, an increasing event rate was observed. Based on our survey, we believed that the limited well injections conducted for evaluation purposes might be responsible for the increase. Horizontal drilling began in 2011, while systematic shale gas hydraulic fracturing in those horizontal wells began in 2014. Since Dec. 2014, the earthquake rate increased dramatically in limited areas surrounding the hydraulic fracturing well pads, where (based on interviews during field surveys) local inhabitants reported feeling quakes at an abnormally high frequency.”

The same authors have used ETAS model to describe the seismicity around the Shangluo shale gas site from 2014 onwards (Lei *et al.*, 2017), marking the beginning of the high-quality relocated earthquake catalogue. They have observed that the earthquakes occurred in faults primed by the existing stress field orientations. The modelling results suggest that “injection-induced earthquakes (not including microseismicity from hydraulic fracture openings) demonstrate no significant differences from normal tectonic earthquakes in aspects such as mechanism solution, source process, and seismic b -value “. They find that the most important difference from tectonic activity is the low productivity (productivity=ability to trigger daughter events relative to the parent event magnitude). They conclude that collective work from academic, oil industry and regulatory communities is necessary to understand more about the mechanisms of hydraulic fracturing seismic activity. They note that short-term injections for shale gas hydraulic fracturing are likely to induce $M4-5$ earthquakes in the Sichuan Basin. The authors cite US, UK and Canada examples that recorded HF events have reached magnitudes in the range 2.0 to 4.6.

Induced seismicity cases around the world have progressed our understanding of how to model such transients hazards, which may sometimes cause disruptions to nearby population centres and therefore require some form of mitigation minimise this risk. Implementation of statistical, and depending on the availability of the necessary high-quality data, physical models for earthquake triggering will provide a basis for ongoing risk assessment (Lee *et al.*, 2019). According to this policy report from an international expert group for injection induced hazards, now published in

Science, “Best practice involves a formal process of risk assessment, with input from competent authorities, and the updating of this assessment as knowledge of the potential hazard evolves. Implementation of a comprehensive risk framework should incorporate scenarios of a triggered large earthquake.

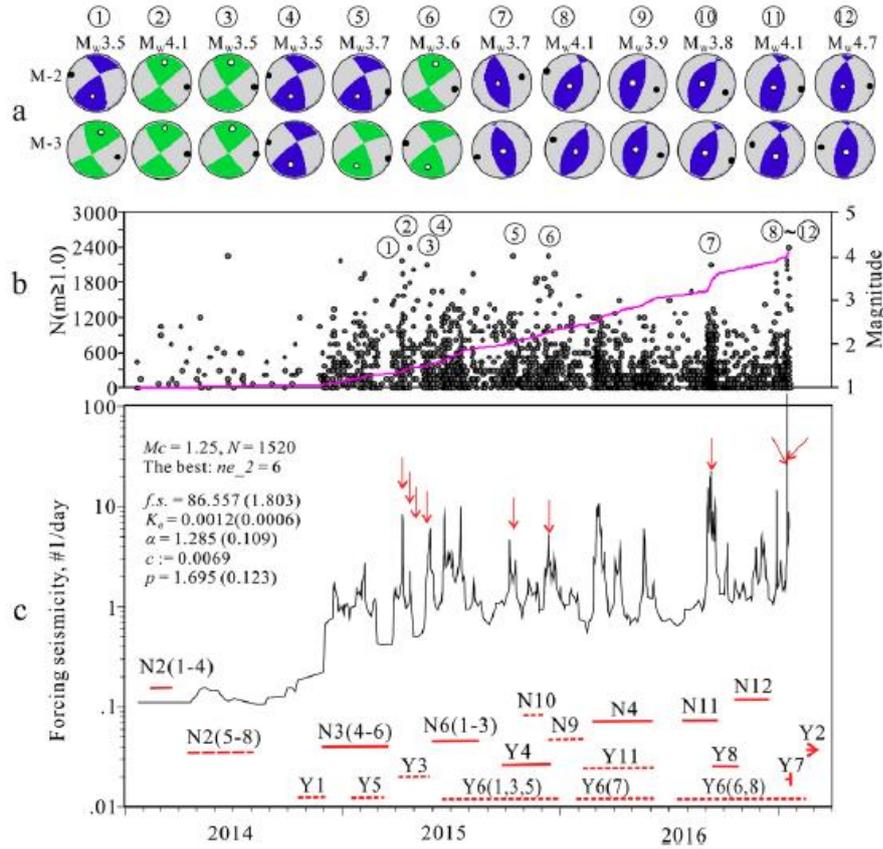


Figure 2-4. ETAS modelling results for event occurrence within the Shangluo shale gas block. Top: rupture parameters from the high resolution relocated earthquake catalogue for the largest earthquakes in the sequence. Middle: Number of observed events Bottom: Expected rate resulting from an ETAS forecast. Figure taken from Lei et al. (2017).

2.2 TECHNICAL DESCRIPTION OF THE ETAS FRAMEWORK

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

$$\lambda(t | H_t) = \mu + \sum_{i: t_i < t} g(t - t_i; M_i), \quad (1)$$

where μ is the background rate, a time-independent 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):

$$g(t; M) = K e^{\alpha(M - M_{cut})} \cdot c^{p-1} (t + c)^{-p} (p - 1), \quad (2)$$

with a normalized temporal distribution as the second term on the right-hand-side, respectively. The parameter K regulates the short-term aftershock productivity by a parent event with magnitude M equal or above a minimum triggering magnitude (M_{cut}), here set to -1 and -1.5; 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) is the modified Omori law (Utsu, 1961) describing the distribution of triggered earthquakes in time in terms of a power law decay with exponent p and a short-term constant c .

In this study, we use a temporal ETAS model, but a spatio-temporal version exists and can be considered in the future. A spatial model would also allow estimates of ground motions but would benefit from well constrained locations obtained via relocation with the double difference technique.

The development of earthquake forecast models includes a validation step where the expected (model) is compared against the observed seismicity (nature). In this performance evaluation, we use well-established statistical metrics introduced by the Collaboratory for the Study of Earthquake Predictability (CSEP – Michael and Werner, 2018). This type of retrospective testing of the predictive skills of earthquake forecasts between statistical and/or physics-based models (e.g. Segou *et al.*, 2013) is a crucial step towards their potential future operational implementation.

We estimate the ETAS parameters by means of the maximum likelihood estimation (MLE) approach, with which we obtain the set of parameters that, given the observations (i.e. a seismicity catalogue with N events), maximize the following log-likelihood function (Zhuang *et al.*, 2002):

$$\log L(\mu, K, c, p, \alpha) = \sum_{i=1}^N \log \lambda(t_i | H_t) - \int_{T_0}^{T_1} \lambda(t) dt, \quad (3)$$

where T_0 and T_1 represent the start and end times for fitting. We use the R package PtProc by Harte (2010) to perform the estimation.

The ETAS model considers the combination of triggering effects from background and evolving seismicity during the learning phase and within each forecast time period for the next forecast window. To improve the ETAS model within the same time interval, and before the next update, we simulate these future events. According to the simulation algorithm (Zhuang and Touati, 2015): (a) the number of future events is Poisson distributed with a mean rate controlled by the productivity law, (b) the occurrence times are sampled from the modified Omori law, (c) magnitudes are drawn from a Gutenberg-Richter distribution with b -value truncated at M_{max} . In tectonic seismicity we take M_{max} consistent with historical seismicity and modern regional strain rates (Rong *et al.*, 2016). For higher generations of triggered events, the simulation process is repeated until the number of potential parent shocks eventually dies out.

2.3 STATISTICAL VALIDATION OF FORECAST MODELS

A number of statistical tests (Schorlemmer *et al.*, 2007; Marzocchi *et al.*, 2012 and references therein) are implemented within the Collaboratory for the Study of Earthquake Predictability (CSEP) in order to efficiently evaluate the spatio-temporal performance of short and long-term earthquake forecast models. We have implemented the modified N-test (Zechar *et al.*, 2010) to compare the total number of observed vs. forecasted earthquakes over a precise time horizon.

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, we use a cumulative mass function F that is constructed from the model simulations:

$$\delta_1 = 1 - F\left((N_{\text{obs}} - 1) \mid N_{\text{fore}}\right), \quad (4)$$

$$\delta_2 = F(N_{obs} | N_{fore}). \quad (5)$$

We then evaluate the N-test by applying a one-sided significance test and we ‘reject’ the forecast if either $\delta_1(t) < \alpha$ or $\delta_2(t) < \alpha$, where $\alpha = 0.025$ is the effective significance value (Zechar et al., 2010). Based on the multiple N-tests over successive forecast periods, 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.

In less formal terms, a “rejected” forecast indicates a potentially significant discrepancy between model and data. Scientific or practical significance of the discrepancy will depend on the context and usage. “Rejected” forecasts may thus still be good enough to be useful.

3 Summary of Operations

The PNR-1z well targets the Bowland shale at a depth of approximately 2,300 m, and runs approximately east-west for 700 m horizontally through the unit. A sliding-sleeve completion method was used, with 41 individual sleeves spaced at intervals of 17.5 m along the well. The hydraulic fracture plan allowed for up to 765 m³ of fluid per sleeve. A “mini-frac” consisting of a few 10s of m³ of fluid was pumped prior to each main stage. The sleeves were numbered from 1 to 41 proceeding from the toe (west) to the heel (east) of the well. A total of 16 sleeves were hydraulically fractured with an additional 18 mini-fracs between 16 October 2018 and 17 December 2018. The sleeves were used in the following order: 1, 2, 3, 12, 13, 14, 18, 22, 30, 31, 32, 37, 38, 39, 40 and 41. The average injected volume for each fracture was 234 m³ and the maximum injected volume was 431 m³. No hydraulic fracturing was carried out between 3 November and 4 December as flow-back from the well took place.

4 Data and Exploratory Analysis

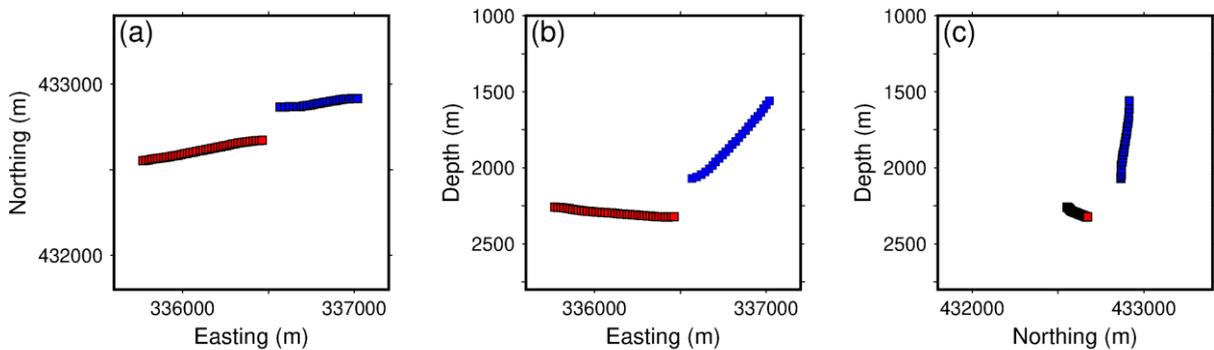


Figure 4-1. Downhole acquisition geometry showing locations of sleeves in well PNR-1z (red squares) and the geophones in PNR-2 (blue squares). (a) plan view, (b) East-West cross-section and (c) North-South cross section.

The microseismicity at PNR-1 was acquired with a downhole geophone array in the adjacent PNR-2 well. This monitoring setup is shown in Figure 4-1. The geophones were located in the heel of the well resulting in approximate path lengths from the sleeves in PNR-1z of between around 200 to 800 m. The geophones recorded continuously from the onset of operations, detecting over 38,000 microseismic events. The event catalogue for this project consisted of origin times, locations and magnitudes of these events as determined by a geophysical processing contractor. No information was available on the method used to locate individual events or the location uncertainties. Typical location uncertainties for downhole microseismic data are around 10-20 m in depth and slightly larger uncertainties in horizontal location. Similarly, although the magnitudes

are reported as moment magnitudes (M_w), no information was available on how this was determined or the magnitude uncertainty. Although a systematic study of event locations is outside the scope of this report we note that the acquisition geometry may result in some bias in location estimates, which may propagate into magnitude estimates.

We encountered two additional issues with the catalogue as provided. Firstly, the precision of the event origin times was limited to 1 second. This lack of precision meant that there were around 1600 events which shared an origin time with at least one other event. Secondly, it was known that the limited dynamic range of the downhole geophones meant that magnitude estimates for events with magnitudes larger than 0.0 M_w were saturated and could not be considered reliable. To address this problem, we found matching events from the catalogue of events detected and located using the dense array surface sensors deployed partly in order to comply with regulatory requirements. This latter catalogue consisted of 172 events with magnitudes given as local magnitude (M_L). We replaced the moment magnitudes for all events in the downhole catalogue with magnitudes greater than 0.0 M_w with the local magnitude estimate for corresponding event in the surface catalogue. Ideally, local magnitudes would have been converted to M_w , however, we find the conversion relationship given in the Cuadrilla Hydraulic Fracture Plan (2019) and based on the relationship derived from the Munafo et al. (2016) does not appear to fit the observed data.

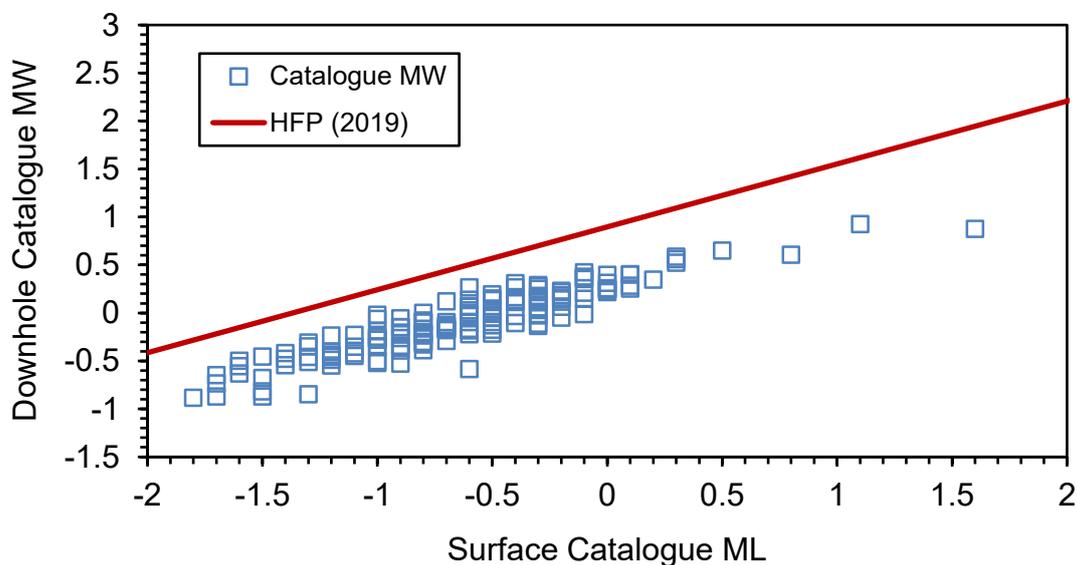


Figure 4-2. Comparison of the M_w estimates from the downhole catalogue with the M_L estimates for matching events in the surface catalogue. The red lines shows the M_L/M_w conversion relationship given the in Cuadrilla HFP (2019)

Figure 4-2 shows a comparison of the M_w estimates from the downhole catalogue with the M_L estimates for matching events in the surface catalogue along with the M_L/M_w conversion relationship given in the Cuadrilla HFP (2019). The latter was derived using moment magnitudes and local magnitudes calculated using surface recordings. We observe a clear difference between the observed magnitudes and the relationship, which appears to be a result of lower moment magnitude values in the downhole catalogue than those used to derive the relationship. Application of this M_L/M_w conversion to the matching events with $M_w > 0.0$ will result in a jump in the magnitude estimates. As a result of this issue, we instead assume an equivalence between M_L and M_w and use the M_L magnitude values for all events with $M_w > 0.0$.

4.1 EVENT BEHAVIOUR IN SPACE AND TIME

Locations for all events in the catalogue are shown in Figure 4-3. Events are coloured by time and move from west to east corresponding to different stages of hydraulic fracturing in the horizontal well PNR-1z. The locations of the events closely correspond to the positions of the sleeves that were hydraulically fractured (coloured squares in Figure 4-3). Event depths are around 2280 m,

but decrease slightly from around 2300 m at the toe of the well to approximately 2250 m closer to the heel.

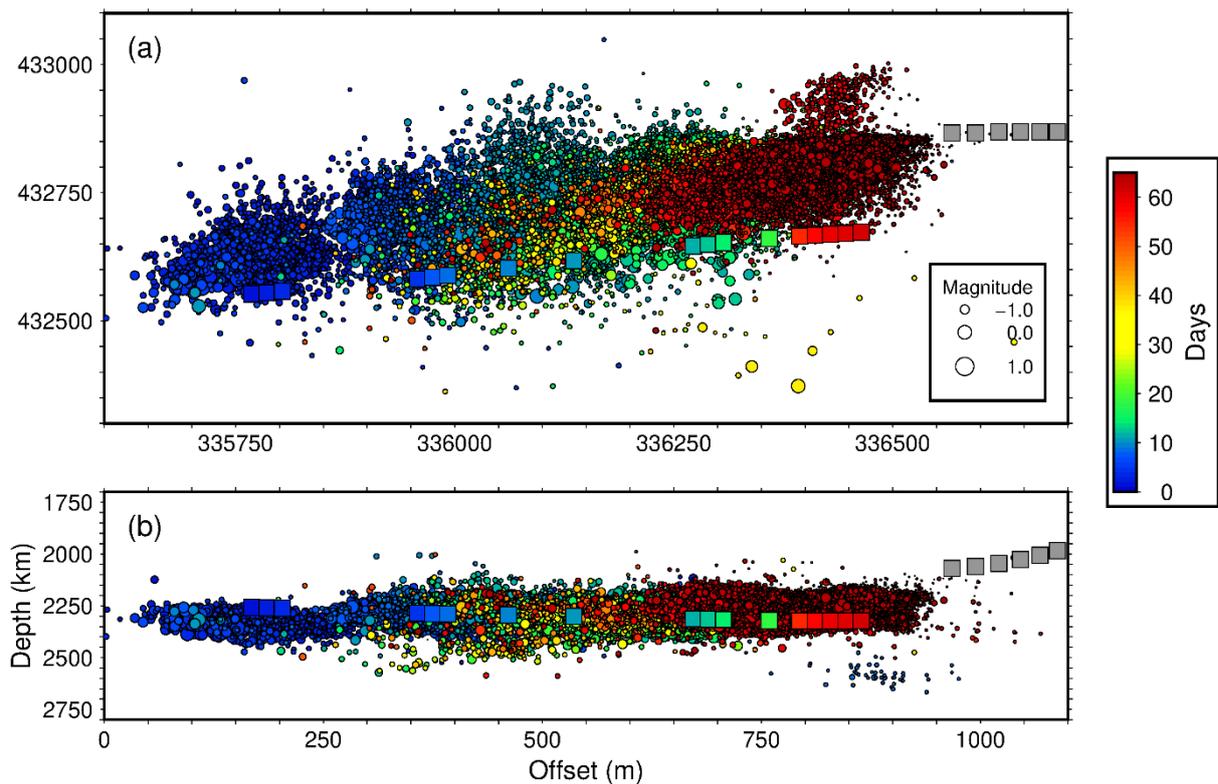


Figure 4-3. (a) Map of all events in the microseismic catalogue. Events are coloured by time and scaled by magnitude. The coloured squares show the locations of the sleeves that were hydraulically fractured. The squares are coloured using the same colour scale as the events. Axes show British National Grid Eastings and Northings. Grey squares show geophone positions. (b) Depth cross-section showing event depths along an east-west profile.

Figure 4-4 shows a histogram of the number of events/hour as a function of time during operations along with the cumulative volume of injected fluid during hydraulic fracturing (blue line). Event rates increase when injection starts and decay rapidly with time after injection stops. There is considerable variation in event rates between individual stages of operations, with some showing relatively high activity rates. For example, sleeves 22, 30, 31 and 32 on 25, 26, 27 and 29 October, all had relatively high levels of detected seismicity. Similarly, sleeves 38, 39 and 40 on 11, 13 and 14 December also have relatively high levels of detected seismicity. These sleeves are all at or closer to the heel (east) end of the horizontal part of the well and all the events with magnitudes greater than 0.5 ML occurred during these hydraulic fracture stages. Conversely, sleeves 1, 2 and 3 on 16, 17 and 18 October at the toe (west) end of the well all have relatively low levels of seismicity, despite similar injected volumes.

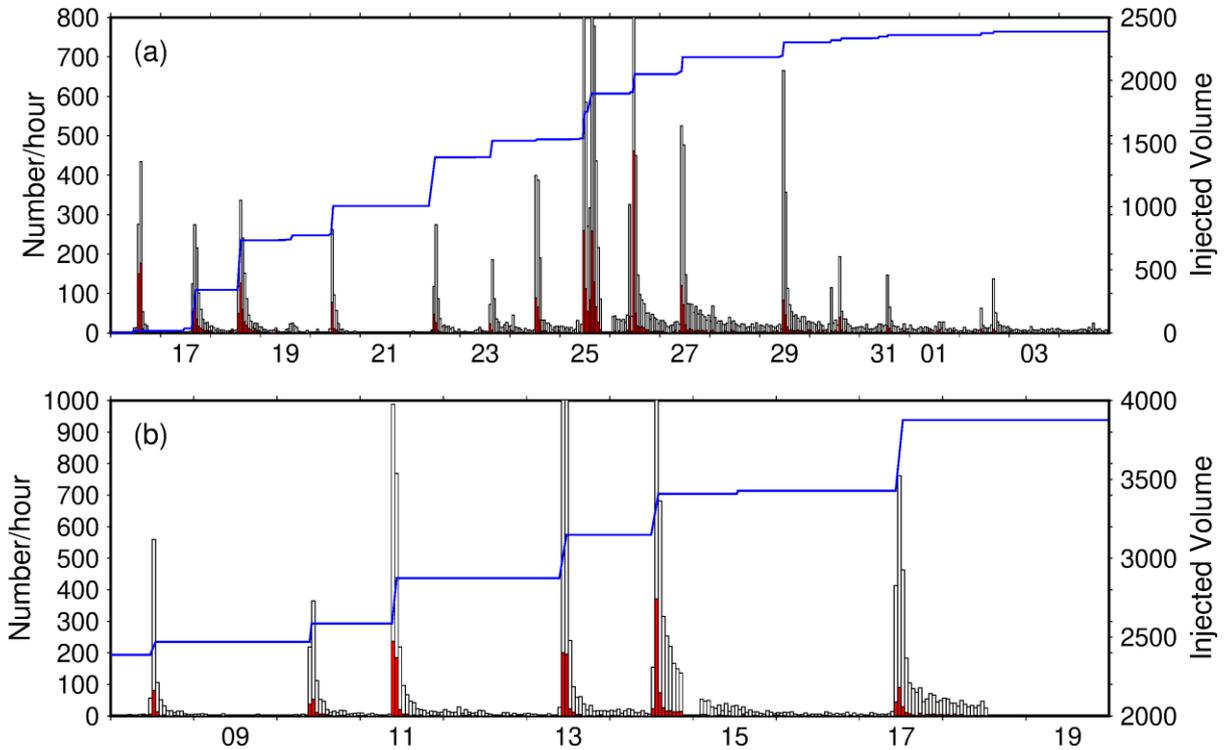


Figure 4-4. Histograms showing the number of events/hour as a function of time during operations. The grey shaded bars show all events, the red bars show events with magnitudes of $-1.5 M_w$ or above. Blue lines show the cumulative volume of injected fluid during hydraulic fracturing operations. (a) shows seismicity in October/November and (b) shows activity in December. No hydraulic fracturing was carried out between 3 November and 4 December.

Similarly, Figure 4-5 shows histograms of the number of events during each stage along with the injected volume of fluid. This allows us to examine the relationship between injected volume and event rates in greater detail. Event rates are observed to increase as soon as injection starts and generally reach a peak towards the end of the period of injection. After injection stops, the rates decay quickly over the subsequent hours. Histograms for those stages that show more complex injection histories, e.g. stages 22 and 30, also show multiple peaks in activity rate that correlate with the observed periods of injection.

4.2 FREQUENCY MAGNITUDE DISTRIBUTION

The relationship between the magnitude and number of earthquakes in a given region and time period generally takes an exponential form that is referred to as the Gutenberg-Richter law (Gutenberg and Richter, 1954), and is commonly expressed as

$$\log_{10} N = a - bM \quad (6)$$

where N is the number of earthquakes above a given magnitude M . The constant a , is a function of the total number of earthquakes in the sample and is known as the earthquake rate. This is often normalised over period of time, such as a year. The constant b gives the proportion of large events to small ones, and is commonly referred to as the b -value. In general, b -values are close to unity for tectonic earthquakes. This means that for each unit increase in magnitude, the number of earthquakes reduces tenfold. However, higher b -values have often been observed for induced seismicity (e.g Bachmann et al, 2011).

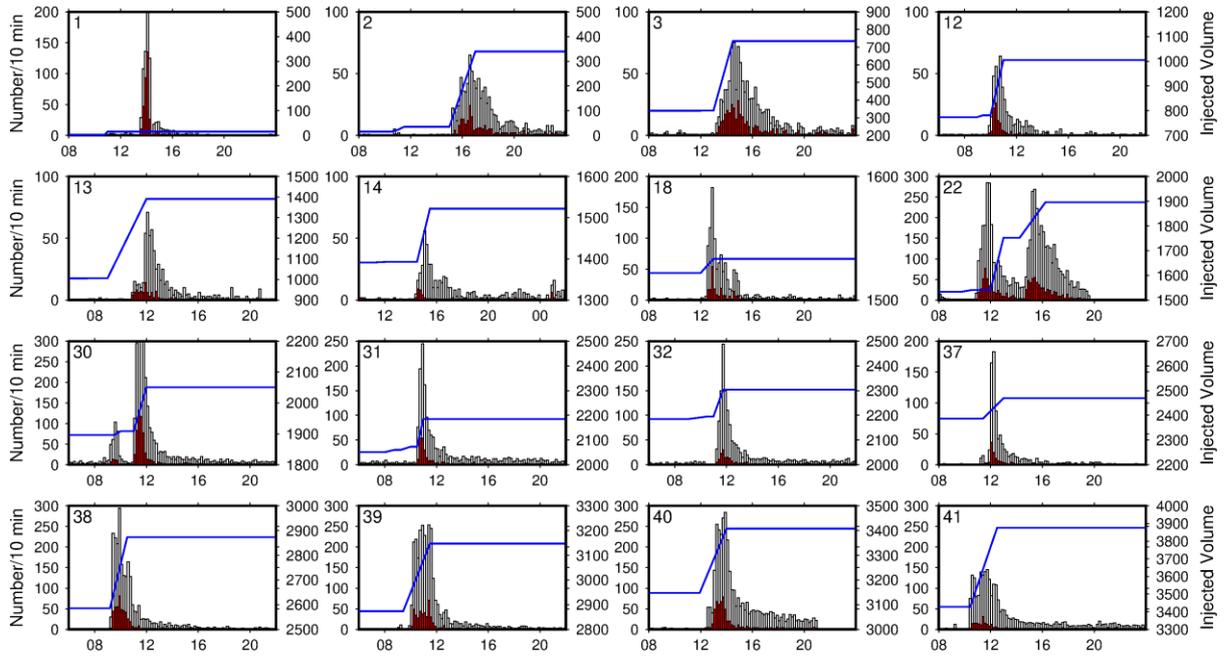


Figure 4-5. Histograms showing the number of events/hours as a function of time during operations. The grey shaded bars show all events, the red bars show events with magnitudes of -1.5 Mw or above. Blue lines show the cumulative volume of injected fluid during hydraulic fracturing operations.

Plotting magnitudes against the logarithm of frequency should therefore give a straight line, where the slope of the line is the b -value and the rate, a , is the value where the line intersects with a given reference magnitude. However, observational data generally shows a roll-off in the number of earthquakes at low magnitudes due to inability to detect small earthquakes. This roll-off in the magnitude-frequency relationship at low magnitudes leads to the concept of a completeness magnitude, M_c , which can be defined as the lowest magnitude at which (approximately) 100% of the earthquakes in a space-time volume are detected (Rydelek and Sacks, 1989). A correct estimate of M_c is crucial since a value too high leads to under-sampling, by discarding usable data, while a value too low leads to erroneous seismicity parameter values and thus to a biased analysis, by using incomplete data.

Following Aki (1965), we calculate the b -value using

$$b = \frac{\log_{10}(e)}{\langle M \rangle - \left(M_c - \frac{\Delta M}{2} \right)} \quad (7)$$

where $\langle M \rangle$ is the mean magnitude of the sample (with $M \geq M_c$) and ΔM is the binning width of the catalogue. Figure 4-6 shows both b -value (a) and activity rate (b) as a function of M_c . This shows that the b -value estimate is strongly dependent on M_c and only starts to stabilize at values of around -1.0, before fluctuating again due to under-sampling at the higher end of the frequency magnitude distribution.

To further assess the completeness of the catalogue we calculated magnitude of completeness, M_c , using the b -value stability method of Cao and Gao (2002). This model is based on the assumption that b -value estimates ascend for cut-off magnitudes less than M_c and remain constant for cut-off magnitudes greater than or equal to M_c , as demonstrated in Figure 4-6. This gives $M_c = -1.0$. We then calculate the b -value for the catalogue using maximum likelihood method of Aki (1965).

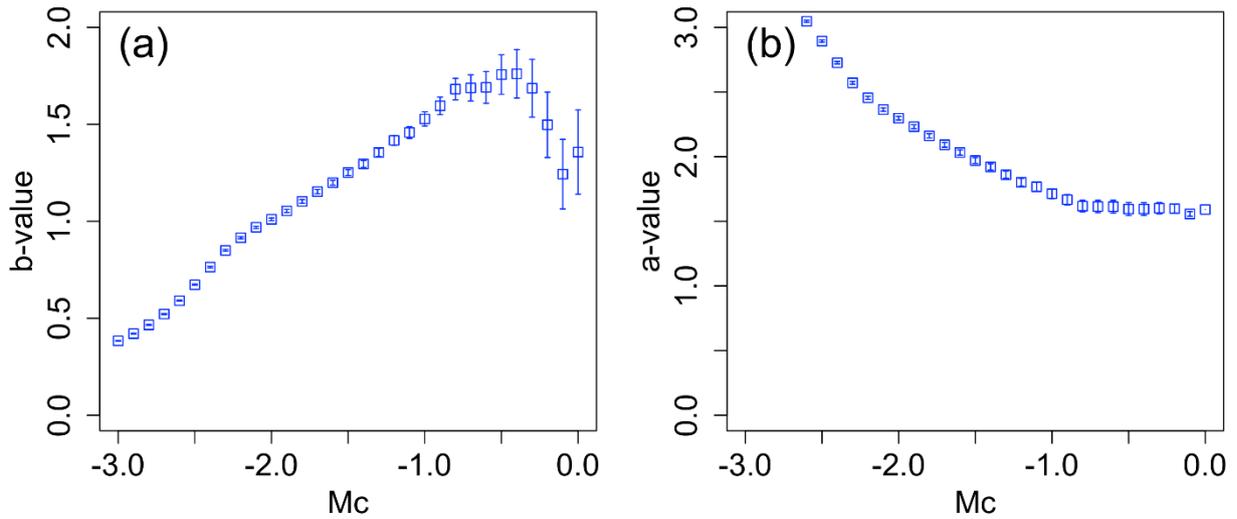


Figure 4-6. b-value (a) and a-values estimates as a function of completeness magnitude M_c for the entire downhole catalogue.

Figure 4-7(a) shows the frequency magnitude distribution for incremental (red) and cumulative data (blue) from the entire down-hole microseismic catalogue, together with the maximum likelihood estimates of the b-value and activity rate for calculated completeness magnitudes of -1.0 (a). This gives a b-value of 1.568 and an activity rate of 1.673. The high b-value for the entire catalogue seems to agree with high b-values calculated for other examples of induced seismicity. However, only around 5 % of the events in the catalogue have magnitudes that are greater than the magnitude of completeness. This greatly reduces the number of events that can be used for the ETAS modelling. As a result, we also calculated a b-value for a slightly less conservative value of completeness magnitude of -1.5, obtaining a b-value of 1.272 and an activity rate of 1.939. Using this value of M_c increase the number of events that can be used for the ETAS forecast to around 20% of the catalogue.

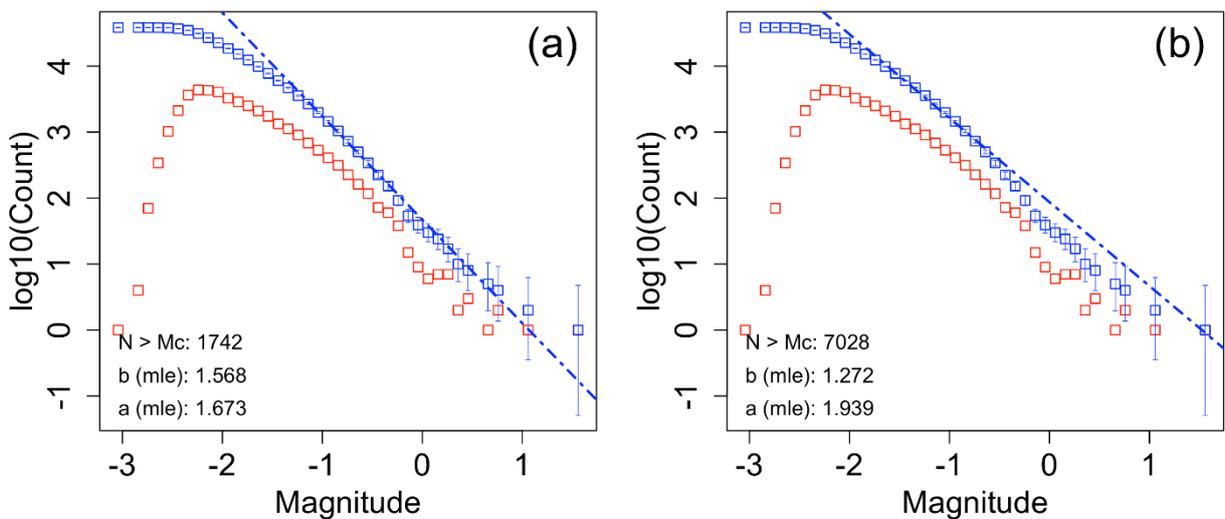


Figure 4-7. Frequency magnitude distributions for incremental (red) and cumulative data (blue). The blue dashed lines show the maximum likelihood estimates of the b-value and activity rate for completeness magnitudes of -1.0 (a) and -1.5 (b).

Finally, we examine how b -values and activity rates vary during individual stages of operations. Figure 4-8 shows frequency magnitude distributions, b -values and activity rates calculated using events that occurred during each stage of operations. The b -values do appear to change from stage to stage, with very high values in the first few stages of operations. Values then appear to decrease as operations progress. However, it is difficult to assign significance to the changes without a more comprehensive analysis of the uncertainties associated with the individual values. As expected, we do see the calculated activity rate increase during those stages where larger number of events were recorded.

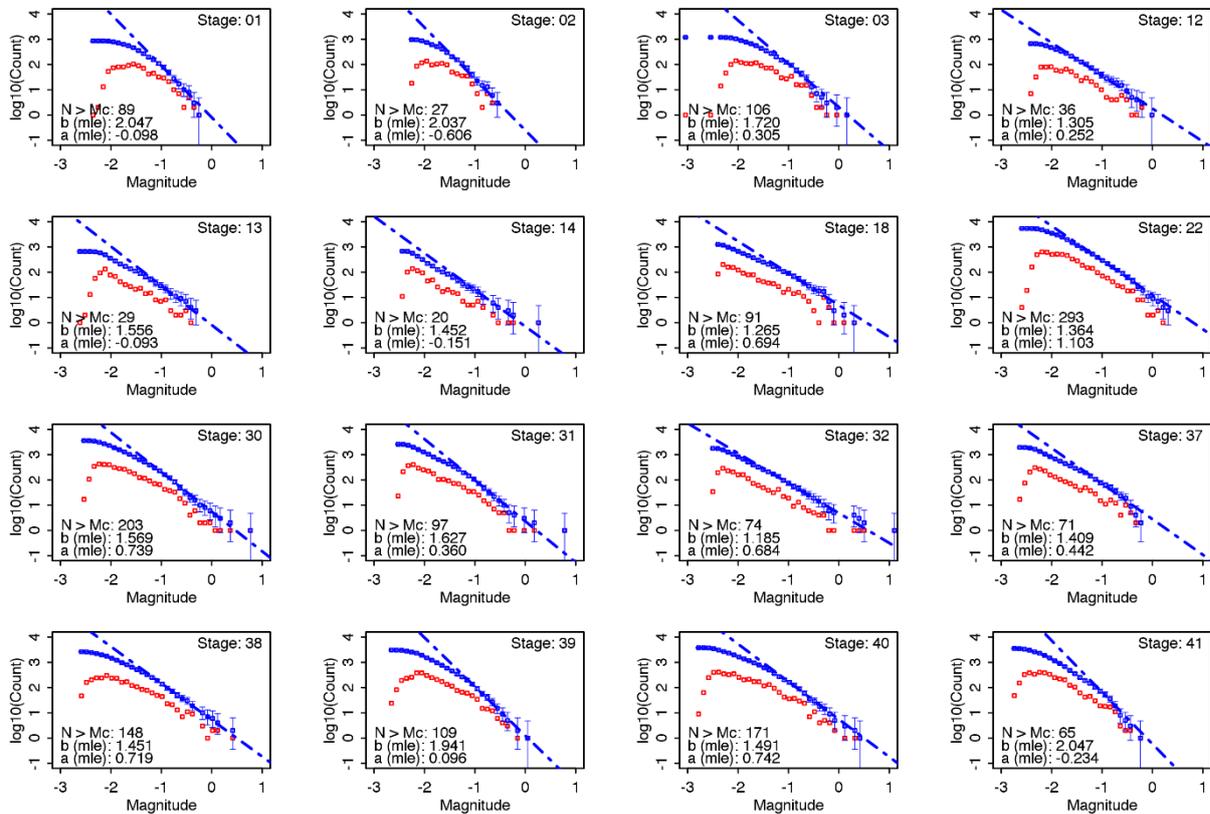


Figure 4-8. Frequency magnitude distributions, b -values and activity rates calculated using events that occurred during each stage of operations. B -values and activity rates were calculated using $M_c = -1.0$.

4.3 DISTRIBUTION OF THE INTER-EVENT TIMES

The distribution of inter-event times provides a useful tool to assess the seismicity patterns and the character of temporal clustering (e.g. Hainzl et al., 2006, Touati et al. 2009, Saichev and Sornette, 2006). Inter-event times refer to the durations between successive events: Given a randomly picked event, how long until the next event?

Features of the distribution can be explained by the underlying seismicity patterns. For example, inter-event times between random and independent (Poissonian) occurrence times follow an exponential distribution. Meanwhile, temporal aftershock clustering according to the Omori law will generate power-law distributed inter-event times with a power-law exponent close to the Omori p -value (e.g. Hainzl et al., 2006). The inter-event times in ETAS model simulations follow a combination of these two in a distribution consisting of three parts: (i) extremely short inter-event times up to a few hours are equally likely (a short-term high plateau corresponding to the Omori law up to the constant c -value), (ii) a mid-term power-law decay due to Omori's power law that identifies temporal clustering, and (iii) a long-term exponential decay that lacks clustering and identifies the independent background rate of spontaneous, unclustered earthquakes. Deviations of the observed distribution from the ETAS model can provide insight into the seismicity patterns and generating mechanisms at PNR-1z.

Figure 4-9 illustrates inter-event time distributions from distinct periods of the operations as well as the distribution of the entire dataset. Several interesting patterns emerge. First, temporal power-law clustering is present. The existence of the temporal clustering provides some justification for the ETAS modelling approach. Panel (d) shows convincing evidence of temporal clustering similar to Omori's law over two orders of magnitude, albeit with a greater p -value than often observed in tectonic seismicity. Interestingly, we observe Omori-type clustering with a p -value closer to tectonic values during the period without stimulation (panel b). In contrast, the seismically productive stage 37 shows steeper decay (panel c).

Second, in contrast to tectonic catalogues, we observe no evidence of the long-term exponential decay in any of the periods. This merely shows that the microseismic catalogue is too short to capture the long-term tectonic background rate, which is not surprising given the low natural rates in the area (Baptie and Luckett, 2019).

Third, the short-term plateau appears to be of different character than the ETAS model prediction and suggests that events in this range are predominantly generated by direct stimulation rather than event-to-event triggering. The difference in the short-term plateau between stages 3 and 37 show that this directly-induced seismicity can vary in intensity and shows characteristic rates. This high-rate unclustered component of the observed seismicity presents challenges for ETAS modelling, which expects short inter-event times to be dominated by triggering. The failure to capture this short-term component contributes to challenges in forecasts of the number of events. This identified model deficiency, however, also suggests potential avenues for future model improvements.

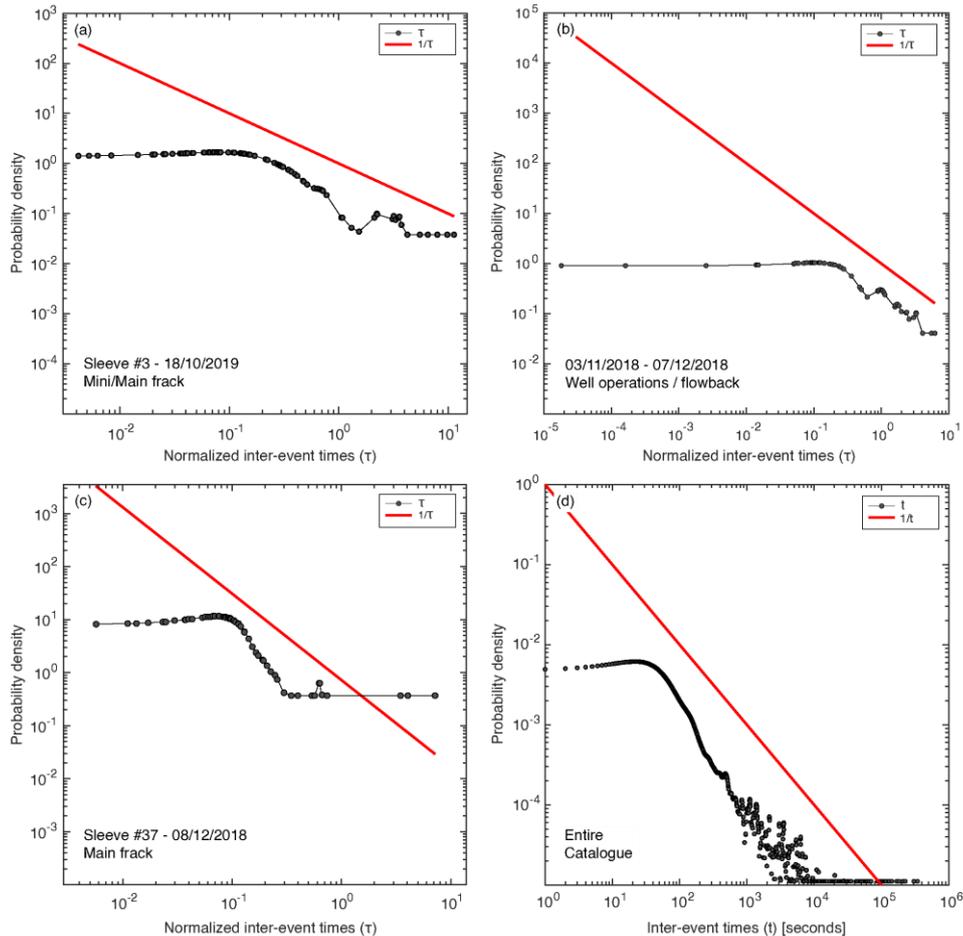


Figure 4-9. Inter-event time distributions of the PNR-1z microseismic catalogue above the completeness magnitude $M_c = -1$. Red line indicates Omori law temporal clustering with p -value of unity. (a) Stage 3, combining a mini and main frack. (b) During the pause in operations between 3 November and 7 December. (c) Stage 37. (d) Entire catalogue. Panels (a)-(c) show inter-event times normalised to the average rate during the period.

5 Model Calibration and Simulation Framework

5.1 MAGNITUDE DISTRIBUTION, B-VALUE AND MAXIMUM MAGNITUDE

Maximum magnitude (M_{\max}) describes the size of the largest possible earthquake in a given region. This is often highly uncertain, particularly in intraplate regions (Holschneider et al., 2011, 2014) where the recurrence interval of large earthquakes is of the order of several hundred to thousands of years due to the low rate of deformation and greatly exceeds the relatively short duration of any catalogue based on historical data.

In the British Isles, the largest earthquake for which a magnitude can be estimated reliably is the magnitude 5.9 M_w Dogger Bank earthquake on 7 June 1931. However, earthquakes of $M_w \geq 6.0$ have occurred in analogous tectonic regions. For example: magnitudes of 8.1 M_w , 7.8 M_w and 8.0 M_w in the 1811-1812 New Madrid earthquake sequence (Johnston and Schweig, 1996); the magnitude 7.3 M_L Charleston earthquake in South Carolina in 1886; the magnitude 7.2 M_L Meeberrie earthquake in Western Australia in 1941 (Johnston et al., 1994), and the magnitude 6.6 \pm 0.5 M_w 1356 Basel earthquake in Switzerland in 1356, which is the largest earthquake ever recorded in north-west Europe (Fäh et al., 2011).

Meletti et al. (2010) follow a statistical approach (see Johnston et al., 1994) to determine the M_{\max} distribution in areas of low seismicity for the 2013 European Seismic Hazard Model (ESHM13) as part of the Seismic Hazard Harmonisation in Europe (SHARE) project (Woessner et al., 2015). We use the same values in this study. The starting point for this method is information from an analysis of the global dataset of seismicity for stable continental regions in Johnston et al. (1994). Then, this is updated with local information available for the study area using a Bayesian approach. In this way, the lack of high seismicity in the study area is compensated for by observations from analogous tectonic regions. The resulting M_{\max} distribution contains magnitudes of 6.5, 6.7, 6.9 and 7.1 with weights of 0.5, 0.2, 0.2 and 0.1 respectively.

To simplify the computation of the numerically intensive ETAS forecasts, we selected an initial single value of the maximum magnitude of 6.5, reflecting the most likely value according to the SHARE approach. We chose this M_{\max} because a more context- and region-specific estimation is difficult and beyond the scope of this project.

5.2 ETAS MODEL CALIBRATION

To obtain the set of parameters that define the conditional intensity function of equations (1) and (2), we fit a temporal ETAS model to the PNR-1z catalogue resulting from the event matching and magnitude assignment described in section 4, for its entire length (15 October to 18 December 2018). We derive the parameters using maximum likelihood estimation (MLE, see equation 3). Following the M_c values reported in section 4.2, we invert two sets of ETAS parameters: the first (ETAS- $1M_{\text{cut}}$) obtained by fixing a minimum triggering magnitude (M_{cut}) of -1.0, the second (ETAS- $1.5M_{\text{cut}}$) by setting $M_{\text{cut}} = -1.5$. For both the inversions, we fix the MLE b-value corresponding to the respective M_{cut} ($b=1.568$ and $b=1.272$ for ETAS- $1M_{\text{cut}}$ and ETAS- $1.5M_{\text{cut}}$, respectively), and we leave all the remaining parameters unconstrained.

We summarize the two sets of ETAS parameters in Table 5-1.

Parameter	ETAS- $1M_{\text{cut}}$	ETAS- $1.5M_{\text{cut}}$	Description
M_{cut}	-1.0 (fixed)	-1.5 (fixed)	Minimum triggering magnitude
μ	1.086	5.367	Background rate
K	0.962	0.946	Productivity parameter
α	-0.442	-0.115	Productivity parameter

c (days)	0.0017	0.0020	Omori-Utsu c-value
p	1.534	1.757	Omori-Utsu p-value
b	1.568 (fixed)	1.272 (fixed)	Gutenberg-Richter b-value
n	0.85	0.91	Branching ratio

Table 5-1. ETAS parameters used for the simulations

Some of the estimated parameters are highly unusual in comparison to typical estimates (e.g., Seif et al, 2017). The productivity exponent α is usually in the range 1.5 to 2.5 and captures the exponential increase of triggering productivity with magnitude. Here, however, the parameter is close to zero and possibly negative, suggesting either little effect of the magnitude on the number of triggered events or possibly a decrease in productivity with magnitude. We interpret this as a result of the challenge to capture the directly induced seismicity, which may not show the usual triggering patterns as also documented in the inter-event time distributions. It is plausible that the model is compensating for the variability of the directly-induced and unclustered rates by eliminating the magnitude-dependent triggering effect and replacing it by strong but magnitude-independent triggering (K is relatively large). In this way, the model could generate high rates after any event independently of its magnitude and attempt to reproduce the high rates observed during injections.

We also observe a higher-than-usual p -value of 1.53 for $M_{\text{cut}}=-1$. A value larger than 1.5 is rare in natural seismicity. The catalogue is much shorter than typical catalogues, suggesting caution in drawing a firm conclusion. However, the high p -value is consistent with the fast decay seen in the inter-event time distributions in particularly active stages, which dominate the statistics. In contrast, the period of relative calm without stimulation shows a more slowly decaying rate with p closer to one. In summary, the high p -value may be a trade-off between the fast decay seen during very active stages and the slower decay seen in quieter stages and in between.

A more detailed uncertainty analysis taking into account parameter covariances is needed before firmer conclusions can be reached, but this was beyond the scope of this project. For example, K and α are known to be anti-correlated (e.g. Seif et al., 2017 and references therein), and the p -value may depend on the length of the catalogue. However, in our experience the observed values are extreme, suggesting that a component of the seismicity is not well captured by the model, while another component (triggered seismicity) is well captured.

5.3 RESIDUAL ANALYSIS (GOODNESS-OF-FIT)

To assess the goodness-of-fit of the model to the data, we performed a residual analysis, which is a standard and powerful tool to assess the ETAS model fit statistically and visually (e.g., Ogata 1988). This is important as parameters can almost always be estimated, but the quality of the fit must be assessed separately.

The basic idea is to compare the continuously changing conditional intensity function (the calibrated model) against the evolving number of observed events. The integrated conditional intensity function (conditional on the steadily accumulating history) should always approximately equal the number of observed events. It can be shown that the fluctuations of a well-calibrated model (i.e. one that fits the data) should be those of a Poisson process of unit rate. Vice versa, significant deviations of the integrated conditional intensity function from a unit Poisson process will highlight discrepancies between model and data. For details, we refer the reader to (Ogata, 1988, Werner, 2008, and references therein).

Figure 5-1 compares the observed seismicity against the calibrated conditional intensity function of ETAS- $1M_{\text{cut}}$. The figure illustrates the rapid variability of the function, which can be thought of as an instantaneous probability of occurrence. The rate shoots up after each event, but decreases

rapidly. During active injection, the contributions from each event to the total rate add up quickly, e.g. on day 10 or 60. The total rate is then orders of magnitude larger than the constant background and is therefore able to anticipate the occurrences of subsequent events in the stage or just after the injection. Isolated events (e.g., on day 50) generate a brief spike but their contributions usually decay too quickly to increase the likelihood of the next event with respect to the background rate.

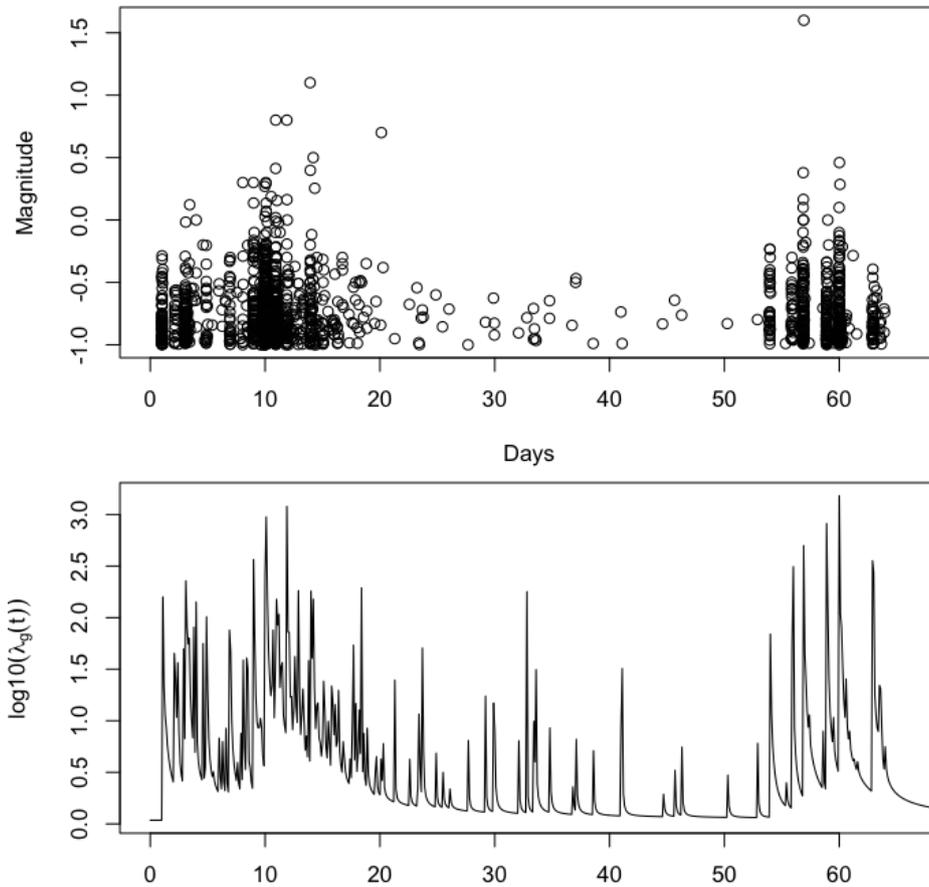


Figure 5-1. Top: Magnitudes above $M_c=-1$ and occurrence times of the PNR-1z microseismicity between August and December 2018. Bottom: calibrated conditional intensity function (instantaneous probability) of ETAS-1 M_{cut} illustrating its variability and the cumulative effect of individual contributions during particularly active periods.

Figure 5-2 shows the residual analysis that compares the expected number of events with the observed number of events. The figure illustrates that the continuous conditional intensity function, when it is updated continuously with the evolving history, is able to match the observed seismicity surprisingly well given the atypical parameter estimates. This is very encouraging. On the other hand, the model is underestimating the rate between event numbers 500 and 1000, approximately between days 9 and 15 of operations.

In practice, forecasts are made at specific times for specific time horizons and will only be updated at regular intervals or at special times (e.g. when larger events happen or during a stage). Such forecasts must be constructed from model simulations and cannot make use, at the time they are issued, of the seismicity during the forecast window. This leads to a loss of predictive skill of the forecasts. In the ETAS framework, this is the only way, however, to generate forecasts that take into account the past history and project over a finite time horizon without instantaneous updating.

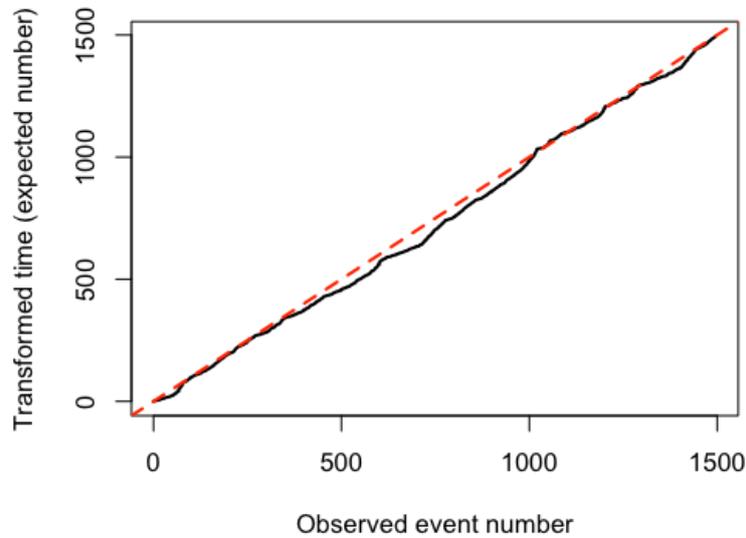


Figure 5-2. Residual analysis showing goodness-of-fit of the calibrated model. The black lines shows the observed versus the expected number of events, obtained by integrating the calibrated conditional intensity function with the continuously updated history. The dashed red line indicates a perfect fit. The goodness-of-fit is good overall, but shows underestimation between event numbers 500 and 1000.

5.4 FORECASTS BASED ON ETAS SIMULATIONS

We fixed the ETAS parameters during the whole forecast horizon, and we used them to simulate a set number of catalogues of synthetic occurrence times and magnitudes. All simulations began at the starting time of the first injection stage at sleeve 1. We tested the performance of two simulation strategies: (a) for ETAS- $1M_{cut}$, simulations updating either every 24 hours or at the beginning of each injection stage (whichever came sooner), and (b) for ETAS- $1M_{cut}$ and ETAS- $1.5M_{cut}$, simulations updating hourly or 15 minutes after the beginning of each injection stage. We included the additional 15 minutes within each stage to allow the model to take into account the early stage-specific activity that could trigger subsequent microseismicity.

For reasons of computational efficiency, we performed 10,000 simulations for the daily forecasts, while we limited the number of simulated catalogues to 1,000 for the hourly forecasts.

6 Performance Evaluation of the Forecasts

In this section, we present the retrospective forecasts in the form of (a) a comparison of observed vs. expected temporal evolution of the seismicity, both as incremental and cumulative number of events with time, and (b) a comparison of observed vs. simulated frequency-magnitude distributions (FMDs). Then, we assess the temporal performance of the forecasts using their rejection ratios (R_{rej}), defined as the percentage of forecast windows in which the observed number did not fall into the 95% predictive model interval. A lower value of R_{rej} corresponds to better predictive skills of the model under test, specifically to simulated ranges that enclose the observations.

6.1 DAILY ETAS-1M_{CUT} MODEL

Figure 6-1a presents the observed and expected number of $M \geq -1$ events per day. We note that ETAS-1M_{cut} projects the onset of the seismicity with a 24 hours delay with respect to the start of our observations, missing the first 77 events due to its relatively low updating frequency (every 24 hours) and the lack of $M \geq -1.0$ parent earthquakes within the first time interval of interest. Following these first few time intervals, the model recovers and estimates expected seismicity rates although it underpredicts the cumulative number by approximately one order of magnitude.

The incremental and cumulative timeseries (Figure 6-1 a-b) reveal that during high-rate periods the model underestimates to the point of being rejected (Figure 6-1c). However, the daily ETAS-1M_{cut} adequately reproduces the temporal decay of earthquakes between subsequent injection stages, when the earthquake rates are low to moderate. During the central part of the experiment, dominated by a 35-day pause of operations period, the seismicity appears mostly governed by event-to-event interactions (November 3rd – December 7th); in this time period the observations always fall within the ETAS-1M_{cut} confidence bounds, hence the forecasts are accepted (green circles in Figure 6-1a). The overall rejection ratio of the model is 19%, i.e. the data will be outside the 95% forecast range in about 1 in 5 forecast periods.

We note that the 95% confidence bounds of ETAS-1M_{cut} always encompass the critical value of zero events, meaning that the model predicts at least a 2.5% probability of no events with magnitude $M \geq -1.0$, including even the most seismically productive days. Such a result reveals the high level of stochasticity of the forecast, which is an intrinsic characteristic of the ETAS framework and reflects a lack of forecast specificity.

6.2 HOURLY ETAS-1M_{CUT} AND ETAS-1.5M_{CUT} MODELS

Figures 6-2 and 6-3 illustrate the performance of the hourly updated ETAS models for $M_{\text{cut}} = -1.0$ (ETAS-1M_{cut}) and $M_{\text{cut}} = -1.5$ (ETAS-1.5M_{cut}), respectively. In the hourly updated model there is only a limited number of new events between successive time windows, therefore we expect a model improvement. In spite of the persistent underestimation of observed rates during high seismicity rate periods, the overall rejection ratios for the ETAS-1M_{cut} and ETAS-1.5M_{cut} are 4% and 8%, respectively (Figures 6-2c and 6-3c). These results underline the importance of frequent model updates in agreement with past research results from natural seismicity (e.g. Segou et al., 2013). Results show that the 1-hour time window implementation is essential for (a) the reduction of the level of initial underestimation of the daily ETAS realizations by a factor larger than 2 and, (b) the rapid convergence of the model towards the observed rates. When we compare the cumulative timeseries of the hourly (Figures 6-2b and 6-3b) vs daily (Figure 6-1b) ETAS realizations, we observe that the hourly forecasts model the seismicity evolution better, especially in the first 3-4 days following the onset of operations. In other words, the rapidly changing seismicity at PNR-1z requires a frequent model update strategy. We do not observe a substantial improvement when including an $M_{\text{cut}} = -1.5$ for the last ETAS implementation. It is possible that completeness issues are influencing the quality of the ETAS-1.5M_{cut} forecasts, but this requires further investigation.

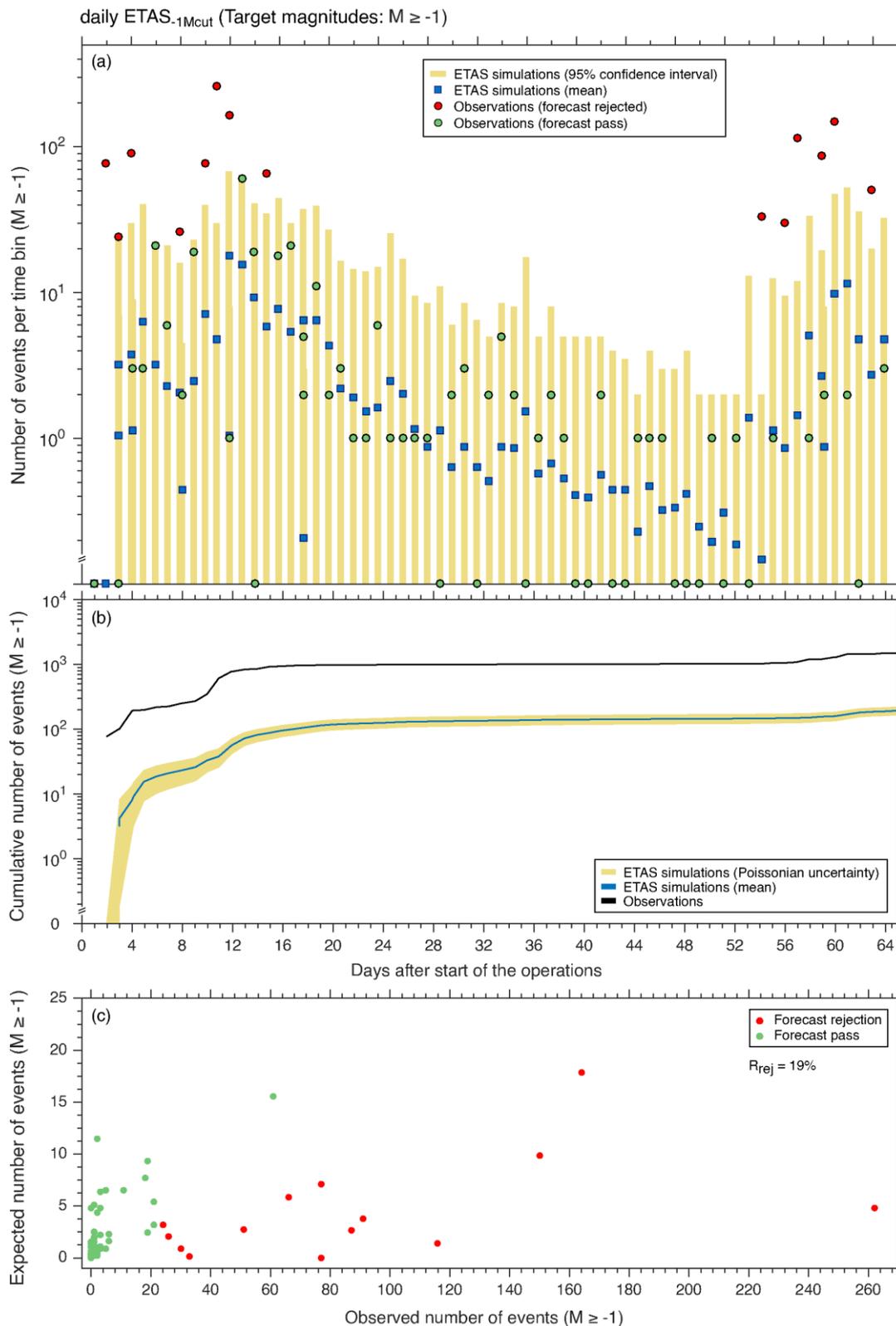


Figure 6-1. Forecast time series and observed vs. expected number of events for the daily updated ETAS_{-1Mcut} model. (a) Observed (circles) and mean forecasted (squares) incremental number of $M \geq -1$ events for the 64 days of operations. The yellow bars represent the 95% confidence bounds of the ETAS forecast. The colour of the circles indicates whether the observed number of events per daily time bin falls within (green) or outside (red) the ETAS confidence bounds. (b) Observed (black line) and mean expected (blue line) cumulative seismicity with shaded areas representing Poissonian uncertainties. (c) Observed vs. expected number of events; red circles indicate forecast rejection, while green ones indicate that the model adequately forecasts the number of events.

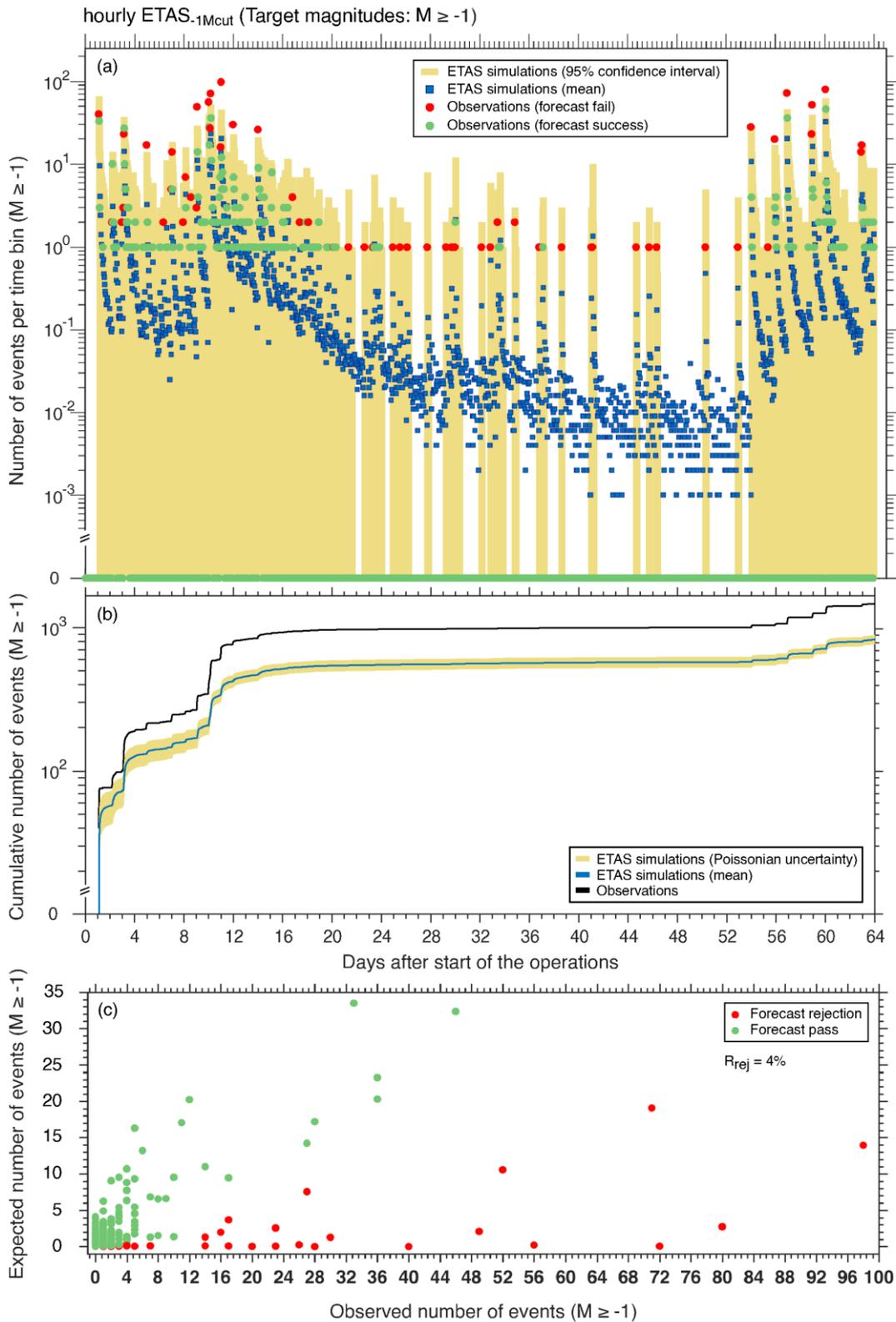


Figure 6-2. Forecast time series and observed vs. expected number of events for the hourly updated ETAS- $1M_{cut}$ model. (a) Observed (circles) and mean forecasted (squares) incremental number of $M \geq -1$ events for the 64 days of operations. The yellow bars represent the 95% confidence bounds of the ETAS forecast. The colour of the circles indicate whether the observed number of events per daily time bin falls within (green) or outside (red) the ETAS confidence bounds. (b) Observed (black line) and mean expected (blue line) cumulative seismicity with shaded areas representing Poissonian uncertainties. (c) Observed vs. expected number of events; red circles indicate forecast rejection, while green ones indicate that the model adequately forecasts the number of events.

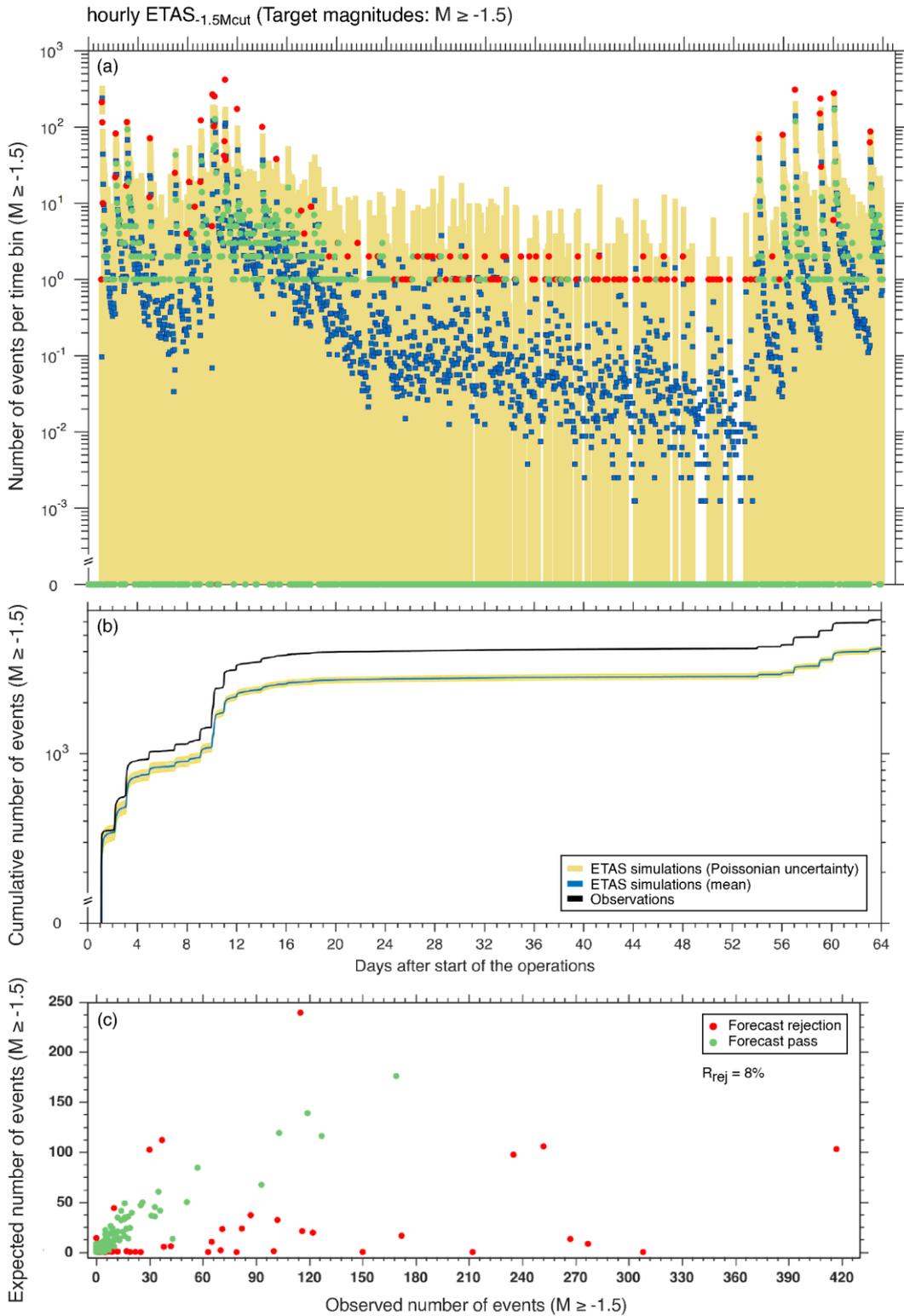


Figure 6-3. Forecast time series and observed vs. expected number of events for the hourly updated ETAS- $1.5M_{cut}$ model. (a) Observed (circles) and mean forecasted (squares) incremental number of $M \geq -1$ events for the 64 days of operations. The yellow bars represent the 95% confidence bounds of the ETAS forecast. The colour of the circles indicate whether the observed number of events per daily time bin falls within (green) or outside (red) the ETAS confidence bounds. (b) Observed (black line) and mean expected (blue line) cumulative seismicity with shaded areas representing Poissonian uncertainties. (c) Observed vs. expected number of events; red circles indicate forecast rejection, while green ones indicate that the model adequately forecasts the number of events.

6.3 COMPARISON OF FREQUENCY MAGNITUDE DISTRIBUTIONS

Here we present the frequency-magnitude distributions (FMDs) resulting from the ETAS model simulations under the daily/hourly ETAS- $1M_{\text{cut}}$ and the hourly ETAS- $1.5M_{\text{cut}}$ implementations. The daily updated model underestimates the observed FMD by an order of magnitude. The results from both hourly updated ETAS models reproduce the observed FMD more closely. Although the model that best fits the observed FMD (ETAS- $1.5M_{\text{cut}}$) generates seismicity up to $M=2.0$ in 95% of simulations, events above $M=2.0$ are rare but can occur. For example, 1 in 1,000 simulations generates a magnitude 3. For comparison, the daily updated ETAS- $1M_{\text{cut}}$ predicts no occurrences above $M=1.2$ within its 95% simulation range, making the observed December $M=1.6$ unlikely if the forecasts were correct. Increasing the number of simulations would continue the downward trend of the mean number in agreement with the log-linear Gutenberg-Richter decay. On the basis of the comparison of the simulated and observed FMDs, we find no evidence to reject the maximum magnitude set from historical data for the UK ($M=6.5$). (However, a range of other maximum magnitudes – both lower and higher – would also be consistent with the observed magnitude distribution.)

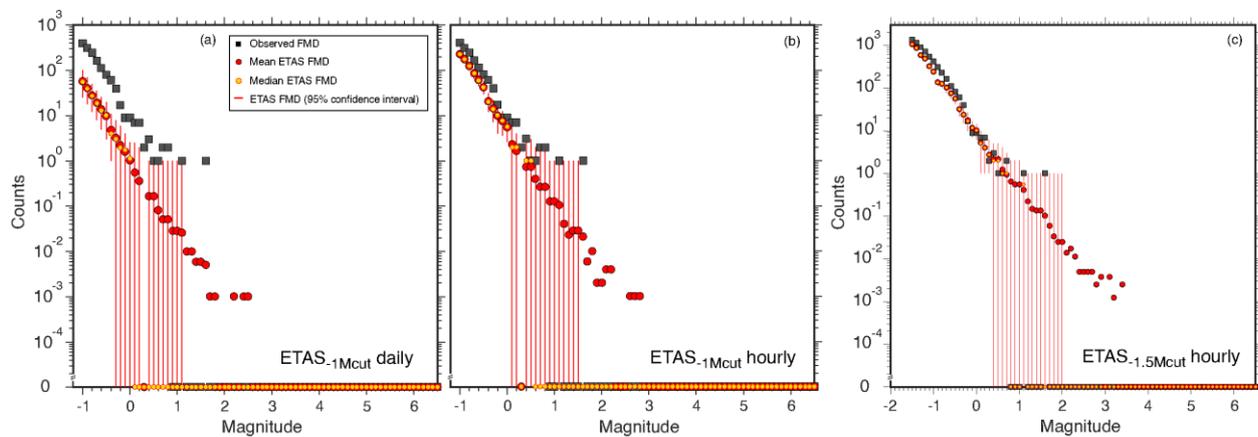


Figure 6-4. Observed vs. simulated frequency-magnitude distributions for the daily and hourly updated ETAS- $1M_{\text{cut}}$ and hourly updated ETAS- $1.5M_{\text{cut}}$ models.

7 Conclusions

The rich microseismicity catalogue resulting from the combined surface and downhole monitoring of the Preston New Road hydraulic fracturing operations offers a unique opportunity to improve our scientific understanding of induced seismicity in low deformation regions. The estimated magnitude of completeness of the catalogue ($M_c=-1.0$) is three orders of magnitude lower than standard earthquake catalogues from networks in high seismic hazard countries. This provides a unique window into the generation and triggering processes down to very small magnitudes.

The effects of magnitude uncertainties, especially for the small magnitude ranges, can cause bias in the estimation of the Gutenberg-Richter frequency-magnitude distribution. This can be corrected when standard errors of magnitude estimates are known (Rhoades and Dowrick, 2000). The comparison between the moment magnitude (M_w) of the PNR downhole microseismic catalogue and the local magnitude (M_L) of the surface catalogue confirms a discrepancy that was previously reported by Kendall et al. (2019). This discrepancy cannot be reconciled using existing M_w to M_L conversions equations such as those developed in the UK for similar magnitude-distance ranges (e.g., Butcher et al., 2017). The lack of a coherent measure of size of a seismic event has a number of important implications. The estimation of the magnitude distribution's b -value and thus the extrapolation to the likelihoods of larger magnitudes will be more uncertain than usual. In

addition, models that forecast seismicity using catalogues as input will also be affected adversely. We recommend additional work to solve this issue.

The calibration of the statistical seismicity model requires a magnitude distribution and five parameters that can be estimated directly from the PNR-1z microseismicity catalogue. We found a good fit of the Gutenberg-Richter (exponential) distribution to the data with a b -value of about 1.5. Formal uncertainties do not express potential systematic bias due to the two magnitude scales. It is conceivable, though unlikely, that the b -value of consistently calibrated magnitudes is closer to the typical tectonic value near unity. We chose a precautionary maximum magnitude ($M_{\max}=6.5$) in keeping with the UK's national seismic hazard map; a more in-depth review of M_{\max} in this context was beyond the scope of this project.

After model calibration, we performed retrospective forecasting experiments, in which we provided the catalogue as input to the model and simulated seismicity repeatedly (at least 1,000 times) for each successive time period over the entire period of operations of PNR-1z. We performed multiple experiments to investigate the effect of two different magnitude thresholds ($M_{\text{cut}}=-1$ and $M_{\text{cut}}=-1.5$) and the updating frequency (daily and at the beginning of each stage; hourly and 15 minutes after beginning of each stage). The forecasts consist of the full statistical distribution of the number of events over each forecast period along with the simulated magnitude distribution. These distributions can be simplified to means, medians and 95% predictive intervals.

The aggregate frequency-magnitude distribution of the simulated seismicity matches the observed microseismicity relatively well. According to the model, the expected occurrence of events with magnitude between 3.0-6.5 Mw is statistically very rare: only 1 out of 1,000 simulations produced seismic events of such magnitude (a magnitude 3 event). In other words, a Gutenberg-Richter distribution with b -value of 1.5 and a tectonic maximum magnitude of 6.5 is consistent with the data, even if only one $M_{L1.5}$ was observed. By the same arguments, a broad range of other maximum magnitudes – both lower and higher than $M_{6.5}$ – are also consistent with the data. The present approach provides little constraint on this critical parameter.

The performance of the model forecasts of both the total number of events during operations and during the successive forecast periods was mixed, with some clear successes and some failures. First, we observed that forecasts that were more frequently updated performed better, because they used information about the early seismic response to injections and could adapt. This is in keeping with operational statistical models of natural seismicity worldwide (e.g. Marzocchi et al., 2016) where ETAS models may provide independent quantifiable outputs that aid expert advice for decision-making purposes during earthquake crises.

Second, the model frequently anticipated the number of seismic events in periods of low to moderate seismicity, but systematically underpredicted when the rate of events during stages was particularly high. The model also underestimated the total number of events over the period of operations. The best performing model, the hourly-ETAS, is rejected in 4% of the time intervals within the entire operational period with the majority of rejection instances corresponding to high observed rates. Nonetheless, the model captured many features of the induced seismicity such as clustering, often tracking the evolution comparatively well. Therefore, the present statistical model may present a useful forecasting tool, especially after suitable modification.

We compared the PNR-1z seismicity patterns with those the model was intended to capture. The microseismicity shows clear evidence of temporal clustering, which is often ascribed to triggering and event-to-event interaction. The model captured this feature well, and it seems to be more prominent during quieter stages, between stages and during the pause in operations in October 2018. We suspect that during these times the dominant seismicity generating mechanism was event-to-event triggering. This describes a critically-stressed faulting process in agreement with the findings of Townend and Zoback (2000). The latter authors, after analysing deep drilling and induced seismicity experiments at several locations worldwide, conclude that the crust remains always in a critical state requiring only small system perturbations to drive fault stability over the cusp of failure even in regions of low long-term intraplate strain rates.

On the other hand, some stages generated high rates of events that were less obviously clustered. This suggests a different dominant mechanism in which events occur independently of each other. A likely physical mechanism is the pore pressure change associated with injections that reduces the effective normal stress and can lead to seismic slip. While the model expects one constant rate of spontaneous events, it currently struggles to model the variability of the high rates and their temporary nature. This opens a promising path for future model development and improvement by modulating the model's background rate with operational parameters.

Further research is critical to improving the predictive power of the model. Testing the model out-of-sample on a new, independent dataset such as the PNR-2 dataset would provide a better evaluation of its skill than testing it on the dataset that it was calibrated on. In addition, embedding several physical mechanisms of seismicity linked with operational parameters is a promising scientific path towards improved operational probabilistic forecasts.

References

British Geological Survey holds most of the references listed below, and copies may be obtained via the library service subject to copyright legislation (contact libuser@bgs.ac.uk for details). The library catalogue is available at: <https://envirolib.apps.nerc.ac.uk/olibcgi>.

Aki, K., (1965). Maximum likelihood estimate of b in the formula $\log N=a-bM$ and its confidence limits, *Bull. Earthq. Res. Inst., Univ. Tokyo*, 43, 237–239.

Bachmann, C., Wiemer, S., Woessner, J., and 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.

Baptie, B. and 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, 9-10 September 2019, Greenwich, London.

Butcher, A., Luckett, R., Verdon, J.P, Kendall, J.M., Baptie, B. and Wookey, J. (2017). Local magnitude discrepancies for near-event receivers: Implications for the UK traffic-light scheme, *Bulletin of the Seismological Society of America* 107 (2), 532-541

Cao, A.M., and Gao, S.S. (2002). Temporal variations of seismic b -values beneath northeastern Japan island arc, *Geophys. Res. Lett.*, 29, doi:10.1029/2001GL013775.

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

Fäh, D, Giardini, D, Kästli, P, Deichmann, N, Gisler, M, Schwarz-Zanetti, G, Alvarez-Rubio, S, Sellami, S, Edwards, B, Allmann, B, Bethmann, F, Wössner, J, Gassner-Stamm, G, Fritsche, S and Eberhard, D. (2011). ECOS-09 Earthquake Catalogue of Switzerland Release 2011 Report and Database. Report SED/RISK/R/001/20110417.

Gutenberg, R. and Richter, C.F. (1954). Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, 34, 185-188.

Hainzl, S., Scherbaum, F., and Beauval, C. (2006). Estimating background activity based on interevent-time distribution, *Bull. Seis. Soc. Am.*, 96(1), 313-320.

Harte, D. (2010). PtProcess: An R package for modelling marked point processes indexed by time, *J. Stat. Softw.*, 35(8), 1-32.

- Holschneider, M, Zöller, G, and Hainzl, S. (2011). Estimation of the maximum possible magnitude in the framework of a doubly truncated Gutenberg-Richter model, *Bull. Seis. Soc. Am.*, 101(4), 1649–1659.
- Holschneider, M, Zöller, G, Clements, R, and Schorlemmer, D. (2014). Can we test for the maximum possible earthquake magnitude?, *J. Geophys. Res. Solid Earth*, 119, 2019–2028.
- Jordan, T. H., Chen, Y. T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., and Zschau, J. (2011). Operational earthquake forecasting. State of knowledge and guidelines for utilization. *Annals of Geophysics*, 54(4).
- Johnston, AC, Coppersmith, KJ, Kanter, LR, and Cornell, CA. (1994). The earthquakes of stable continental regions, EPRI Report, Electric Power Research Institute, Palo Alto, California.
- Johnston, AC, and Schweig, ES. (1996). The enigma of the New Madrid earthquakes of 1811-1812, *Ann. Rev. Earth Planet. Sci.*, 24, 339-384.
- Kendall, J.M., Butcher, A., Stork, A.L., Verdon, J.P., Luckett, R. and Baptie, B.J. (2019). How big is a small earthquake? Challenges in determining microseismic magnitudes, *First Break*, 37, 2, 51-56.
- 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., and Langenbruch, C. (2019). Managing injection-induced seismic risks, *Science*, 364 (6442), 730–32.
- Lei X., Huang D., Su J, Jiang G., Wang X., Wang H., Guo X. and Fu H. (2017). Fault reactivation and earthquakes with magnitudes of up to Mw4.7 induced by shale-gas hydraulic fracturing in Sichuan Basin, China. *Scientific Reports*, 7, 7971
- Lei X., Wang Z. and 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.
- Llenos, A., and Michael, A. (2013). Modelling earthquake rate changes in Oklahoma and Arkansas: possible signatures of induced seismicity, *Bull. Seismol. Soc. Am*, 103(5), 2850-2861.
- 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. Geophys. Res. Solid Earth*, 124. <https://doi.org/10.1029/2019JB017874>
- Marzocchi, W., Murru, M., Lombardi, A. M., Falcone, G., and 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>.
- Marzocchi, W., Sandri, L., Heuret, A., and Funicello, F. (2016). Where giant earthquakes may come. *J. Geophys. Res. Solid Earth*, 121, 7322– 7336.
- Meletti, C, d’Amico, V, and Martinelli, F, (2010). Homogeneous determination of maximum magnitude. Seismic Hazard Harmonization in Europe (SHARE) Deliverable 3.3, Rome, Italy, pp.20 <http://hdl.handle.net/2122/6530>
- Mena, B., Wiemer, S. and Bachmann, C. (2013). Building robust models to forecast the induced seismicity related to geothermal reservoir enhancement, *Bull. Seismol. Soc. Am*, 103(1), 383-393.
- Michael, A. J. and 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>.
- Munafò, I., Malagnini, L. and Chiaraluce, L. (2016). On the Relationship between Mw and ML for Small Earthquakes. *Bulletin of the Seismological Society of America*, 106. 10.1785/0120160130.

- Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes, *J. Am. Stat. Assoc.*, 83(401), 9–27.
- 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>.
- Rhoades, D. A., and Dowrick, D.J. (2000). Effects of magnitude uncertainties on seismic hazard estimates. Proc. 12th World Conference on Earthquake Engineering, 30 Jan-4 Feb 2000, Wellington, NZ. Paper 1179.
- Rong, Y., Bird, P. and 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>.
- Rydelek, P. A., and Sacks, I. S. (1989). Testing the completeness of earthquake catalogs and the hypothesis of self-similarity, *Nature*, 337, 251–253.
- Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., and 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>.
- Saichev, A., and Sornette, D. (2007). Theory of earthquake recurrence times. *J. Geophys. Res.*, 112, B04313.
- Segou, M., Parsons, T. and 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>.
- Segou, M. and Parsons, T. (2016), Impending earthquakes beneath Katmandu and the Himalayan front after the 25 April 2015 M=7.8 Nepal mainshock. *Seismol. Res. Lett.*, 87, 816-825.
- Seif, S., Mignan, A., Zechar, J. D., Werner, M. J. and 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>.
- Touati, S., Naylor, M. and Main, I. G. (2009). Origin and nonuniversality of the earthquake interevent time distribution. *Physical Review Letters*, 102(16), 168501.
- Townend, J. and Zoback, M.D. (2000). How faulting keeps the crust strong, *Geology*, 28, 5, 399-402.
- Utsu, T. (1961). A statistical study on the occurrence of aftershocks. *Geophys. Mag.*, 30, 521-605.
- Van der Elst, N J., Page, M.T., Weiser, D.A., Goebel, T.H.W. and Hosseini, S.M. (2016), Induced earthquake magnitudes are as large as (statistically) expected. *J. Geophys. Res. Solid Earth*, 121, 4575– 4590.
- Werner, M. J. (2008). On the fluctuations of seismicity and uncertainties in earthquake catalogs: Implications and methods for hypothesis testing. PhD Thesis. University of California, Los Angeles.
- Werner, M.J., Helmstetter, A., Jackson, D.D. and 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>.
- Woessner, J, Danciu, L, Giardini, D., Crowley, H, Cotton, F, Grünthal, G, Valensise, G, Arvidsson, R, Basili, R, Betül Demircioglu, M, Hiemer, S, Meletti, C, Musson, Rwm, Rovida, An, Sesetyan, K, Stucchi, and the Seismic Hazard Harmonization in Europe (SHARE) Consortium (2015). The 2013 European Seismic hazard model: key components and results. *Bull. Earth. Eng.*, doi.org/10.1007/s10518-015-9795-1.

Zechar, J. D., Gerstenberger, M. C., and 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>.

Zhuang, J., Ogata, Y., and 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>.

Zhuang, J., and Touati, S. (2015). Stochastic simulation of earthquake catalogs, Community Online Resource for Statistical Seismicity Analysis. <https://doi.org/10.5078/corssa-43806322>.