An empirically constrained forecasting strategy for induced earthquake magnitudes using extreme value theory

James P. Verdon[1*], Leo Eisner[2]

1. School of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, United Kingdom.

2. Seismik s.r.o., Kubišova 1265/8 Praha, Czech Republic.

* Corresponding Author. Email: James.Verdon@bristol.ac.uk, Tel: 0044 117 331 5135. Orcid: 0000-0002-8410-2703

Keywords

Induced seismicity; earthquake forecasting

Acknowledgements

James Verdon's contribution to this study was funded by the Natural Environment Research Council (NERC) under the SeisGreen Project (Grant No. NE/W009293/1)

Declaration of Competing Interests

Both authors have acted and continue to act as independent consultants for a variety of organisations including hydrocarbon operating companies and governmental organisations on issues pertaining to induced seismicity. None of these organisations had any input into the conception, development, analysis or conclusions of this study.

ABSTRACT

Induced seismicity magnitude models seek to forecast upcoming magnitudes of induced earthquakes during the operation of subsurface industries such as hydraulic fracturing, geothermal stimulation, wastewater disposal, and carbon capture and storage. Accurate forecasting models could guide operational decision-making in real time, for example operations could be reduced or paused if forecast models indicate that magnitudes may exceed acceptable levels. Robust and transparent testing of forecasting models is required if they are to be adopted by operators and regulators of such industries. We develop and test a suite of models based on extreme value estimators to forecast the magnitudes of upcoming induced seismic events based on observed seismicity. We apply these models to multiple induced seismicity cases from wastewater disposal in Oklahoma and in western Texas, as well as other cases of seismicity caused by subsurface fluid injection in North America, Europe, and China. In total, our testing dataset consists of more than 80 individual sequences of induced seismicity. We find that all the models produce strong correlation between observed and modelled magnitudes, indicating that the forecasting provides useful information about upcoming magnitudes. However, some models are found to systematically over-predict the observed magnitudes, while others tend to under-predict. As such, the combined suite of models can be used to define upper and lower estimators for the expected magnitudes of upcoming events, as well as empirically constrained statistical expectations for how these magnitudes will be distributed between the upper and lower values. We conclude by demonstrating how our empirically constrained distribution can be used to produce probabilistic forecasts of upcoming induced earthquake magnitudes, applying this approach to two recent cases of induced seismicity.

1. INTRODUCTION

Cases of induced seismicity have grown rapidly over the past two decades, associated with the growth and expansion of oilfield technologies such as hydraulic fracturing, wastewater disposal (WWD), and natural gas storage (NGS). Emerging low-carbon energy technologies such as geothermal and carbon capture and storage, which entail the injection of fluids into the subsurface, also carry the potential to generate induced seismicity.

In severe cases, induced seismicity has caused damage to nearby buildings and infrastructure, and injuries to nearby people (e.g., Lee et al., 2019; Lei et al., 2019; Campbell et al., 2020). Even where induced event magnitudes are insufficient to cause damage, they are nevertheless a source of public concern (e.g., Evensen et al., 2022). A failure to adequately manage induced seismicity during development of subsurface geo-energy projects has led to the cancellation of individual projects and sites, and limits or even moratoria being imposed on entire industries. The need to develop methods to quantify induced seismicity hazard during operations, primarily by estimating what magnitudes of earthquakes are likely to be generated, is clear.

Our aim in this study is to forecast the growth in earthquake magnitudes as induced seismicity sequences develop. We do this by tracking the magnitudes of new record-breaking events – events that are larger than any previous event within a sequence. We refer to these record-breaking magnitudes as M_{NRB} hereafter. The growth of record-breaking events is of particular importance to operators and regulators of subsurface industries, since their magnitudes will usually determine the largest ground motions that are generated, and therefore the largest impact to nearby buildings, infrastructure, and people. If we are able to accurately forecast upcoming record-breaking magnitudes (and preferably, a probability distribution thereof), this could enable operators to make decisions to ensure the safety of their activities by, for example, reducing, ceasing, or applying other mitigation actions to their operations if it becomes likely that unacceptably high magnitudes will be generated.

1.1. Observed versus physically possible induced seismicity magnitudes

The largest record-breaking event within an induced seismicity sequence is, by definition, the largest event within that sequence. The largest observed magnitude during a sequence of induced seismicity (or a forecast thereof) is commonly referred to as M_{MAX} (e.g., Hallo et al., 2014; van der Elst et al., 2016; Eaton and Igonin, 2018; Verdon and Bommer, 2021). This is different from the M_{MAX} parameter used in tectonic seismic hazard assessment, where it denotes the largest magnitude earthquake that is physically possible given the particular tectonic circumstances in question (e.g., Mueller, 2010). The largest possible magnitude represents a truncation to the Gutenberg and Richter (1944) magnitude-frequency distribution (G-R hereafter). We refer to this truncation magnitude as M^T_{MAX} to differentiate these terms.

In making this distinction, we recognise that there is a fundamental difference between tectonic and induced seismicity (Bommer, 2022). Tectonic seismicity is driven by processes acting over geological timescales. Theoretically, all tectonic earthquake populations will eventually be truncated at M^T_{MAX} if we are only able to wait for long enough observation times. In contrast, induced seismicity is driven by a human-induced perturbation that is of limited spatial extent and temporal duration. We are therefore able to observe induced sequences in their entirety, from start to finish. The largest induced event that actually occurs (M_{MAX}) will probably not correspond to the largest possible event at which the G-R distribution would truncate (M^T_{MAX}) unless a sufficient number of induced events have been generated (Zöller and Holschneider, 2016; van der Elst et al., 2016; Eaton and Igonin, 2018).

There are some cases of induced seismicity, usually in settings with fairly specific and unique geomechanical conditions, where truncations of the G-R distribution have been observed (e.g., Verdon et al., 2018). However, for most sequences of induced seismicity there has been little robust evidence of truncations to the G-R distribution at high magnitudes, as would be observed if M^T_{MAX} were regularly being reached (e.g., van der Elst et al., 2016; Watkins et al., 2023). It is therefore reasonable in most cases to treat the magnitudes of an ongoing induced seismicity sequence as being drawn from an unbounded G-R distribution unless specific evidence to the contrary is available.

Furthermore, the accumulation of tectonic strain that drives tectonic earthquakes is assumed to be relatively constant (with respect to the timescales of our observations). In contrast, the human-made perturbations that drive induced seismicity may quickly increase in scale and spatial extent during operations, for example as injection continues in a given well. As a result, induced seismicity sequences may be expected to grow as injection progresses.

While van der Elst et al. (2016) suggested that the order in which induced earthquakes occur is random, subsequent analyses of induced seismicity sequences have shown evidence for progression of event magnitudes as sequences have grown (e.g., Skoumal et al., 2018; Verdon and Bommer, 2021; Watkins et al., 2023). Whereas estimates of the maximum possible magnitude, M^T_{MAX}, should be constant, as this parameter is controlled by underlying physical conditions (e.g., the size and frictional properties of nearby faults), forecasts of M_{MAX} during induced seismicity may be time-dependent, since we should expect a different maximum magnitude event to occur if, for example, we were to inject a given volume of fluid for only 1 month, versus injecting the same volume of fluid every month for a period of years.

1.2. Forecasting induced seismicity magnitudes

A range of methods to forecast magnitudes during induced seismicity sequences have been developed. One approach is to use numerical geomechanical simulations of subsurface processes (e.g., Rutqvist et al., 2013; Verdon et al., 2015; Dempsey and Suckale, 2017). However, such modelling is often difficult to apply in practice since a detailed characterisation of the subsurface is required to generate a model. For many cases, the causative faults on which induced seismicity occurred were not visible in geophysical surveys acquired prior to the onset of industrial activities (e.g., Eaton et al., 2018; Cesca et al., 2021; Nantanoi et al., 2022). Even where faults are successfully imaged, quantification of their mechanical and frictional properties, as required for accurate numerical geomechanical modelling, can be challenging.

The alternative to physics-based numerical modelling is to use statistics-based approaches. For these methods the observed population of seismic events is characterised statistically, and the statistical models are then used to make forecasts of the ongoing seismicity. A commonly used approach is to characterise a relationship between the rate of seismicity and the volume of fluids injected into (or removed from) the subsurface at an early stage of operations (e.g., McGarr, 1976, Shapiro et al., 2010; Hallo et al., 2014; Mancini et al., 2021). The future seismicity can then be forecast by extrapolating this relationship to a future planned injection (or production) volume. This approach has been used to forecast seismicity and guide decision-making for several notable cases of induced seismicity, including the Helsinki St1 Deep Heat project (Kwiatek et al., 2019), the Weyburn Carbon Capture and Storage Project (Verdon, 2016), and during hydraulic fracturing of the Preston New Road shale gas wells in Lancashire, UK (Clarke et al., 2019; Kettlety et al., 2021). Verdon et al. (2024) published a comprehensive appraisal of the performance of the Shapiro et al. (2010) and Hallo et al. (2014) models across a wide range of WWD-induced seismicity case studies.

1.3. Forecasting induced seismicity magnitudes using extreme value estimators

An alternative approach relies solely on the characterisation of the earthquake population, without any reference to injection or production rates or any other subsurface information. This approach, applied by Mendecki (2016) for mining induced seismicity, is based on the theory of extreme value estimators developed by Cooke (1979) and is related to methods developed to estimate tectonic M^T_{MAX} values from observed natural earthquake populations (e.g., Kijko, 2004). The relative simplicity of this method, since it does not require any operational or geological information, is an attractive aspect of this approach. A limitation is the need for a catalog of observed seismicity to make a forecast. However, for cases of induced seismicity we are often able to observe the seismicity to a low magnitude of completeness if dedicated monitoring systems are installed before the start of operations.

Mendecki (2016) applied two approaches to forecasting induced seismicity magnitudes using the order statistics theory of Cooke (1979). For a random sample of n magnitude observations, M^O, drawn from a constant underlying distribution, the upper limit for future such observations can be estimated as:

$$M_{UL} = 2M_n^O - \sum_{i=1}^{n-1}\left[\left(1-\frac{i}{n}\right)^n - \left(1-\frac{i+1}{n}\right)^n\right] M_{n-i}^O \tag{1}$$

where M^O_i represents the event magnitudes sorted into size order, from smallest to largest, such that M^O_n is the largest event observed to date, which we refer to as M^O_{MAX}.

Alternatively, one can consider the jumps in magnitude between events, ΔM^O, since an estimate for the next largest event can be obtained by adding the estimated maximum jump, ΔM_{MAX}, to the observed largest event. We refer to this estimate as the "jump-limited" magnitude:

$$M_{JL} = M_{MAX}^O + \Delta M_{MAX} \tag{2}$$

The maximum jump is calculated using the same formulation as Equation 1, but applied to the distribution of magnitude jumps:

$$\Delta M_{MAX} = 2\Delta M_{n_j}^O - \sum_{i=1}^{n_j-1}\left[\left(1-\frac{i}{n_j}\right)^{n_j} - \left(1-\frac{i+1}{n_j}\right)^{n_j}\right] \Delta M_{n_j-i}^O \tag{3}$$

where ΔM^O_i represents the magnitude jumps ordered from smallest to largest, and n_j is the number of jumps. There are several ways in which these methods can be applied in practice to forecast induced event magnitudes (see Section 2 for further details). For example, since these estimators can be applied to any quantity, the input to these equations can be magnitudes, seismic moments, or potencies.

Our aim in this study is to forecast the magnitudes of new record-breaking events during induced seismicity sequences (M_{NRB}). The two magnitude estimators defined above, M_{UL} and M_{JL}, provide a means by which this can be done. We might normally expect M_{NRB} values to follow the jump-limited estimator, since this explicitly describes the jumps to new record-breaking magnitudes. However, there is a possibility that the next event to occur is at (or close to) the upper limit value as given by the M_{UL} estimator. We therefore might expect to find, in practice, a distribution of M_{NRB} observations, with most cases falling close to the M_{JL} values, but with some events falling closer to the M_{UL} estimate. Hence, our approach is to combine our estimates of M_{UL} and M_{JL} to produce a combined estimator for M_{NRB}.

We note that in forecasting record-breaking events, the implicit assumption is that induced event magnitudes will continue to grow during a sequence. In reality, induced seismicity sequences may stabilise and decrease, either as pressures stabilise in large, open reservoirs (e.g., Verdon et al., 2024), or in response to successful mitigating actions taken by operators. Clearly, forecasting methods that include an implicit assumption that new record-breaking magnitudes will occur may not be appropriate in such circumstances. In Section 5.3 we discuss how it might be possible to identify when an induced seismicity sequence is decaying such that forecasting new record-breaking events is no longer appropriate. Likewise, the methods presented in Equations 1 – 3 do not provide any temporal constraint – when might a new record-breaking event be expected to occur? Again, in Section 5.3 we provide some discussion as to how temporal constraints could be introduced.

1.4. The need for performance assessment of induced seismicity forecasting models

If induced seismicity forecasting models are to be used to guide decision-making at active industrial sites, then there is a clear need for robust, transparent testing of such models. Only through robust testing can we gain confidence in the performance of models such that they can be relied on to guide operational decisions that, on the one hand, may compromise significant financial investments (if projects are abandoned due to potential induced seismicity hazard), while on the other hand could compromise public safety (if larger magnitude events are allowed to occur without mitigation). The public often takes a strong interest in the occurrence of induced seismicity, and so model testing must be transparent and reproducible as a loss of trust of public in ability to safely conduct underground energy operations easily results in loss of social license to operate and rejection of future projects.

Empirical testing of forecasting models can go beyond simple assessments of performance since results can be used to feed back into future forecasts. In our case, we anticipate that record-breaking magnitudes will follow the M_{JL} estimator but we allow for the possibility that magnitudes could jump to the upper limit M_{UL} value. As such, the M_{JL} and M_{UL} values may provide lower and upper estimates for M_{NRB}, respectively. A suite of models could be combined to produce an overall estimate (and preferably, a probability distribution thereof) for upcoming induced event magnitudes. An overall estimate from a suite of models should consider the observed performances of the different modelling strategies as applied to large numbers of induced seismicity case studies.

1.5. Study objectives

The objective of this study is to provide a systematic assessment of the performance of the M_{UL} and M_{JL} estimators as applied to a large number of cases of injection-induced seismicity. We evaluate several different ways in which these methods can be applied, for example using earthquake magnitudes versus potencies as the inputs to Equations 1 – 3; and using all observed events and jumps as inputs versus only the events and jumps that represent new record-breaking events (see Section 2). In doing so, we investigate the influence of these different formulations on the resulting M_{NRB} forecasts and quantitatively compare their respective performances.

Our observations across a large number of induced seismicity sequences provide empirical data on the behaviour of record-breaking magnitudes relative to the M_{UL} and M_{JL} estimators. These observations allow us to define an empirically constrained estimator for M_{NRB}, where the next record-breaking magnitude is expected to fall within a statistical distribution that is defined based on the M_{UL} and M_{JL} estimates.

2. Methods

Equations 1 – 3 describe two approaches to estimating induced event magnitudes. M_{UL} describes the expected upper limit magnitude based on the population of observed events to date. M_{JL} defines the expected next record-breaking magnitude based on the population of magnitude jumps, with the largest expected magnitude jump being added to the largest observed event to date.

For both of these estimates, calculations can use either the earthquake magnitudes or seismic moments, M_O (or potencies, $P = M_O/G$, where G is the shear modulus). Hereafter, we refer to results computed using magnitudes with the subscript $_{MM}$, and results computed using potencies with the subscript $_{MO}$. Furthermore, the magnitudes and magnitude jumps used as inputs to Equations 1 – 3 can be taken from the entire event catalog, where $M^O{}_i$ represents the entire event population sorted into size order and $\Delta M^O{}_i$ represents the magnitude (or potency) jump between every event when the entire population is sorted into magnitude order, with $\Delta M^O{}_i$ then being sorted into size order. Alternatively, one can use an event population that consists only of the record-breaking events as they appear in a sequence, where $M^O{}_i$ represents the record-breaking events sorted into size order, and $\Delta M^O{}_i$ represents the jumps between the record-breaking events. Hereafter, we refer to calculations using the entire event population resorted into size order with the subscript $_{AE}$ (for all events) and calculations using only the record-breaking events as $_{RB}$ (for record-breaking events). These combinations mean that we have a total of 8 possible ways in which induced event magnitudes can be estimated. These are summarised in Table 1.

Table 1: Summary of different model implementations used for M_{NRB} forecasting

Model No.	Model Name	Upper Limit [UL] or Jump-Limited [JL] formula	All Events in Size Order [AE] or Record Breaking only [RB]	Magnitudes [MM] or Potencies [MO]
1	$M_{UL\_RB\_MM}$	UL	RB	MM
2	$M_{UL\_RB\_MO}$	UL	RB	MO
3	$M_{UL\_AE\_MM}$	UL	AE	MM
4	$M_{UL\_AE\_MO}$	UL	AE	MO
5	$M_{JL\_RB\_MM}$	JL	RB	MM
6	$M_{JL\_RB\_MO}$	JL	RB	MO
7	$M_{JL\_AE\_MM}$	JL	AE	MM
8	$M_{JL\_AE\_MO}$	JL	AE	MO

We note that dedicated microseismic monitoring arrays often produce large numbers of events (e.g., Verdon and Budge, 2018), but even for a very large catalog ranging across several orders of magnitude we often observe only a few record-breaking events. Thus, the methods based on record-breaking versus all events represent different approaches to statistical estimates. By definition, the record-breaking method excludes aftershocks as these are smaller than, and occur after, a mainshock and therefore do not contribute to record-breaking series. However, the approach based on all events includes aftershocks in the evaluation of the maximum magnitude while representing whole sequence.

Given the different ways in which these estimators can be applied to induced seismicity sequences, there is a clear need to produce a quantitative comparison of their relative performance in forecasting magnitudes during induced seismicity sequences. Several studies have applied various versions of the M_{UL} and/or M_{JL} methods to cases of induced seismicity (Cao et al., 2020; Verdon and Bommer, 2021; Watkins et al., 2023; Schultz et al., 2023a; Cao et al., 2024). In general, these studies have produced results that show that, at least from a qualitative perspective, these methods do provide useful forecasting potential. Whereas Mendecki (2016) formulated these methods in terms of seismic potency, all of the later studies have used earthquake magnitudes. Cao et al. (2020) applied the M_{UL} and M_{JL} methods to the seismicity induced by gas production at Groningen and to a case of hydraulic fracturing-induced seismicity in North America. In their calculations, they used all events and jumps within the catalogs, not just record-breaking ones.

Verdon and Bommer (2021) applied the M_{JL} approach to a compilation of 22 instances of hydraulic fracturing-induced seismicity, and Watkins et al. (2023) applied the M_{JL} approach to 27 cases of seismicity induced by WWD and NGS. Like Cao et al. (2020), Verdon and Bommer (2021) and Watkins et al. (2023) used the jumps between all events (when sorted into size order), not just the jumps to new record-breaking events.

Cao et al. (2024) applied the M_{JL} approach to 15 cases of induced seismicity (mostly consisting of the same hydraulic fracturing sequences examined by Verdon and Bommer, 2021), but using as input to their model only the population of jumps that created new record-breaking events. Schultz et al. (2023a) applied the M_{JL} approach to the sequence of WWD-induced seismicity at Musreau Lake, Alberta. Like Cao et al. (2024), they used as inputs only the population of jumps that created new record-breaking events.

For all the above studies, the assessment of model performance has been somewhat unsystematic. Mendecki (2016) demonstrated his methods by application to a single example of mining-induced seismicity but did not make any quantitative assessment of model performance. Likewise, Cao et al. (2020) and Schultz et al. (2023a) simply compared the evolution of the observed earthquakes with the changing M_{NRB} estimates, noting that the models generally did a reasonable job of fitting the observed magnitudes. Verdon and Bommer (2021) and Watkins et al. (2023) produced cross-plots of modelled versus observed M_{MAX} (the largest magnitude within each sequence), while Cao et al. (2024) compared modelled and observed magnitudes each time a new record-breaking event occurred (M_{NRB}). These plots showed evidence for correlation between observed and modelled magnitudes, but also showed that at times the M_{JL} model can underestimate M_{NRB}. As such, there has not yet been any effort to systematically quantify the performance of these methods, either between the different methods, or for the same method between different sites. In the following section we introduce the datasets that we use to assess the performance of each method, before presenting our results in Section 4.

3. DATASETS

3.1. Oklahoma and southern Kansas

WWD in central and northern Oklahoma and southern Kansas (OK-KS hereafter) has increased significantly over the past two decades, driven primarily by a move towards hydrocarbon production from reservoirs with high water fractions, with the produced water then requiring disposal (Rubenstein and Mahani 2015). WWD, primarily into the deep Arbuckle Formation, has caused significant amounts of induced seismicity (Weingarten et al. 2015), including some of the largest induced events to have ever been recorded from fluid injection activities, such as the M 5.6 Prague (Keranen et al., 2013) and M 5.8 Pawnee (Yeck et al., 2017) sequences. Induced seismicity in Oklahoma has also been caused by hydraulic fracturing (e.g., Holland, 2013; Skoumal et al., 2018; Verdon and Rodríguez-Pradilla, 2023), particularly in the Anadarko Basin. However, our focus here is on central and northern Oklahoma and southern Kansas, where the bulk of the seismicity is caused by WWD.

In this study we use the earthquake catalog published by Park et al. (2022), who used the PhaseNet deep learning model (Zhu and Beroza, 2019) to detect earthquakes recorded by publicly available seismic networks in the OK-KS region. The deep learning model produced a significant increase in event detection, improving detection thresholds by at least 1 magnitude unit over pre-existing earthquake catalogs for the region. We adopt a minimum magnitude of completeness of $M_C = 1.5$, based on the magnitude-frequency relationships plotted in Figure 2 of Park et al. (2022). To estimate potencies from the given magnitudes, we adopt a single value of $G = 20$ GPa (this value is adopted for all sequences in our study).

There are 70 earthquakes in the Park et al. (2022) catalog with magnitudes ≥ 4.0. Some of these events occur in close spatial proximity to each other such that they can be considered to be part of the same sequence. Park et al. (2022) identified clear, discrete fault structures that were responsible for hosting most of the larger magnitude events. These structures typically had lengths of between 5 – 20 km (see Figures 1 and 2 of Park et al., 2022). Where multiple M ≥ 4.0 events were located within 10 km of each other, we treated these as being part of the same sequence of induced events. In doing so, we identified 24 individual sequences in which induced event magnitudes reached or exceeded M 4.0 (see Figure 1). We take these 24 sequences as test datasets for our analysis. For each case, we define a 20 × 20 km square around the $M \geq 4.0$ event (or the largest event for sequences which contain more than one $M \geq 4.0$ event). All earthquakes within this square are taken as representing part of the sequence and used to perform our M_{MAX} forecasting. The $M \geq 4.0$ events, and the 20 × 20 km squares around them, are shown in Figure 1. The choice of dimensions (20 × 20 km) was somewhat arbitrary, but we found that such dimensions were usually sufficient to capture the bulk of the seismic events that occurred on each of the discrete fault strands that hosted larger events, as identified by Park et al. (2022).

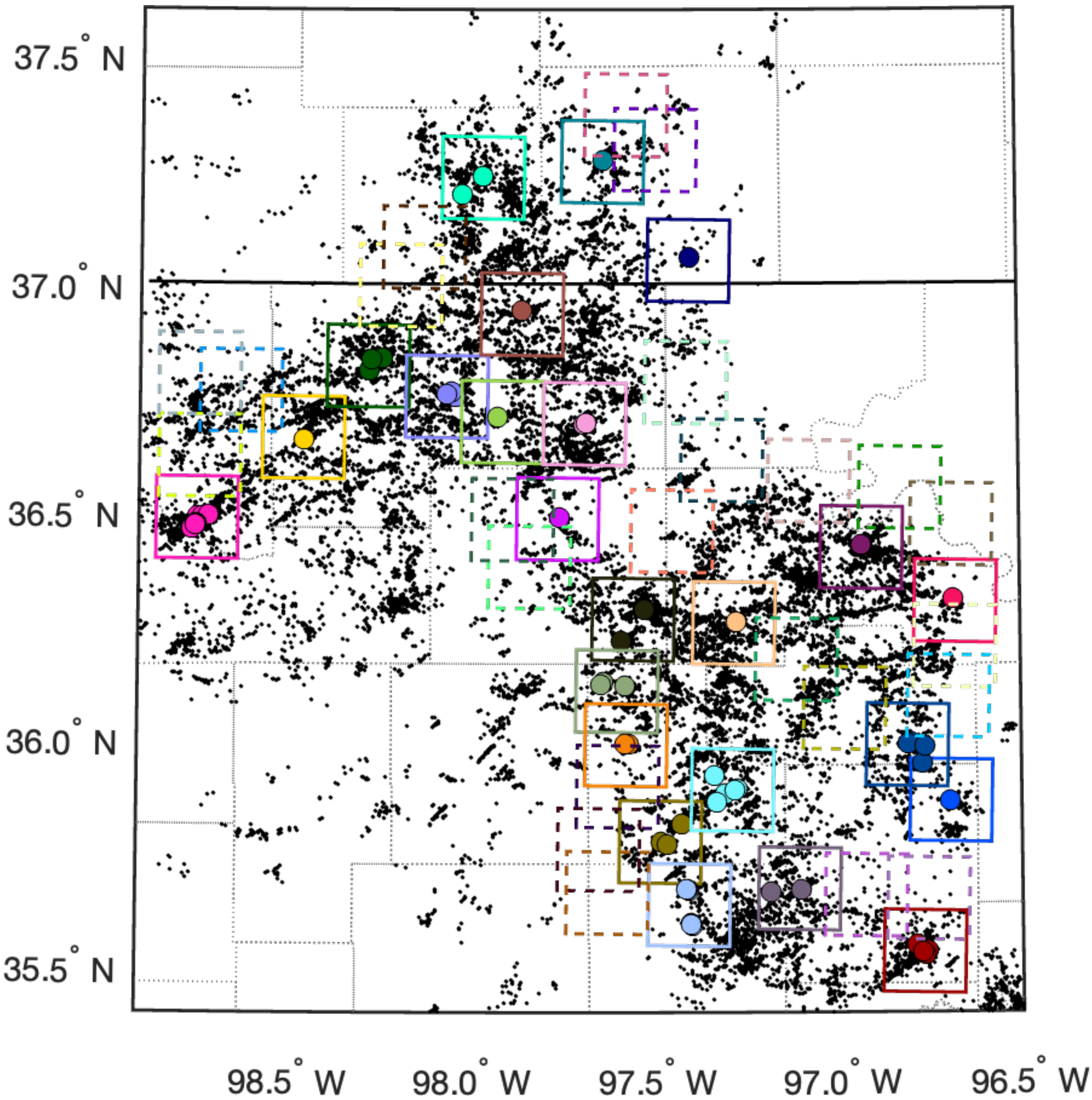


Figure 1: Map of the OK-KS study area. Black dots show all earthquakes with $M \geq 1.5$ and coloured circles show events with $M \geq 4.0$. The solid boxes show the 20 × 20 km blocks around each of the sequences containing $M \geq 4.0$ events, while the dashed boxes show 20 × 20 km blocks in which 500 events were recorded with no $M \geq 3.5$ events. The box colours used in this figure correspond to the marker colours used in Figure 3.

In testing induced seismicity forecasting models, there can be a tendency to focus on cases where larger magnitude events occurred, since these cases tend to attract the most attention (from the public and policy makers, as well as from academics). However, comprehensive testing should include sequences that did not reach larger magnitudes, since our objective is to develop models that can differentiate between sequences that do, and that do not, escalate to higher magnitude events. Hence, in addition to the 24 sequences with $M \geq 4.0$ events, we identify the same number of cases where magnitudes did not exceed M 3.5, selecting twenty-four 20 × 20 km blocks at random within the study area that contained at least 500 events but no events with $M \geq 3.5$. To do so, we randomly generated block positions and rejected those that did not meet these criteria, continuing until we had 24 cases. The 24 blocks without larger magnitude events are also shown in Figure 1.

There is some overlap between the different blocks that are treated hereafter as discrete induced seismicity sequences, meaning that some events are included in more than one forecast. This will create some partial dependence between results from individual sequences. However, in our view a smaller event that is mid-way between the future locations of two different larger events could be reasonably considered to be a precursor to either or both, and so it is reasonable that such events could be included within the forecasts for both larger events, and this partial dependence cannot therefore be avoided.

3.2. Permian Basin, western Texas

Induced seismicity has been recognised in the Permian Basin of western Texas (WTX hereafter) since the 1970s (Davis and Pennington, 1989). Rates of seismicity in the basin have increased substantially since 2015 (Skoumal et al., 2020), associated with WWD and hydraulic fracturing. Given the co-location of these activities, distinguishing causality between WWD and hydraulic fracturing can be challenging, although the bulk of the seismicity is thought to have been caused by WWD (Grigoratos et al., 2022). Three $M \geq 5.0$ events have been induced in this basin: the March 2020 M 5.0 event near to the city of Pecos in Reeves County (Skoumal et al., 2021), the November 2022 Coalson Draw M 5.4 event in western Reeves County, and the December 2022 M 5.2 event in Martin County, just to the north of the city of Midland (Hennings and Young, 2023).

In this study we use the TexNet earthquake catalog (Savvaidis et al., 2019), with data running from the start of 2017 until April 2023. We computed the minimum magnitude of completeness by evaluating the lowest magnitude at which the cumulative magnitude-frequency distribution was consistent with the G-R distribution, as assessed by the Kolmogorov-Smirnov test with an acceptance criterion of 10 % (Clauset et al., 2009), which gave $M_C = 2.0$. There are 48 events for which $M \geq 4.0$ (Figure 2). Our examination of the temporal and spatial evolution of the seismicity identified 11 individual sequences in which induced event magnitudes reached or exceeded M 4.0. Much like for our OK-KS datasets, we define 20 × 20 km squares around each sequence and use all events within these blocks to perform our M_{NRB} forecasting. We then identify an equal number (i.e., 11) of 20 × 20 km blocks containing at least 100 events (we use a lower criterion here recognising the lower number of events in the TexNet catalog compared to the Park et al. (2022) catalog for OK-KS) but no events larger than M 3.5, in order to test M_{NRB} model performance for cases where larger magnitude events did not occur.

3.3. Watkins et al. (2023) sequences

Watkins et al. (2023) published M_{MAX} forecasts using the $M_{JL\_SO\_MM}$ formulation for more than 20 individual sequences of WWD and NGS-induced seismicity. Some of the Watkins et al. (2023) sequences are already included in our OK-KS and WTX datasets described in the previous sections (Reeves and Cogdell in Texas, Cushing, Fairview, Guthrie-Langston, Pawnee and Prague in Oklahoma, Milan and Harper in Kansas), while for some older sequences with lower levels of monitoring, the largest events occurred before a sufficient number of events were available to compute M_{NRB} estimates (e.g., the Cordel sequence in Alberta). This left 16 additional sequences which we were able to include in our analysis, including: the Azle-Reno, Dallas-Fort Worth, Venus, Timpson and

Irving sequences in eastern Texas (Hennings et al., 2021; Frohlich et al., 2014); the Guy-Greenbrier sequence in Arkansas (Horton, 2012); the Youngstown sequence in Ohio (Kim, 2013); the Paradox Valley, Greeley and Raton Basin sequences in Colorado (Block et al., 2014; Yeck et al., 2016; Nakai et al., 2017); the Eagle West, Graham, and Musreau Lakes sequences in western Canada (Horner et al., 1994; Hosseini and Eaton, 2018; Li et al., 2022); the Rongchang sequence in the Sichuan Basin (Wang et al., 2020); the Castor project in the Gulf of Valencia, Spain (Cesca et al., 2021); and the Puerto Gaitán sequence, Colombia (Molina et al., 2020). For each of these sequences, we use the earthquake catalogs published in the Supplementary Materials of Watkins et al. (2023). We refer to these sequences as the W23 cases hereafter.

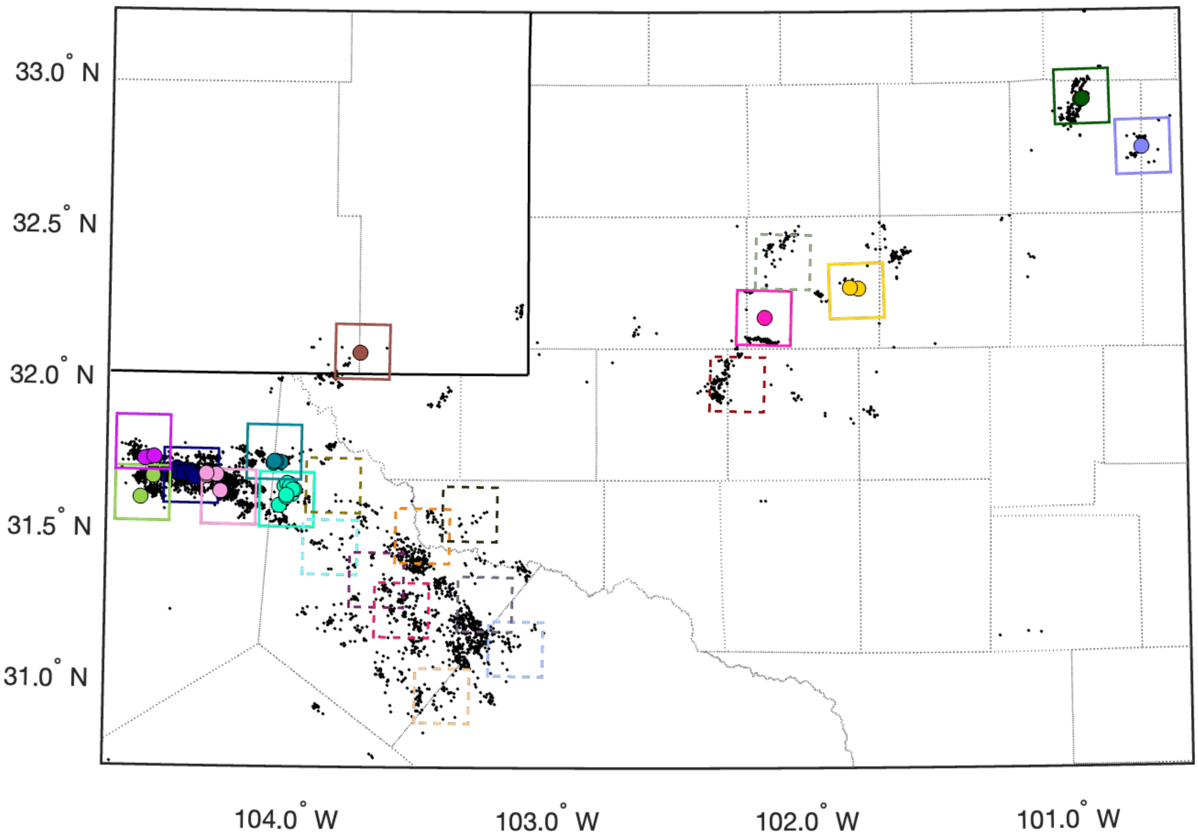


Figure 2: Map of the western Texas study area. Black dots show all earthquakes with $M \geq 1.0$ and coloured circles show events with $M \geq 4.0$. The solid boxes show the 20 × 20 km blocks around each of the sequences containing $M \geq 4.0$ events, while the dashed boxes show 20 × 20 km blocks in which 100 events were recorded with no $M \geq 3.5$ events. The box colours used in this figure correspond to the marker colours used in Figure 4.

3.4. Application

For the OK-KS and WTX datasets we compute M_{NRB} values at intervals of 0.5 months, starting at the time when at least 10 events above the magnitude of completeness within the sequence have been recorded, and continuing for the duration of the available catalog. For the W23 sequences, the timespans of each sequence are highly variable – we therefore compute M_{NRB} values at 1,000 evenly-spaced intervals between the first and final event within each sequence. At each time step we estimate the next record breaking magnitude in a pseudo-prospective manner, using all the events in the sequence that occurred prior to a given time to estimate M_{NRB} for the next time interval.

Our objective is to assess the forecast performance as each sequence evolves. We therefore make comparisons between observed and modelled magnitudes each time there is a new largest event within the sequence. Each new largest event within the sequence is treated as an observed record-breaking event, M^{O}_{NRB}. The M^{O}_{NRB} values are compared against the M_{NRB} values calculated at the timestep prior to when the M^{O}_{NRB} event occurred. For the calculations made using potencies, the modelled values are converted back to magnitude to facilitate a comparison with the observed magnitudes.

4. RESULTS

Figures 3, 4 and 5 show our results, comparing the observed and forecast M^{O}_{NRB} and M_{NRB} values, using each of the 8 methods described in Table 1, for the sequences from OK-KS (Figure 3), WTX (Figure 4), and the W23 sequences (Figure 5). In total we have applied our models to 86 sequences (48 in OK-KS, 22 in WTX, 16 from W23), with a combined total of 331 individual record-breaking events within these sequences (205 from OK-KS, 72 from WTX, 54 from W23). The time evolution of every individual sequence, and the corresponding modelled M_{NRB} values, are provided in the supplementary materials (Section S3).

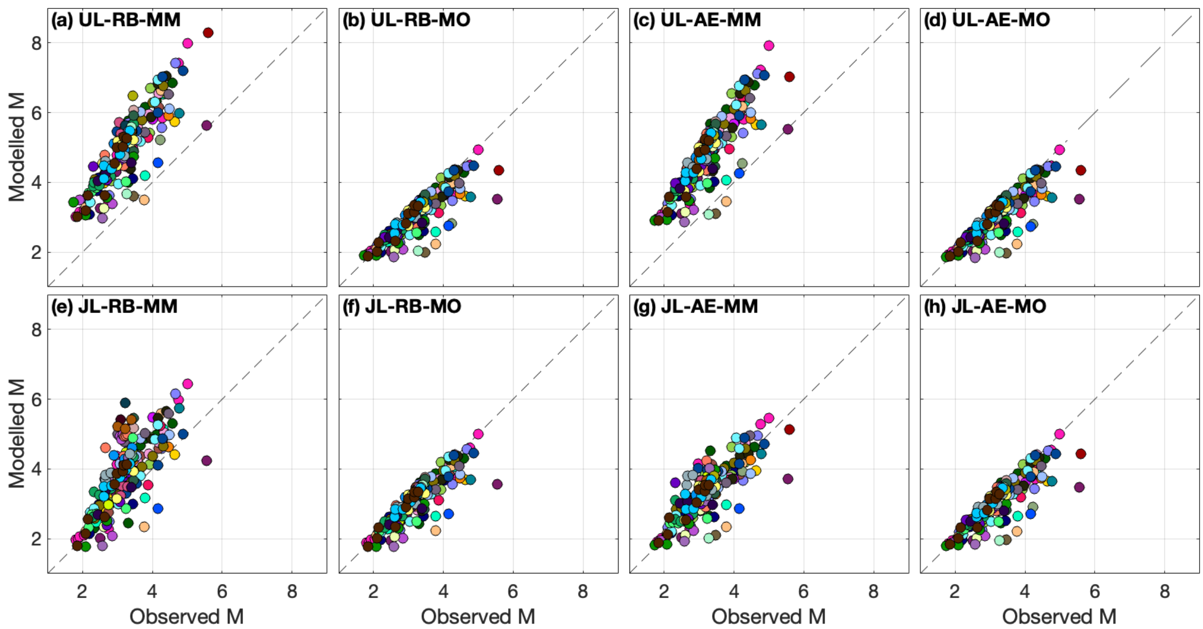


Figure 3: Results for OK-KS sequences comparing observed and modelled magnitudes for each of the M_{NRB} forecasting methods listed in Table 1. Marker colours correspond to sequences within each box shown in Figure 1.

We quantify the model performance using several metrics. We compute the root-mean-squared (RMS) error between modelled and observed magnitudes, σ_{RMS}, the Pearson correlation coefficient between modelled and observed magnitudes, r, and the gradient of the line of (least squares) best-fit, m. A well-performing model should minimise σ_{RMS} and maximise r, and have a best-fit gradient close to 1.0, implying a 1:1 relationship between M_{NRB} and M^{O}_{NRB}. Additionally, in most applications we anticipate that M_{NRB} forecasting will be used to guide operational decision making in order to avoid unwanted large events. It is therefore of particular importance that models do not make large underpredictions, such that the actual seismicity significantly exceeds what has been forecast by the model. We therefore compute N_{UP}, the percentage of M^{O}_{NRB} instances where the forecast M_{NRB} value was a significant underprediction with $M_{NRB} < M^{O}_{NRB} - 0.5$. These metrics are listed in Table 2 for the OK-KS, WTX and the W23 sequences respectively.

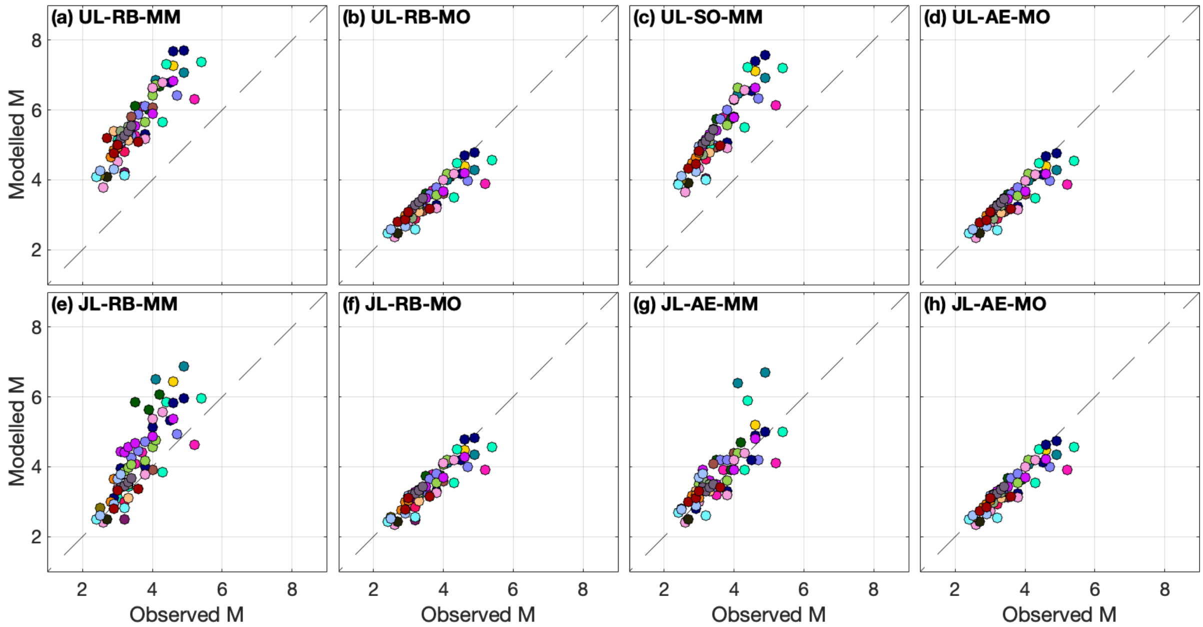


Figure 4: Results for WTX sequences comparing observed and modelled magnitudes for each of the M_{NRB} forecasting methods listed in Table 1. Marker colours correspond to sequences within each box shown in Figure 2.

In general, we observe strong correlation between the modelled and observed M_{NRB} values, implying that these methods all provide useful forecasting information for induced seismicity magnitudes, and could therefore be used as part of a decision-making strategy to manage induced seismicity. The performance of these models is generally better than that found by Verdon et al. (2024) for commonly used volume-based forecasting models, having higher correlation coefficients between modelled and observed magnitudes, lower RMS errors (except for the $M_{UL\_RB\_MM}$ and $M_{UL\_RB\_MM}$ models, see below), and fewer cases where models produced significant underpredictions of upcoming magnitudes.

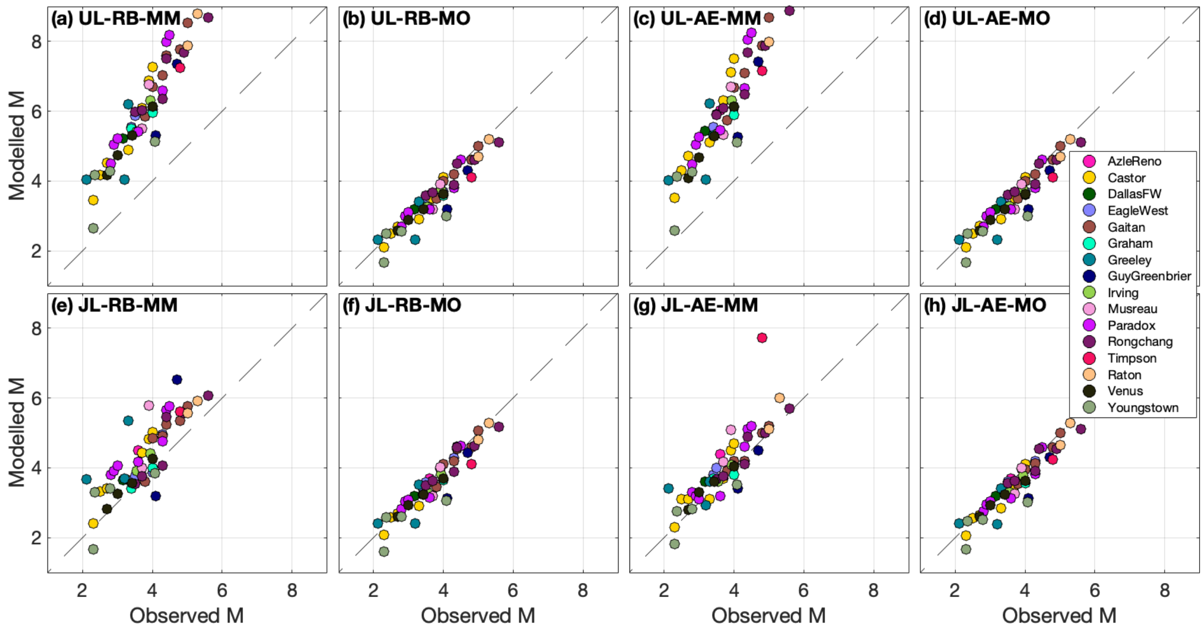


Figure 5: Results for the W23 sequences comparing observed and modelled magnitudes for each of the M_{NRB} forecasting methods listed in Table 1.

Table 2: Performance metrics for OK-KS, WTX and W23 sequences.

Model	σ_{RMS}	***r***	***m***	N_{UP} **[%]**
		OK-KS		
$M_{UL\_RB\_MM}$	1.84	0.86	1.27	0
$M_{UL\_RB\_MO}$	0.41	0.86	0.76	14.2
$M_{UL\_AE\_MM}$	1.67	0.86	1.24	0
$M_{UL\_AE\_MO}$	0.41	0.85	0.76	14.2
$M_{JL\_RB\_MM}$	0.93	0.75	1.11	3.4
$M_{JL\_RB\_MO}$	0.37	0.87	0.82	12.7
$M_{JL\_AE\_MM}$	0.47	0.81	0.85	7.3
$M_{JL\_AE\_MO}$	0.41	0.85	0.78	14.6
		WTX		
$M_{UL\_RB\_MM}$	2.06	0.90	1.23	0
$M_{UL\_RB\_MO}$	0.32	0.92	0.78	12.5
$M_{UL\_AE\_MM}$	1.84	0.91	1.26	0
$M_{UL\_AE\_MO}$	0.32	0.92	0.78	12.5
$M_{JL\_RB\_MM}$	0.89	0.83	1.35	2.8
$M_{JL\_RB\_MO}$	0.32	0.91	0.81	12.5
$M_{JL\_AE\_MM}$	0.54	0.80	0.98	5.6
$M_{JL\_AE\_MO}$	0.32	0.91	0.79	12.5
		W23		
$M_{UL\_RB\_MM}$	2.37	0.93	1.62	0
$M_{UL\_RB\_MO}$	0.34	0.94	0.93	11.1
$M_{UL\_AE\_MM}$	2.43	0.92	1.66	0
$M_{UL\_AE\_MO}$	0.34	0.94	0.93	11.1
$M_{JL\_RB\_MM}$	0.81	0.83	1.04	3.7
$M_{JL\_RB\_MO}$	0.34	0.93	0.94	11.1
$M_{JL\_AE\_MM}$	0.59	0.85	1.05	3.7
$M_{JL\_AE\_MO}$	0.34	0.94	0.93	9.3

More detailed inspection of Figures 3 – 5 and Table 2 leads us to the following conclusions, all of which are consistent between the OK-KS, WTX, and W23 sequences.

Using re-sorted magnitudes or just record-breaking events does not significantly change forecasting performance. The use of the entire earthquake catalog, versus solely using record-breaking events (or jumps to record-breaking events), was a key point of difference between Cao et al. (2020), Verdon and Bommer (2021) and Watkins et al. (2023) on the one hand, and Cao et al. (2024) and Schultz et al. (2023a) on the other. However, comparison of panels (a) vs (c), (b) vs (d), (e) vs (g), and (f) vs (h) of Figures 3 – 5 show that these different implementations in fact produce very similar results. From examination of Equations 1 and 3 this outcome is unsurprising since only the first few terms of the weighting applied to the summation of the magnitudes (or jumps), given by:

$$W_i = \left(1 - \frac{i}{n}\right)^n - \left(1 - \frac{i+1}{n}\right)^n \tag{4}$$

are significant (Mendecki, 2016). The first weightings correspond to the largest magnitudes (or magnitude jumps), which tend to be magnitudes (or jumps) that produce record-breaking events.

Figure 6 plots the value of W_i as a function of *i and n*. The weighting term drops to values of 0.01 or less after the 4th term in the summation (the weighting applied to the 4th-largest magnitude or jump). The fact that only a few values are required to produce stable magnitude estimates is an additional advantage of this approach, since it can be applied even where only a few initial events have been observed in a new sequence.

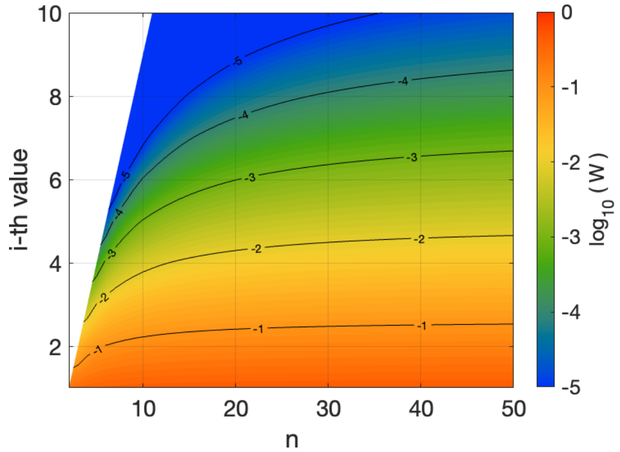


Figure 6: Value of the weighting W applied within the summation term in Equations 1 and 3 (as defined in Equation 4) as a function of i. The contours here show values of $log_{10}(W)$. For any value of n, the weighting for terms where i > 4 is less than 0.01.

Upper limit models using magnitude provide a credible upper limit. The $M_{UL\_AE\_MM}$ and $M_{UL\_RB\_MM}$ models (panels (a) and (c) in Figures 3 – 5) did not produce any significant underpredictions ($N_{UP} = 0$). This is notable given that we have applied it to 86 individual earthquake sequences. Hence, the *UL_MM* values (upper limit calculations using magnitudes) do seem to provide a credible upper limit to induced earthquake magnitudes.

However, while these values never produced underpredictions, they did not provide a good fit to the evolution of record-breaking magnitudes within sequences, tending to produce significant overpredictions in most cases. This is to be expected since the M_{UL} method is formulated to estimate the largest possible value within a distribution, not the expected next record-breaking event. As a result, the $M_{UL\_AE\_MM}$ and $M_{UL\_RB\_MM}$ models gave the largest σ_{RMS} values, and best-fit relationships with the gradient *m* significantly higher than 1.0. That said, the correlation coefficients for the $M_{UL\_AE\_MM}$ and $M_{UL\_RB\_MM}$ models are not significantly worse than those of other models, implying that the scatter between modelled and observed magnitudes is no worse than for the other models, just the fit is not along the 1:1 line, resulting in systematic overprediction.

Next record-breaking models using magnitudes produce the highest scatter. Although the $M_{JL\_AE\_MM}$ and $M_{JL\_AE\_MM}$ models (panels (e) and (g) in Figures 3 – 5) produced reasonable fits between observed and modelled magnitudes, with the gradient *m* close to 1.0, these models had the lowest correlation coefficients of all the models, and the highest σ_{RMS} values with the exception of the overpredicting $M_{UL\_AE\_MM}$ and $M_{UL\_RB\_MM}$ models, as described above. The $M_{JL\_AE\_MM}$ and $M_{JL\_RB\_MM}$ models therefore produced the highest scatter between modelled and observed magnitudes and may therefore have the least utility in forecasting. This is ironic given that this approach has been the most widely used to date, forming the basis of results presented by Cao et al. (2020; 2024), Verdon and Bommer (2021), Watkins et al. (2023) and Schultz et al. (2023a).

Potency-based models have the least scatter, but significantly underpredict on occasion. All four of the models that used earthquake potencies, $M_{UL\_AE\_MO}$, $M_{UL\_RB\_MO}$, $M_{JL\_AE\_MO}$, and $M_{JL\_RB\_MO}$ (panels (b), (d), (f) and (h) in Figures 3 – 5) produced very similar results. These models had the lowest σ_{RMS} values and highest correlation coefficients, indicating that these models had low scatter and the closest match between modelled and observed magnitudes. However, these models also produced the largest number of underpredictions, with between 10 – 15 % of events being underpredicted by more than 0.5 magnitude units. We surmise that in most cases where sequences are evolving relatively gently, the potency-based models perform well. However, they do not perform as well in capturing the more unusual sequences where a sharp increase in magnitudes takes place.

5. DISCUSSION

5.1. Towards an empirically constrained probabilistic model

Our results show that the upper limit magnitude-based models, $M_{UL\_AE\_MM}$ and $M_{UL\_RB\_MM}$, provided credible upper bounds for the actual event magnitudes, having no significant underpredictions after application to a large number of sequences. However, in most cases these models overpredicted the observed events. In contrast, the potency-based models ($M_{UL\_AE\_MO}$, $M_{UL\_RB\_MO}$, $M_{JL\_AE\_MO}$, and $M_{JL\_RB\_MO}$) generally produced a good fit to the observed magnitudes, but occasionally produced significant underpredictions.

From this, it is reasonable to propose a composite approach to forecasting event magnitudes where $M_{UL\_AE\_MM}$ or $M_{UL\_RB\_MM}$ is used provide an upper estimator for the expected magnitude of the next record-breaking event and $M_{UL\_AE\_MO}$, $M_{UL\_RB\_MO}$, $M_{JL\_AE\_MO}$, or $M_{JL\_RB\_MO}$ is used to provide a lower estimator for the expected magnitude. Hereafter, we use $M_{UL\_RB\_MM}$ for the upper estimator and $M_{JL\_AE\_MO}$ for the lower estimator, referred to hereafter as M_{UE} and M_{LE} respectively.

The probability distribution of event magnitudes between these estimators can be evaluated through empirical calibration with our observed seismicity. For each event, we normalise each observed record-breaking event magnitude relative to the M_{LE} and M_{UE} estimators at the time of the event's occurrence:

$$M_N^O = \frac{M_{NRB}^O - M_{LE}}{M_{UE} - M_{LE}} \tag{5}$$

We then examine the distribution of these normalised magnitudes – where do events typically fall with respect to the upper and lower magnitude estimators? Our results for each of our studies are shown in Figure 7. The distributions of $M^O{}_N$ are consistent between the three sets of sequences that we studied. Most values are close to 0, i.e., they match the modelled lower estimator values, $M^O{}_{NRB} = M_{LE}$. However, the distribution has a tail of higher values extending towards 1, i.e., where observed magnitudes reach towards the higher estimator values, $M^O{}_{NRB} = M_{UE}$.

We examine the fit of various statistical distributions to our observations, including lognormal, a Gumbel, and Generalised Extreme Value (GEV) distributions. We further test the performance of these distributions when applied to synthetically generated sequences. These results are shown in our Supplementary Materials (Sections S1 and S2). The consistency found for $M^O{}_N$ between our different case studies and synthetic models enables us to construct an empirically constrained probabilistic model for induced seismicity forecasting using extreme value estimators. We find that our observations are reasonably approximated either by a shifted lognormal distribution with a mean of μ_{LN} = -1.4, a deviation of σ_{LN} = 0.6, and a shift of δ_{LN} = 0.2, or a GEV distribution with shape parameter k_{GEV} = 0.23, scale parameter σ_{GEV} = 0.1, and location parameter μ_{GEV} = 0.0. Hereafter we use the GEV distribution as providing the best fit to our combined observations (see Supplementary Materials Section S1).

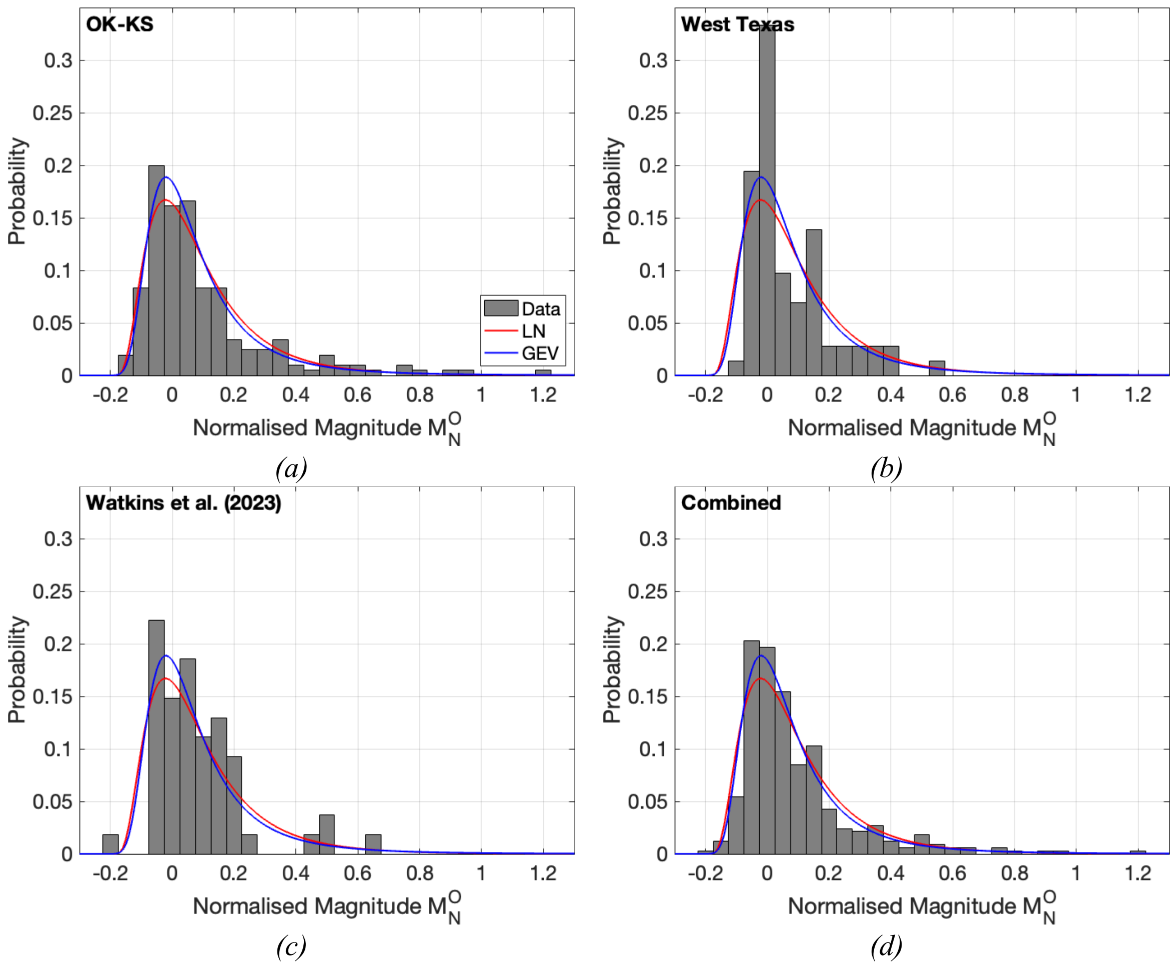


Figure 7: Distribution of normalised observed magnitudes M^O_N (bars), where the observed magnitudes are normalised relative to the modelled upper and lower estimators, for the OK-KS (a), WTX (b), and W23 (c) sequences, and for all observations combined (d). The red and blue lines show the shifted lognormal and GEV distributions that we adopt to approximate the observed distributions.

For a given sequence of seismicity, we compute the M_{UE} and M_{LE} estimators at a given time. Having computed M_{UE} and M_{LE}, we can compute the probabilities for the next largest magnitude event that will occur in the sequence. We use Equation 5 to normalise magnitudes relative to M_{UE} and M_{LE}, and then estimate the probability of occurrence for any magnitude event from the GEV distribution with scale, shape and location parameters described above.

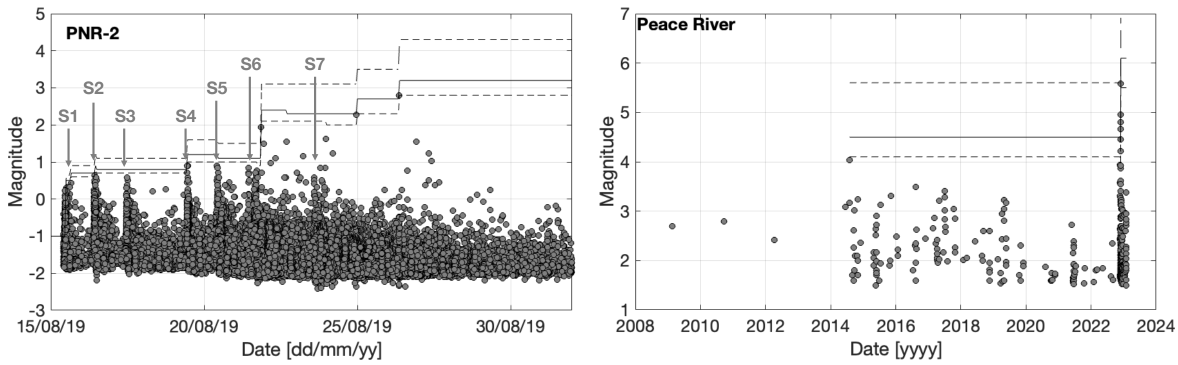
Our synthetic testing (Supplementary Materials Section S2) shows that the observed distributions are consistent with situations where no upper truncation is applied to the G-R distribution from which the events are drawn (or where the magnitude of truncation is much larger than the observed event sizes, such that is has, in effect, no impact on the simulated magnitudes). Where a truncation is applied to our synthetic tests, the M^O_N values are systematically shifted towards the lower estimator ($M^O_N = 0$), such that the representative distributions defined above are no longer appropriate. The similarities between our observed distributions and those generated by an untruncated model, alongside past studies which have generally failed to find significant evidence for magnitude truncations in most induced seismicity cases (e.g., van der Elst et al., 2016) suggest that our approach is reasonable with respect to this caveat. However, if clear upper truncations to the G-R distribution are observed for induced seismicity sequences (e.g., Verdon et al., 2018), then alternative methods for M_{MAX} estimation, such as those that explicitly assume an upper-truncated G-R distribution (e.g., Kijko and Sellevoll, 1989; Pisarenko et al., 1996; Holschneider et al., 2011), may be preferable.

5.2. Application to out-of-sample cases

We demonstrate this approach by application to two notable cases of induced seismicity: from hydraulic fracturing at the Preston New Road PNR-2 well in Lancashire, England in 2019 (Kettlety et al., 2021), and from seismicity associated with WWD activities in north-western Alberta, near to the town of Peace River (Schultz et al., 2023b). The PNR-2 sequence is notable because its occurrence led the UK government to impose a moratorium on hydraulic fracturing, primarily because of the perceived inability to "accurately predict the probability or magnitude of earthquakes linked to fracking operations" (BEIS, 2019).

The Peace River sequence reached a magnitude of *M* 5.6 in November 2022. If induced (the nature of this event is still disputed, see Salvage et al., 2024) it would be the largest magnitude induced event in the Western Canada Sedimentary Basin. This sequence is useful for our purposes since, given when it occurred, it was not included in the sequences compiled by Watkins et al. (2023), and so it represents an out-of-sample test, since the sequences in W23 were used to generate our empirically constrained distribution of M^O_N.

For the PNR-2 sequence, we use the corrected moment magnitudes published by Kettlety and Butcher (2022) – note that these M_W values are different from the M_L values published by Kettlety et al. (2021). For the Peace River sequence, we use earthquakes from the Alberta Geological Survey database (AGS, 2020). Our results are shown in Figure 8, where the observed seismicity is compared with the forecast values. The solid lines in Figure 8 show the magnitude that has a 50 % chance of exceedance by the next record-breaking event, M_{50}, while the dashed lines show M_{95} and M_{05} (i.e., the magnitude that has a 95 % chance of being exceeded, and the magnitude that has a 5 % chance of being exceeded by the next record-breaking event).



(a) *(b)*

Figure 8: Application of the empirically constrained forecasting model to the Preston New Road PNR-2 (a) and Peace River (b) sequences. Observed events are marked with grey dots. The solid line marks M_{50}, while the dashed lines mark M_{05} and M_{95}. For PNR-2, the bursts of seismicity associated with each discrete hydraulic fracturing interval (Stages 1-7) are marked with grey arrows.

For the Peace River case, the forecast values are stable for the duration of the sequence. The *M* 5.6 event that occurs is close to the M_{05} value, indicating a 5 % likelihood of this magnitude being reached or exceeded.

For PNR-2, the M 2.8 event is well within the forecast range, and close to the M_{50} value at the time it occurred. Hydraulic fracturing at PNR-2 was conducted as a series of discrete injection stages, typically lasting between 1 – 2 hours, with only one injection stage taking place each day. Stage 7 was the last stage to have been stimulated, with the M 2.8 event occurring roughly 72 hours after this stage had been completed (Kettlety et al., 2021). The forecast values prior to Stage 7 are therefore of particular interest since these values could have informed the operational decision to perform this stage. At the time that injection of Stage 7 began, the likelihood of reaching or exceeding M 2.8 was 12 %. The forecasting model therefore provides a reasonable characterisation of the hazard at the time that the decision to proceed with Stage 7 was made.

Interestingly, the event that most exceeds the forecast is the M 1.9 event that followed Stage 6. At the start of injection of Stage 6, the likelihood of reaching or exceeding M 1.9 was only 1 %. Kettlety et al. (2021) identified that Stage 6 saw a significant change in geomechanical behaviour in the reservoir, with microseismicity beginning to occur along the fault structure that ultimately hosted the M 2.8 event. Kettlety et al. (2021) interpreted the microseismicity prior to Stage 6 as being associated with hydraulic fracture propagation (and the reactivation of some natural fracture networks), whereas microseismicity from Stage 6 onwards begins to represent the onset of reactivation of a critically stressed fault.

This highlights one of the challenges with induced seismicity forecasting – where a sudden change in the underlying geomechanical behaviour takes place, events from prior to this change may not be useful in forecasting subsequent behaviour. As described in our Methods, the M_{UL} and M_{JL} estimators assume that record-breaking magnitudes are sampled from a stationary underlying distribution. We note that this caveat also applies to other induced seismicity forecasting methods that assume constant scaling between injection rates and induced seismicity rates (e.g., Shapiro et al., 2010; Hallo et al., 2014; Mancini et al., 2021).

It is unclear the degree to which this assumption should be expected to hold for induced seismicity sequences. For WWD, injection rates are typically constant over years, creating a slow and steady pressure increase, such that a relatively constant underlying distribution of seismicity might be expected. However, Verdon et al. (2024) found evidence for accelerating rates of seismicity relative to injection volumes during the early stages of WWD-induced seismicity onset, which then stabilised at later times.

The successful performance of the M_{UL} and M_{JL} estimators in our study suggests that the assumption of stationarity is sufficiently satisfied, at least on the timescale of intervals between record-breaking events in these WWD-induced sequences. In contrast, for hydraulic fracturing at PNR-2, the microseismicity associated with hydraulic fracture propagation during the earlier stages does not do a good job of forecasting what happened as the larger fault began to reactivate. Once this fault reactivated, the forecasting model using the seismicity from this point onwards does a good job of forecasting the subsequent seismicity that developed.

These observations show that care should be taken to fully interpret and understand the geomechanical behaviours that can be manifested in microseismic event observations when using statistical models to forecast induced seismicity. It may be necessary to assess whether the underlying assumptions – such as stationarity and constancy of scaling between injection rate and seismicity rate – are reasonable in a particular case. These assumptions may not be appropriate in situations, such as at PNR-2, where a new fault structure is encountered by a growing injection pulse and begins to reactivate.

5.3. Time dependent forecasting

The forecasting methods developed here do not provide any estimate of whether a new record-breaking event will occur and, if so, when it will occur. The timing of the next record-breaking event could be estimated from the growing number of earthquakes within a sequence. The expected number of record-breaking events, N_{rb}, in a population of n events can be approximated, assuming that the events are independent and drawn from a constant underlying distribution, as (Arnold et al., 1998; Nevzorov, 2001):

$$N_{rb} \approx \ln(n) + 0.577215 \tag{6}$$

with the variance given by:

$$\mathrm{Var}(N_{rb}) = \ln(n) - 1.0677 \tag{7}$$

The number of record-breaking events relative to the total number of events within the sequence could therefore be used to indicate whether another record-breaking event might be imminent. Further investigation of this possibility is clearly merited.

Perhaps more importantly, the methods developed here, which are based on the concept of record-breaking events, imply that M_{MAX} for a sequence of induced seismicity will be ever-increasing, unless and until clear evidence of an upper truncation to the G-R distribution emerges. In practice, many sequences of induced seismicity generated by long-term injection have shown time-dependent behaviour where magnitudes increased during the first years of injection, but then stabilised and decreased over time (Rodríguez-Pradilla et al., 2022; Watkins et al., 2023; Verdon et al., 2024).

As sequences stabilise and abate, magnitude forecasts based on extreme value estimators will cease to be appropriate. Clearly, some means of estimating the point at which the rates and magnitudes of induced seismicity are no longer increasing is required. One method may be to compare the numbers of record-breaking events when the sequence is run forwards versus when the sequence is run in a time-reversed order (Mendecki, 2016). If the earthquake sequence is sampling from an underlying stationary distribution, then we would expect the same number of record-breaking events whether the sequence is run forwards or backwards. If there are significantly more record-breaking events when the sequence is run forwards, then this would imply that the hazard is increasing, while if there are significantly more record-breaking events when the sequence is run in reverse, then this would imply that the hazard is abating. Again, further investigation of this concept is clearly merited.

6. Conclusions

We have assessed the performance of induced seismicity forecasting models for M_{NRB} using methods based on extreme value estimators. These models can be implemented in a number of different ways, and we have quantitatively compared the performance of these implementations. We compiled a database of over 80 individual sequences of induced seismicity against which comparisons of model performance were made. We found that using all events within a catalog or just the record-breaking events made little difference to the forecasting results, since the models are primarily sensitive to the largest magnitude events in the sequence.

Estimates of M_{NRB} using the upper limit method with event magnitudes tended to overestimate the observed magnitudes. However, unlike other models, this model never significantly underpredicted the observed seismicity, so it has use in defining an upper estimate for M_{NRB}. The models which used earthquake potency instead of magnitude produced the closest overall fit to the observed magnitudes, but on occasion did produce significant underestimates of the observed magnitudes. The potency-based models seldom produced overpredictions of the observed magnitudes.

Based on these observations, we conclude that the upper limit magnitude-based model and the jump-limited potency-based models can be combined to give upper and lower estimators for the upcoming events within an induced seismicity sequence. We found that most of the observed events were much closer to the lower magnitude estimator. We used these observations to define an empirically constrained probability distribution for expected magnitudes relative to the upper and lower estimators. This distribution was consistent between the different populations of induced seismicity sequences compiled for our analysis, as well as for sequences that were generated synthetically.

We applied this forecasting approach to two out-of-training-sample (i.e., not used in defining our empirically constrained distribution) sequences of induced seismicity. We find that in both cases our modelling approach does a good job of characterising the induced seismicity that occurred. However, the example from PNR-2 again highlights one of the major challenges in forecasting induced seismicity: where rapid changes in the underlying geomechanical processes occur (such as a when a different fault begins to be perturbed), seismicity from earlier within the sequence may not be useful for forecasting once this change has occurred.

Data and Resources

The earthquake catalog for Oklahoma was sourced from Park et al. (2022), where the catalog is provided as a digital supplement. The earthquake catalog for Texas was sourced from the TexNet database at https://www.beg.utexas.edu/texnet-cisr/texnet/earthquake-catalog (last accessed 14/02/2024). The earthquake catalogs for the sequences described by Watkins et al. (2023) are available as a digital supplement to that paper. The catalog for PNR-2 is available from the UK National Geoscience Data Centre at https://webapps.bgs.ac.uk/services/ngdc/accessions/index.html#item173104 (last accessed 14/02/2024). The catalog for the Peace River sequence was sourced from the Alberta Earthquake Dashboard at https://ags-aer.shinyapps.io/Seismicity_waveform_app/ (last accessed 14/02/2024).

Acknowledgements

JPV's contribution to this research was funded by the NERC SeisGreen grant (Grant Number NE/W009293/1). We are grateful to Dr Robert Shcherbakov and an anonymous reviewer for their thorough and constructive reviews of our original manuscript.

References

AGS, 2020. Alberta Earthquake Dashboard: Alberta Geological Survey. Available at: https://ags.aer.ca/publication/iam-005 (last accessed 14/02/2024).

Arnold, B.C., N. Balakrishnan, H.N. Nagaraja, 1998. Records. John Wiley and Sons, New York.

BEIS, 2019. Government ends support for fracking: Press Release from the UK Department of Business, Energy and Industrial Strategy. Available at: https://www.gov.uk/government/news/government-ends-support-for-fracking (last accessed 15/08/2023.

Block, L.V., C.K. Wood, W.L. Yeck, V.M. King, 2014. The 24 January 2013 M_L 4.4 earthquake near Paradox, Colorado, and its relation to deep well injection: Seismological Research Letters 85, 609-624.

Bommer, J.J., 2022. Earthquake hazard and risk analysis for natural and induced seismicity: towards objective assessments in the face of uncertainty: Bulletin of Earthquake Engineering 20, 2825-3069.

Campbell, N.M., M. Leon-Corwin, L.A. Ritchie, J. Vickery, 2020. Human-induced seismicity: risk perceptions in the state of Oklahoma: The Extractive Industries and Society 7, 119-126.

Cao, N-T., L. Eisner, Z. Jechumtálová, 2020. Next record-breaking magnitude for injection induced seismicity: First Break 38, 53-57.

Cao, N-T., L. Eisner, Z. Jechumtálová, J.P. Verdon, U. bin Waheed, 2024. Upper limit magnitudes for induced seismicity in energy industries: Geophysical Prospecting, *accepted*.

Cesca, S., D. Stich, F. Grigoli, A. Vuan, J.Á. López-Comino, P. Niemz, E. Blanch, T. Dahm, W.L. Ellsworth, 2021. Seismicity at the Castor gas reservoir driven by pore pressure diffusion and asperities loading: Nature Communications 12, 4783.

Clarke, H., J.P. Verdon, T. Kettlety, A.F. Baird, J-M. Kendall, 2019. Real time imaging, forecasting and management of human-induced seismicity at Preston New Road, Lancashire, England: Seismological Research Letters 90, 1902-1915.

Clauset, A., C.R. Shalizi, M.E.J. Newman, 2009. Power-law distributions in empirical data: Society of Industrial and Applied Mathematics Reviews 51, 661-703.

Cooke, P., 1979. Statistical inference for bounds of random variables: Biometrica 66, 367-374.

Davis, S.D., and W.D. Pennington, 1989. Induced seismic deformation in the Cogdell oil field of west Texas: Bulletin of the Seismological Society of America 79,1477-1494.

Dempsey, D. and J. Suckale, 2017. Physics-based forecasting of induced seismicity at Groningen gas field, the Netherlands: Geophysical Research Letters 44, 7773-7782.

Eaton, D.W., and N. Igonin, 2018. What controls the maximum magnitude of injection-induced earthquakes: The Leading Edge 37, 135-140.

Eaton, D.W., N. Igonin, A. Poulin, R. Weir, H. Zhang, S. Pellegrino, G. Rodriguez, 2018. Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors: Seismological Research Letters 89, 1641-1651.

Evensen, D., A. Varley, L. Whitmarsh, P. Devine-Wright, J. Dickie, P. Bartie, H. Napier, I. Mosca, C. Foad, S. Ryder, 2022. Effect of linguistic framing and information provision on attitudes towards induced seismicity and seismicity regulation: Scientific Reports 12, 11239.

Frohlich, C., W. Ellsworth, W.A. Brown, M. Brunt, J. Luetgert, T. MacDonald, S. Walter, 2014. The 17 May 2012 M 4.8 earthquake near Timpson, East Texas: An event possibly triggered by fluid injection: Journal of Geophysical Research 119, 581-593.

Grigoratos, I., A. Savvaidis, E. Rathje, 2022. Distinguishing the causal factors of induced seismicity in the Delaware basin: hydraulic fracturing or wastewater disposal: Seismological Research Letters 93, 2640-2658.

Gutenberg, B., and C.F. Richter, 1944. Frequency of earthquakes in California: Bulletin of the Seismological Society of America 34, 185-188.

Hallo, M., I. Oprsal, L. Eisner, M.Y. Ali, 2014. Prediction of magnitude of the largest potentially induced seismic event: Journal of Seismology 18, 421-431.

Holland, A.A., 2013. Earthquakes Triggered by Hydraulic Fracturing in South-Central Oklahoma: Bulletin of the Seismological Society of America 103, 1784-1792.

Hennings, P.H., and M.H. Young, 2023. The TexNet-CISR collaboration and steps toward understanding induced seismicity in Texas: in R.C. Buchanan, M.H. Young, K.E. Murray eds., Recent Seismicity in the Southern Midcontinent, USA: Scientific, Regulatory, and Industry Responses: Geological Society of America Special Paper 559, 53-71.

Hennings, P.H., J.P. Nicot, R.S. Gao, H.R. DeShon, J.E. Lund Snee, A.P. Morris, M.R. Brudzinski, E.A. Horne, C. Breton, 2021. Pore pressure threshold and fault slip potential for induced earthquakes in the Dallas-Fort Worth area of north central Texas: Geophysical Research Letters 48, e2021GL093564.

Holschneider, M., G. Zöller, S. Hainzl, 2011. Estimation of the maximum possible magnitude in the framework of a doubly truncated Gutenberg-Richter model: Bulletin of the Seismological Society of America 101, 1649-1659.

Horton, S., 2012. Injection into subsurface aquifers triggers earthquake swarm in central Arkansas with potential for damaging earthquake: Seismological Research Letters 83, 250-260.

Horner, R.B., J.E. Barclay, J.M. MacRae, 1994. Earthquakes and hydrocarbon production in the Fort St. John area of northeastern British Columbia: Canadian Journal of Exploration Geophysics 30, 39-50.

Hosseini, B.K., and D.W. Eaton, 2018. Fluid flow and thermal modeling for tracking induced seismicity near the Graham disposal well, British Columbia (Canada): SEG 88th Annual Conference, Anaheim CA, Expanded Abstracts 4987-4991.

Keranen, K.M., H.M. Savage, G.A. Abers, E.S. Cochran, 2013. Potentially induced earthquakes in Oklahoma: USA: links between wastewater injection and the 2011 M_W 5.7 earthquake sequence: Geology 41, 699-702.

Kettlety, T., and A. Butcher, 2022. Local and moment magnitudes of Preston New Road seismicity, 2018-2019: National Geological Data Centre, DOI: 10.5285/709cbc2f-af5c-4d09-a4ea-6deb5aa8c5d8.

Kettlety, T., J.P. Verdon, A. Butcher, M. Hampson, L. Craddock, 2021. High-resolution imaging of the M_L 2.9 August 2019 earthquake in Lancashire, United Kingdom, induced by hydraulic fracturing during Preston New Road PNR-2 operations: Seismological Research Letters 92, 151-169.

Kijko, A., 2004. Estimation of the maximum earthquake magnitude, m_{max}: Pure and Applied Geophysics 161, 1655-1681.

Kijko, A., and M.A. Sellevoll, 1989. Estimation of earthquake hazard parameters from incomplete data files. Part I: Utilization of extreme and complete catalogs with different threshold magnitudes: Bulletin of the Seismological Society of America 79, 645-654.

Kim, W.-Y., 2013. Induced seismicity associated with fluid injection into a deep well in Youngstown, Ohio: Journal of Geophysical Research 118, 3506-3518.

Kwiatek, G., T. Saamo, T. Ader, F. Bluemle, M. Bohnhoff, M. Chendorain, G. Dresen, P. Heikkinen, I. Kukkonen, P. Leary, M. Leonhardt, P. Malin, P. Martinez-Garzon, K. Passmore, P. Passmore, S. Valenzuela, C. Wollin, 2019. Controlling fluid-induced seismicity during a 6.1-km-deep geothermal stimulation in Finland: Science Advances 5, eaav7224.

Lee, K.K., W.L. Ellsworth, D. Giardini, J. Townend, S. Ge, T. Shimamoto, I-W. Yeo, T-S. Kang, J. Rhie, D-H. Sheen, C. Chang, J-U. Woo, C. Langenbruch, 2019. Managing injection-induced seismic risks: Science 364, 730-732.

Lei, X., Z. Wang, J. Su, 2019. The December 2018 M_L 5.7 and January 2019 M_L 5.3 earthquakes in south Sichuan Basin induced by shale gas hydraulic fracturing: Seismological Research Letters 90, 1099-1110.

Li, T., Y.J. Gu, J. Wang, R. Wang, J. Yusifbayov, M. Reyes Canales, T. Shipman, 2022. Earthquakes induced by wastewater disposal near Musreau Lake, Alberta, 2018-2020: Seismological Research Letters 93, 727-738.

Mancini, S., M.J. Werner, M. Segou, B. Baptie, 2021. Probabilistic forecasting of hydraulic fracturing-induced seismicity using an injection-rate driven ETAS model: Seismological Research Letters 92, 3471-3481.

McGarr, A., 1976, Seismic moments and volume changes: Journal of Geophysical Research 81, 1487-1494.

Mendecki, A.J., 2016. Mine Seismology Reference Book. Institute of Mine Seismology, Somerset West, South Africa.

Molina, I., J.S. Velásquez, J.L. Rubenstein, A. Garcia-Aristizabal, V. Dionicio, 2020. Seismicity induced by massive wastewater injection near Puerto Gaitán, Colombia: Geophysical Journal International 223, 777-781.

Mueller C.S., 2010 The influence of maximum magnitude on seismic-hazard estimates in the central and eastern United States: Bulletin of the Seismological Society of America 100, 699-711.

Nakai, J.S., M. Weingarten, A.F. Sheehan, S.L. Bilek, S. Ge, 2017. A possible causative mechanism of Raton Basin, New Mexico and Colorado earthquakes using recent seismicity patterns and pore pressure modelling: Journal of Geophysical Research 122, 8051-8065.

Nantanoi, S., G. Rodríguez-Pradilla, J.P. Verdon, 2022. 3D-seismic interpretation and fault slip potential analysis from hydraulic fracturing in the Bowland Shale, UK: Petroleum Geoscience 28, petgeo2021-057.

Nevzorov, V.B., 2001. Records: Mathematical Theory. American Mathematical Society, Providence, Rhode Island.

Park, Y., G.C. Beroza, W.L. Ellsworth, 2022. Basement fault activation before larger earthquakes in Oklahoma and Kansas: The Seismic Record 2, 197-206.

Pisarenko, V.F., A.A. Lyubushin, V.B. Lysenko, T.V. Golubeva, 1996. Statistical estimation of seismic hazard parameters: Maximum possible magnitude and related parameters: Bulletin of the Seismological Society of America 86, 691-700.

Rodríguez-Pradilla, G., D.W. Eaton, J.P. Verdon, 2022. Basin-scale multi-decadal analysis of hydraulic fracturing and seismicity in western Canada shows non-recurrence of induced runaway fault-rupture: Scientific Reports 12, 14463.

Rubenstein, J.L., and A.B. Mahani, 2015. Myths and facts on wastewater injection, hydraulic fracturing, enhanced oil recovery, and induced seismicity: Seismological Research Letters 86, 1060-1067.

Rutqvist, J., A.P. Rinaldi, F. Cappa, G.J. Moridis, 2013. Modeling of fault reactivation and induced seismicity during hydraulic fracturing of shale-gas reservoirs: Journal of Petroleum Science and Engineering 107, 31-44.

Salvage, R.O., D.W. Eaton, C.M. Furlong, J. Dettmer, P.K. Pedersen, 2024. Induced or natural? Toward rapid expert assessment, with application to the M_W 5.2 Peace River earthquake sequence: Seismological Research Letters 95, 758-772.

Savvaidis, A., B. Young, G.D. Huang, A. Lomax, 2019. TexNet: A statewide seismological network in Texas: Seismological Research Letters 90, 1702-1715.

Schultz, R., Y. Park, A. Leonardo, A. Suarez, W.L. Ellsworth, G.C. Beroza, 2023a. En-echelon faults reactivated by wastewater disposal near Musreau Lake, Alberta: Geophysical Journal International 235, 417-429.

Schultz, R., J.-U. Woo, K. Pepin, W.L. Ellsworth, H. Zebkar, P. Segall, Y.J. Gu, S. Samsonov, 2023b. Disposal from in situ bitumen recovery induced the M_L 5.6 Peace River earthquake: Geophysical Research Letters 50, e2023GL102940.

Shapiro, S.A., C. Dinske, C. Langenbruch, F. Wenzel, 2010. Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations: The Leading Edge 29, 304-308.

Skoumal, R.J., R. Ries, M.R. Brudzinski, A.J. Barbour, B.S. Currie, 2018. Earthquakes induced by hydraulic fracturing are pervasive in Oklahoma: Journal of Geophysical Research 123, 10918-10935.

Skoumal, R.J., A.J. Barbour, M.R. Brudzinski, T. Langenkamp, J.O. Kaven, 2020. Induced seismicity in the Delaware Basin, Texas: Journal of Geophysical Research 125, e2019JB018558.

Skoumal, R.J., J.O. Kaven, A.J. Barbour, C. Wicks, M.R. Brudzinski, E.S. Cochran, J.L. Rubinstein, 2021. The induced Mw 5.0 March 2020 West Texas seismic sequence: Journal of Geophysical Research 126, e2020JB020693.

van der Elst, N.J., M.T. Page, D.A. Weiser, T.H.W. Goebel, S.M. Hosseini, 2016. Induced earthquake magnitudes are as large as (statistically) expected: Journal of Geophysical Research 121, 4575-4590.

Verdon, J.P., 2016. Using microseismic data recorded at the Weyburn CCS-EOR site to assess the likelihood of induced seismic activity: International Journal of Greenhouse Gas Control 54, 421-428.

Verdon, J.P. and J. Budge, 2018. Examining the capability of statistical models to mitigate induced seismicity during hydraulic fracturing of shale gas reservoirs: Bulletin of the Seismological Society of America 108, 690-701.

Verdon, J.P. and J.J. Bommer, 2021. Green, yellow, red, or out of the blue? An assessment of Traffic Light Schemes to mitigate the impact of hydraulic fracturing-induced seismicity: Journal of Seismology 25, 301-326.

Verdon, J.P. and G. Rodríguez-Pradilla, 2023. Assessing the variability in hydraulic fracturing-induced seismicity occurrence between North American shale plays: Tectonophysics 859, 229898.

Verdon, J.P., A.L. Stork, R.C. Bissell, C.E. Bond, M.J. Werner, 2015. Simulation of seismic events induced by CO_2 injection at In Salah, Algeria: Earth and Planetary Science Letters 426, 118-129.

Verdon, J.P., J-M. Kendall, A. Butcher, R. Luckett, B.J. Baptie, 2018. Seismicity induced by longwall coal mining at the Thoresby Colliery, Nottinghamshire, U.K.: Geophysical Journal International 212, 942-954.

Verdon, J.P., B. Pullen, G. Rodríguez-Pradilla, 2024. Growth and stabilization of induced seismicity rates during long-term, low pressure fluid injection: Proceedings of the Royal Society Transactions A, *in press.* DOI: 10.1098/rsta.2023-0183.

Wang, Z., X. Lei, S. Ma, X. Wang, Y. Wan, 2020. Induced earthquakes before and after cessation of long-term injections in Rongchang gas field: Geophysical Research Letters 47, e2020GL089569.

Watkins, T.J.M., J.P. Verdon, G. Rodríguez-Pradilla, 2023. The temporal evolution of induced seismicity sequences generated by long-term, low pressure fluid injection: Journal of Seismology 27, 243-259.

Weingarten, M., S. Ge, J.W. Godt, B.A. Bekins, J.L. Rubinstein, 2015. High-rate injection is associated with the increase in U.S. mid-continent seismicity: Science 348, 1336-1340.

Yeck, W.L., A.F. Sheehan, H.M. Benz, M. Weingarten, J. Nakai, 2016. Rapid response, monitoring and mitigation of induced seismicity near Greeley, Colorado: Seismological Research Letters 87, 837-847.

Yeck, W.L., G.P. Hayes, D.E. McNamara, J.L. Rubinstein, W.D. Barnhart, P.S. Earle, H.M. Benz, 2017. Oklahoma experiences largest earthquake during ongoing regional wastewater injection hazard mitigation efforts: Geophysical Research Letters 44, 711-717.

Zhu, W., and G.C. Beroza, 2019. PhaseNet: a deep-neural-network-based seismic arrival-time picking method: Geophysical Journal International 216, 261-273.

Zöller, G. and M. Holschneider, 2016. The maximum possible and the maximum expected earthquake magnitude for production-induced earthquakes at the gas field in Groningen, the Netherlands: Bulletin of the Seismological Society of America 106, 2917-2921.