Geophysical Journal International

Geophys. J. Int. (2011) **186,** 793–807 doi: 10.1111/j.1365-246X.2011.05068.x

Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems

C. E. Bachmann,[1] S. Wiemer,[1] J. Woessner[1] and S. Hainzl[2]

[1]*Swiss Seismological Service, ETH Zurich, Sonneggstrasse* 5, 8092 *Zurich, Switzerland. E-mail: bachmann@sed.ethz.ch*
[2]*GFZ German Research Centre for Geosciences, Telegrafenberg,* 14473 *Potsdam, Germany*

Accepted 2011 May 4. Received 2011 May 4; in original form 2010 September 30

SUMMARY

Geothermal energy is becoming an important clean energy source, however, the stimulation of a reservoir for an Enhanced Geothermal System (EGS) is associated with seismic risk due to induced seismicity. Seismicity occurring due to the water injection at depth have to be well recorded and monitored. To mitigate the seismic risk of a damaging event, an appropriate alarm system needs to be in place for each individual experiment. In recent experiments, the so-called traffic-light alarm system, based on public response, local magnitude and peak ground velocity, was used. We aim to improve the pre-defined alarm system by introducing a probability-based approach; we retrospectively model the ongoing seismicity in real time with multiple statistical forecast models and then translate the forecast to seismic hazard in terms of probabilities of exceeding a ground motion intensity level. One class of models accounts for the water injection rate, the main parameter that can be controlled by the operators during an experiment. By translating the models into time-varying probabilities of exceeding various intensity levels, we provide tools which are well understood by the decision makers and can be used to determine thresholds non-exceedance during a reservoir stimulation; this, however, remains an entrepreneurial or political decision of the responsible project coordinators. We introduce forecast models based on the data set of an EGS experiment in the city of Basel. Between 2006 December 2 and 8, approximately 11 500 m^3 of water was injected into a 5-km-deep well at high pressures. A six-sensor borehole array, was installed by the company Geothermal Explorers Limited (GEL) at depths between 300 and 2700 m around the well to monitor the induced seismicity. The network recorded approximately 11 200 events during the injection phase, more than 3500 of which were located. With the traffic-light system, actions where implemented after an M_L 2.7 event, the water injection was reduced and then stopped after another M_L 2.5 event. A few hours later, an earthquake with M_L 3.4, felt within the city, occurred, which led to bleed-off of the well. A risk study was later issued with the outcome that the experiment could not be resumed. We analyse the statistical features of the sequence and show that the sequence is well modelled with the Omori–Utsu law following the termination of water injection. Based on this model, the sequence will last 31+29/−14 years to reach the background level. We introduce statistical models based on Reasenberg and Jones and Epidemic Type Aftershock Sequence (ETAS) models, commonly used to model aftershock sequences. We compare and test different model setups to simulate the sequences, varying the number of fixed and free parameters. For one class of the ETAS models, we account for the flow rate at the injection borehole. We test the models against the observed data with standard likelihood tests and find the ETAS model accounting for the on flow rate to perform best. Such a model may in future serve as a valuable tool for designing probabilistic alarm systems for EGS experiments.

Key words: Probabilistic forecasting; Earthquake interaction, forecasting, and prediction; Statistical seismology.

INTRODUCTION

Enhanced Geothermal Systems (EGSs) represent an attractive source of alternative energy with a low carbon footprint and few environmental concerns. EGSs are commonly known as 'hot-dry-rock' or 'hot-fractured-rock' technique and refer to a technology that uses hydraulic stimulation of a hot ($T > 100\,^{\circ}$C) but comparably impermeable ($\kappa < 10^{-16}$ m^2) rock mass at depth ($Z > 3$ km) to create an artificial geothermal reservoir. These systems are most economically viable in areas with a steep geothermal gradient because drilling costs increase exponentially with depth. Proximity to end users of heat and electricity gained from depth is desirable to minimize energy loss through distance.

In a 'hot-dry-rock' experiment, the rock mass is stimulated hydraulically by pumping fluids through an injection well under high pressure into the target area at depth (e.g. Smith 1983; Tenzer 2001). The increase in pore pressure as the fluid propagates away from the injection well fractures the host rock locally. Generating fractures in the target rock mass simultaneously causes microseismicity through the fracturing process, defining the paths of the fluids to flow through and heat up. The spatial distribution of the micro-earthquakes provides important clues about the volume and orientation of the fractured rock at depth. Highly sensitive seismic monitoring techniques are routinely applied at EGS sites to map the spatial and temporal development of the stimulated volume and to characterize the geothermal reservoir (e.g. Wohlenberg & Keppler 1987; Haering *et al.* 2008). Once a sufficiently large reservoir (volume >1 km^3) has developed, a second well is typically drilled into the stimulated volume. Water then flows between the two wells; hot water is extracted from the production borehole and engineered to an energy resource. Artificially creating fractures in rock is a necessary component of an EGS; this process bears the risk of producing not only micro-earthquakes but also possibly moderate-to-large magnitude earthquakes that could cause damage (Giardini 2009; Kraft *et al.* 2009).

One of the first purely commercially oriented EGS projects was initiated in Basel, Switzerland, in 1996 by an industry consortium (GeoPower Basel) (Haering *et al.* 2008). The consortium led by Geothermal Explorers Limited (GEL) was well aware of the possibility of inducing earthquakes strong enough to be felt. To monitor earthquake activity and be prepared for hazard and risk mitigation actions, GEL adapted a 'traffic-light' system first proposed by Bommer *et al.* (2006) for the 'Berlin' geothermal project in El Salvador. The four stage traffic-light system was based on three components: (1) Public response, (2) observed local magnitude and (3) peak ground velocity (PGV). Haering *et al.* (2008) explain the different stages in detail; we only provide an abbreviated version here. According to the three components, the injection of fluids would be (1) continued as planned (green), (2) continued but not increased (yellow), (3) stopped and bleed-off stimulation pressure started (orange) or (4) stopped and bleed-off to minimum wellhead pressure started (red), where bleed-off implies to actively pump water out of the borehole. The traffic-light system is defined *ad hoc* and mainly based on expert judgment; however, it was the one single system that had been used in similar projects before (Bommer *et al.* 2006) and others did not exist at the time.

The fluid injection started on 2006 December 2. Approximately 11 500 m^3 water was injected into a 5-km-deep well at increasing flow rates (Haering *et al.* 2008) to stimulate the reservoir. The seismicity was monitored by a six-sensor borehole array at depths between 300 m and 2700 m around the injection well (GEL) and by up to 30 seismic surface stations in the Basel area (SED) (Fig. 1).

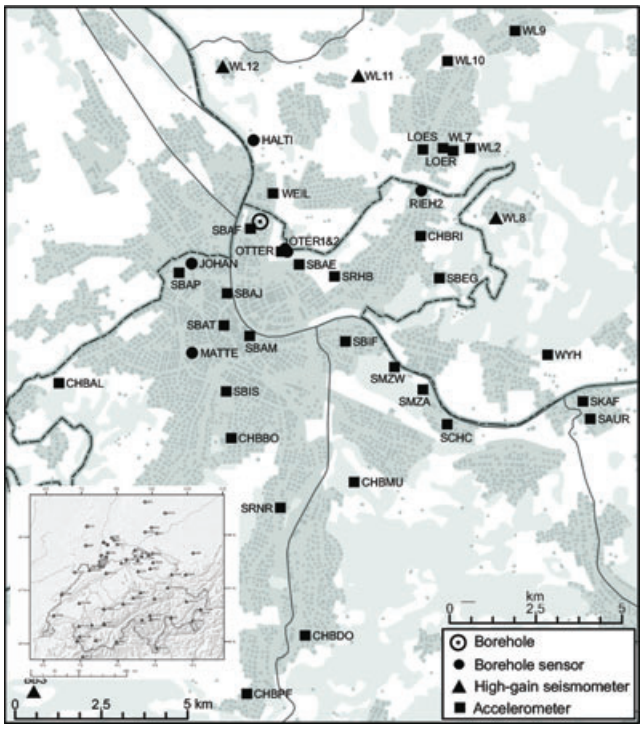

Figure 1. Overview of the study region with all seismic stations. Different symbols show borehole and strong motion stations maintained by either Geothermal Explorers or the Swiss Seismological Service. The inset indicates the location of all seismic stations in Switzerland with the high density of stations around Basel.

More than 11 200 events were recorded, more than 3500 of which were located by GEL (Fig. 2). The gradual increase in flow rate and wellhead pressure was accompanied by a steady increase in seismicity, both in terms of event rates and magnitudes (Fig. 3). In the early morning hours of December 8, after water had been injected at maximum rates in excess of 50 l s^{-1} and at wellhead pressures of up to 29.6 MPa for about 16 hr (Haering *et al.* 2008), a magnitude $M_{\rm L}$ 2.6 event occurred within the reservoir. This triggered the 'orange' alarm level, so that the injection pressure was reduced around 2006 December 8, 4:00 a.m., and fully stopped on the same day at 11:33 a.m (Haering *et al.* 2008). However, an $M_{\rm L}$ 3.4 event occurred 5 hr later, widely felt within the city of Basel. Slight non-structural damage, such as fine cracks in plaster, corresponding to an intensity of V on the European Macroseismic Scale (EMS98), has been claimed by many homeowners, with a damage sum, estimated and to a large extent already paid by insurance, of US $7 million (Kraft *et al.* 2009). About 1 hr after the $M_{\rm L}$ 3.4 event, bleed-off was initiated by opening the injection well, and hydrostatic downhole pressure was reached within 4 d. Following this, the seismicity slowly decayed. However, three additional felt earthquakes with $M_{\rm L} > 3$ occurred 1–2 months after bleed-off. More than 3 yr later, sporadic seismicity inside the stimulated rock volume is still being detected by the downhole instruments. The EGS project was on hold for more than 2 yr, awaiting the completion of an independent risk analysis study. The results of this study, completed in 2009 November (Baisch *et al.* 2009a), suggest that the risk of further felt and potentially damaging events is substantial. Public authorities thus decided that the project cannot be continued.

This well-monitored induced seismic sequence provides an excellent opportunity to improve the understanding of the physics

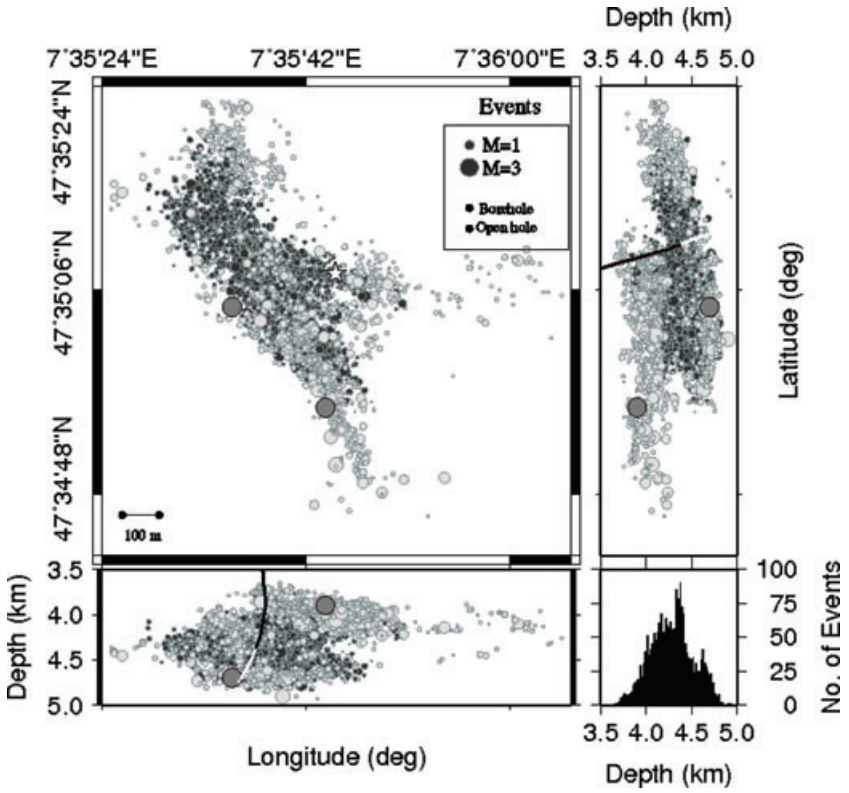

Figure 2. Distribution of the events in plane view (top left panel) and as depth distributions (EW lower panel and NS right panel). Circle sizes are scaled by magnitude; events with magnitudes above 3 are highlighted with darker colours. Events in black occurred during the injection and events in grey after water injection was terminated. The borehole is indicated; the darker part is cased and the lighter part is open.

of EGS. The Swiss Seismological Service (SED) at ETH Zurich, for example, is investigating the Basel data set in the framework of the multidisciplinary research project GEOTHERM (www.geotherm.ethz.ch) of which this study is a part. Several results have already been published (e.g. Deichmann & Ernst 2009; Deichmann & Giardini 2009; Kraft *et al.* 2009; Ripperger *et al.* 2009) and additional studies currently being conducted to be published.

The pressure reduction and eventual bleed-off of the system during the critical days around 2006 December 8 was consistent with the actions stipulated in the traffic-light systems; however, the public outcry, ongoing legal actions and ultimate termination of the Basel EGS project highlights clearly that the traffic-light system was, at least in this case, not a sufficient monitoring and alerting approach. A major goal of this study is to develop and test alternative probability-based statistical approaches to the traffic-light system. To achieve this goal, we first need to evaluate the performance of available statistical models that suitably describe and forecast features of induced seismicity and are readily able to forecast future seismic activity.

To model induced seismicity, a variety of approaches have been proposed in the past. Primarily physics-based approaches (Kohl & Megel 2007; Shapiro *et al.* 2007; Shapiro & Dinske 2009; Baisch *et al.* 2009c) suggest a hydro-mechanical model of fluid migration through the rock matrix and cracks. Events in such a model are triggered directly by shear failure of favourably oriented natural joints as a response of normal stress reduction due to high-pressure fluid injection (e.g. Baisch *et al.* 2009b,c). Such an approach also forms one of the logic tree branches of the Basel risk study and has been applied to other EGS projects, such as Soultz-sous-Forêts (Baisch *et al.* 2009b) or Cooper Basin (Baisch *et al.* 2009c). In this study, however, we focus on statistical approaches that describe the time-dependent seismic hazard as a combination of empirical observation and statistical modelling of the observed seismicity. A model in this sense defines the total number and frequency–magnitude distribution of the future seismicity in a given time window. The aim of the models is to best forecast the seismicity but not necessarily to understand the detailed physics of the ongoing processes. In a statistical framework, the seismic sequence triggered by fluid injection can be understood and described as a point process, as for any other sequence of clustered earthquakes. In addition to fluid-triggered events, each event potentially triggers 'daughter' earthquakes by the static and dynamic stress changes induced by their rupture. The overall seismicity can be described as a cascading, or epidemic, process, responding to an external forcing function.

The advantage of using statistical models to describe the seismicity during EGS stimulation is that the models are comparatively simple and well established in statistical seismology (e.g. Ogata 1988; Hainzl & Ogata 2005; Gerstenberger *et al.* 2005; Woessner *et al.* 2010). We show in this paper that the entire Basel sequence can in fact be well described using statistical models. We argue that such models should be considered as a starting point or reference model for any assessment of the time-dependent seismic hazard.

We adopt and compare two model frameworks widely used in the domain of time-varying earthquake forecasting on timescales of hours and days:

(1) The Reasenberg & Jones model, which is the basis of the Short Term Earthquake Probabilities (STEP) model (Gerstenberger

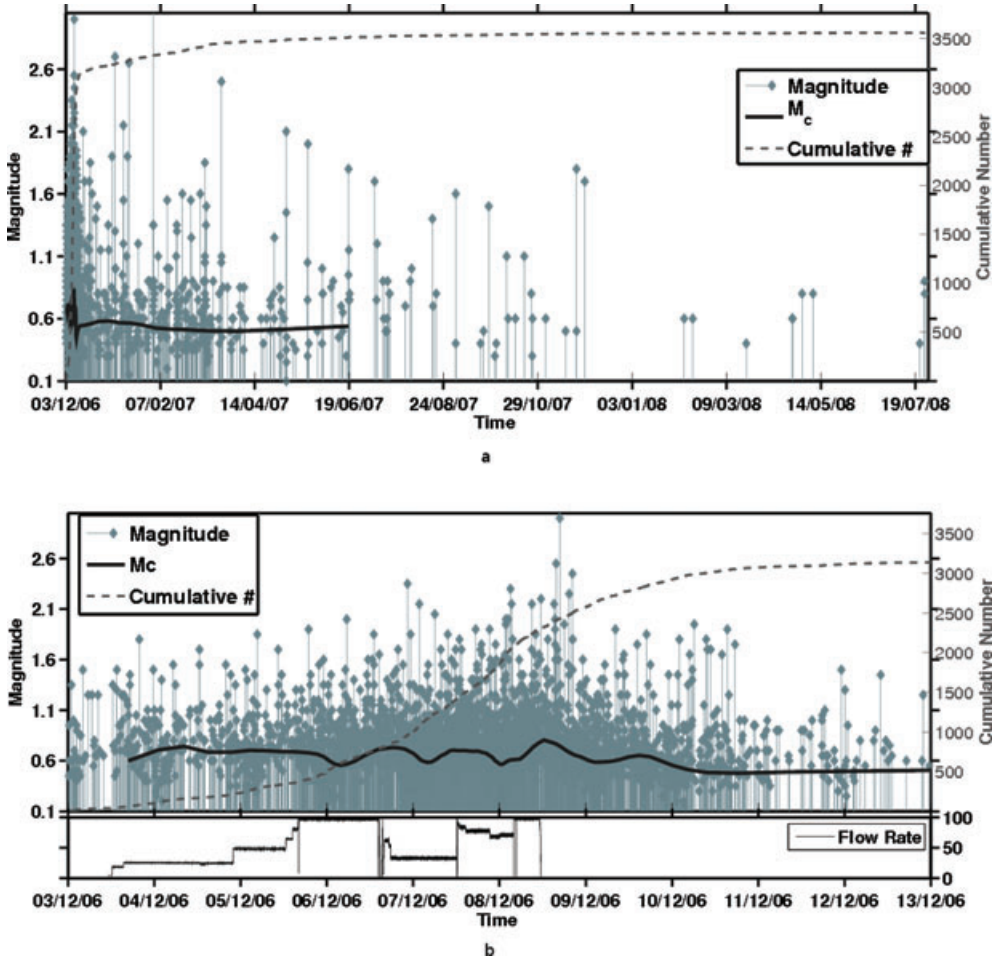

Figure 3. Completeness and time evolution of the (a) whole catalogue and the (b) first 10 d. The black solid line indicates the variation of the completeness of this catalogue over time; it varies the most during the first days and then becomes constant. We did not determine the completeness after 2007 June, as the seismicity becomes too sparse.

et al. 2005, 2007; Woessner *et al.* 2011). STEP-based hazard maps were first implemented for time-dependent hazard forecasts in California and are available online at the United States Geological Survey (USGS) (http://earthquake.usgs.gov/eqcenter/step/).

(2) The Epidemic Type Aftershock Sequence (ETAS) model (Ogata 1988; Hainzl & Ogata 2005). Based on ETAS, we also develop a model that takes the time-dependent pumping rate as an external forcing term into account.

None of these models has so far been applied to EGS-related seismicity but they are well established and tested in retrospective and in fully prospective forecasting experiments at regional scale (www.cseptesting.org; Schorlemmer *et al.* 2010; Woessner *et al.* 2011), as well as on the scale of aftershocks sequences (Hainzl *et al.* 2009; Cocco *et al.* 2010; Woessner *et al.* 2011).

We compare the forecasting ability of a total of eight models with standard likelihood testing approaches introduced by the Collaboratory for the Study of Earthquake Predictability (CSEP; Zechar *et al.* 2010; Schorlemmer *et al.* 2010; Woessner *et al.* 2011). We propose that a quantitative approach to model testing and evaluation is valuable as the ultimate goal is to establish models that can be used for regulatory guidelines.

DATA

The seismicity analysed in this study was recorded by six permanent downhole borehole geophones, operated by GEL. Four of the geophones were installed at intermediate depths between 300 and 600 m below the surface, one was at 1180 m and the deepest one at 2740 m. Prior analysis of this network led to the conclusion that two geophones at intermediate depth had a minor influence on the resolution and therefore they were not routinely processed (Dyer *et al.* 2008). Fig. 1 indicates the location of the stations involved in the data acquisition. In addition to the borehole network, the SED maintained a dense network of seismometers and accelerometers; their locations are also shown in Fig. 1.

The GEL network recorded over 11 200 events from 2006 December to 2008 July; 3500 had a good-quality signal and were located. Two magnitudes were provided: (1) local earthquake magnitudes (M_L) were provided by the SED for the largest ∼190 earthquakes only; and (2) a moment magnitude (M_W) was determined by the microseismic network of GEL (Dyer *et al.* 2008). Since the moment magnitudes are available for all 3500 events, we will be using them throughout this report. Bethmann *et al.* (2007) showed that the moment magnitudes provided by GEL are comparable to the moment magnitudes calculated by the SED. Fig. 2 indicates the locations

of all events located by GEL; earliest events occurred around the casing shoe and then migrated away from the opening, while latest events occurred to the east of the borehole. Dyer *et al.* (2008) and Haering *et al.* (2008) describe the evolution of seismicity in more detail.

Fig. 3 shows the time evolution of the events; the top panel shows the sequence until 2008 July 25 and the middle left panel a close-up from 2006 December 2 to December 12. Grey stems indicate single events, with magnitudes ranging from M_W 0.1 to M_W 3.1, where the M_W 3.1 event is equal to the earlier mentioned M_L 3.4 event which occurred at 2006 December 8, 4:48 p.m. Three additional events with magnitudes above three occurred in 2007 January and February. The lowest panel in Fig. 3 includes the applied flow rate. The injected water was increased step wise and reached a maximum of about 100 l s^{-1} before it was reduced. Dyer *et al.* (2008) describe the water injection and the pressure evolution in more detail.

METHOD

Fitting the overall parameters of the sequence: *b*-values and duration

In a first step, we analyse the monitoring completeness and bulk statistical parameters of the sequence. Of specific interest in this context is an estimate of the time required for the seismicity to return to the background rate. This estimate has, possibly, an impact on the continued need to monitor the site.

To estimate seismicity parameters, we first need to analyse the magnitude of completeness, (M_c), of the catalogue as a function of time. We apply the maximum curvature method (Wiemer 2000; Woessner & Wiemer 2005) and require 150 events to estimate the M_c for one sample. Most computations in this study are based on adaptations of the software package ZMAP (Wiemer 2001). The completeness estimates span a period starting about mid-day 2006 December 3 until 2007 June, after which the seismicity becomes too sparse (Fig. 3, as indicated by the black line). $M_c(t)$ varies between $0.5 \pm 0.07 \leq M_c \leq 0.9 \pm 0.11$ during the first 9 d of the time series (Fig. 3). These variations are most probably due to unpicked events during times of highest activity, because smaller events are hidden in the coda of larger ones. After a little more than 1 month, $M_c(t)$ remains constant at 0.5 ± 0.07. For the overall completeness of the entire sequence, we choose in the subsequent analysis a conservative value of 0.9, the maximum observed in any time period.

We then estimate activity rates and relative earthquake size distribution during the injection phase and the post-injection period. Using a maximum likelihood fit, we determine the *a*- and *b*-values of the Gutenberg–Richter (GR) law (Gutenberg & Richter 1942), which describes the relation between the frequency of earthquakes and the magnitudes of an event.

$$\log N = a - bM, \tag{1}$$

where N is the number of events with magnitudes larger or equal to M, a describes the productivity of the sequence and b the ratio of small to large events.

We find a substantial decrease of the *b*-value from 1.56 ± 0.05 to 1.15 ± 0.07 for co- and post-injection events, respectively, using the above-determined M_c of 0.9. All uncertainties are computed by bootstrapping the data set 100 times and fitting the parameters values to the bootstrap samples. The annual *a*-values for the same periods change from 6.08 to 3.31 (Fig. 4, colours as in Fig. 2).

Once the injection of water under high pressures stops, the seismicity decays gradually in the following weeks and months (Fig. 5), quite similar to any tectonic aftershock sequence. For aftershock sequences, the rate is usually well described by the Omori–Utsu law (Utsu 1961; Ogata 1999).

$$\lambda(t, M_c) = \frac{k(M_c)}{(t + c)^p}, \tag{2}$$

where t is the time elapsed after the main shock, c and p are empirical parameters, characteristic for a specific sequence and $k(M_c)$ is a function of the number of events with magnitudes above the completeness magnitude M_c.

We investigate whether the Omori–Utsu law can provide an acceptable fit to the data of the post-injection period. We fit the events with $M_w \geq 0.9$, that occurred between the end of the injection (December 8, at 11:33 a.m) and day 200 of the sequence. We find $p = 1.33 \pm 0.06$, $c = 0.38 \pm 0.061$ d and $k = 86.6 \pm 9.81$ (Fig. 5) as the mean parameters for 1000 bootstrap models. The two-sample Kolmogorov–Smirnov (Conover 1972; Woessner *et al.* 2004) test, testing whether the cumulative rate of the data and the fitted Omori–Utsu law belong to the same distribution, is not rejected at the significance level of 0.05. This indicates a good fit of the Omori–Utsu law to our data.

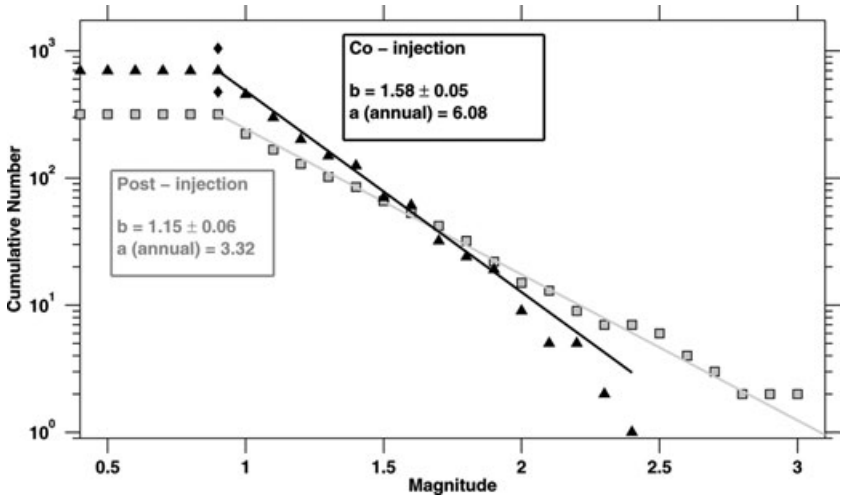

Figure 4. Gutenberg–Richter frequency–magnitude relation for two different sequences. Darker squares show events during the injection, lighter triangles mark events after the termination of water injection. Gutenberg-Richter parameters are indicated for each sequence.

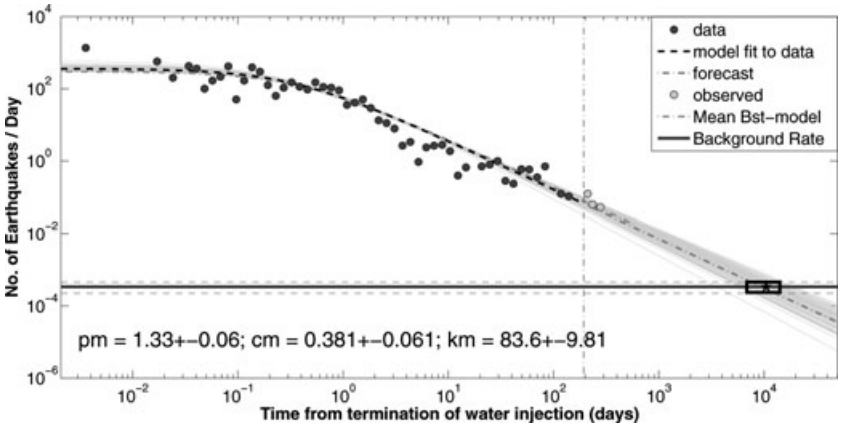

Figure 5. Decay of the sequence after the termination of the water injection. A modified Omori–Utsu law is fitted to the sequence to determine its duration; circles represent the data. We only fit the first 200 d of data (dark circles); events after this time (grey circles) fall within the uncertainty. The background of this region and the uncertainty are indicated at the bottom of the figure. Where they intercept with the model, we find the duration of the sequence. The black box indicates the uncertainty and the black star the best fit. We find a duration of 31 +29/−14 years.

To estimate the expected duration of the sequence, we additionally determine the background seismicity rate for this region. No event has ever been located within the small stimulated volume of about 1 km^3 during the 25-yr history of recording microseismicity in Switzerland. We therefore use as a proxy the seismic activity rate of the seismogenic source zone of Basel used in the determination of the Swiss Seismic hazard in 2004 (Giardini *et al.* 2004; Wiemer *et al.* 2009). The rate is normalized to the size of the stimulated volume (more specifically, to its 2-D areal extension). This assumption is consistent with the definition of a seismic source as a zone of equal seismic potential in seismic hazard assessment. The seismogenic source zone has an a-value of 2.31, a b-value of 0.9 and spans over an area of 1741 km^2. The area affected by the injection is 1.6 km^2 which leads to a background rate of $R_b = 3.38\,\mathrm{e}^{-4}$ events per day with $M_\mathrm{W} \geq 0.9$. In other words, an $M_\mathrm{W} \geq 0.9$ event should occur naturally only about every 8 yr. If we solve the Omori–Utsu law (eq. 2) for the duration, we get (Woessner 2005)

$$t_a = \left[\frac{k}{R_b}\right]^{\frac{1}{p}} - c. \tag{3}$$

With these values, we obtain a duration of the 'aftershock' sequence of $t_a = 31 + 29/-14$ yr, where the uncertainties are obtained from the bootstraps in Fig. 5.

Forecasting models

We apply two different classes of model, which we introduce in the following paragraphs. Both models have been used to forecast aftershock sequence and are modified by us to be applied to an induced seismic sequence.

The Reasenberg & Jones model

Reasenberg & Jones (1989, 1990, 1994) combine the GR law (eq. 1) and the Omori-Utsu law (eq. 2) to determine the probability of triggering earthquakes. They express the rate λ of aftershocks with magnitudes larger than M_c at a given time t after the main shock with magnitude M_m as

$$\lambda(t, M_\mathrm{c}) = \frac{10^{\tilde{a}+b(M_\mathrm{m}-M_\mathrm{c})}}{(t+c)^p}, \tag{4}$$

where $\tilde{a}$ is given as

$$\tilde{a} = A_0 - \log\left[\int_S^T (t+c)^{-p}\mathrm{d}t\right], \tag{5}$$

where

$$A_0 = a - b \times M_\mathrm{m} \tag{6}$$

and a, b, p and c are the same constants as in eqs (1) and (2), respectively.

The probability P of one or more events occurring in the magnitude range ($M_1 \leq M < M_2$, with $M_1 \geq M_\mathrm{c}$) and the time range ($S \leq t < T$) is then given as (Wiemer 2000)

$$P = 1 - \exp\left[-\int_S^T \lambda(t, M)\,\mathrm{d}t\right]. \tag{7}$$

Here, we treat each individual event as a main shock and sum the rates λ in each time bin to obtain the joint probability P of observing one or more event in the given magnitude range.

Epidemic Type Aftershock Models

We use an epidemic type aftershock (ETAS) model of Ogata (1988). The rate of aftershocks induced by an event occurring at time t with magnitude M_i is given by

$$\lambda_i(t) = \frac{K}{(c+t-t_i)^p} 10^{\alpha(M_i - M_\mathrm{min})} \tag{8}$$

for time $t > t_i$. The parameters c and p are empirical parameters (compare eq. 4) and K and α describe the productivity of the sequence. The total occurrence rate is the sum of the rate of all preceding earthquakes and a constant background rate λ_0.

$$\lambda(t) = \lambda_0 + \sum_{[i:t<t_i]} \lambda_i(t) \tag{9}$$

We consider that the forcing term should depend on the applied injection flow rate F_r. According to Shapiro & Dinske (2009), we can model the fluid-triggered event rate as proportional to the injection rate. We therefore modify the background to be

$$\lambda_0(t) = \mu + c_f \times F_r(t) \tag{10}$$

with c_f and μ being free parameters.

Forecasts with the ETAS model have to include secondary aftershock triggering during the prediction period. For that purpose, we use the mean of 10 000 Monte Carlo simulations based on the inverse transform method by Felzer *et al.* (2002) to obtain a stable forecast. The contribution in each magnitude bin is simply obtained by applying the GR frequency–magnitude relation to this rate with a given *b*-value.

For the Basel case study, we test different model versions:

(i) In two versions (E1 and E3), we set $c_f = 0$ to see whether the model without the physically reasonable dependence on the flow rate has a similar prediction power.

(ii) Fitting too many parameters to a limited data set can also lead to a reduced prediction power of the model. Thus we fixed in two versions (E1 and E2) as many parameters as possible to generic parameters known from previous model applications.

(iii) In versions E4 and E5, we fit c_f in addition to all other parameters. To forecast rates for time t_i, we use $F_r(t_{i-1})$ for E4 and $F_r(t_i)$ for E5.

MODELLING APPROACHES

Before we apply and compare the statistical forecast models, we define a common framework in which we apply and test the models. This involves choosing the testing period, the updating strategy and the magnitude range in which to test the forecast. Although these choices are somewhat arbitrary, they can potentially have a significant effect on the outcome of the testing, and they reflect to some extent the requirements of end users. We use the experience of the RELM and CSEP experiments to define the 'rules of the game' and retrospective testing of aftershock sequences (Field 2007; Schorlemmer *et al.* 2010; Woessner *et al.* 2011).

For both model classes, the RJ-models (R0–R2) and the ETAS-models (E1–E5), we apply two modelling approaches: (1) Use one set of pre-determined parameters from the entire sequence; and (2) Update model parameter values with successively extending the period for assembling data by 6 hr. Models R1, E1 and E2 fall into class (1); we use generic parameters defined for other sequences and make forecasts with those in 6-hr bins. Models R2 and E3–E5 fall into class (2); here, we start with the same generic parameters, but after the first 6 hr of data, we fit the parameters to the data. We use increasing time bins and fit the parameters again after every time bin. For example, we fit the parameters to the first 12 hr of events to forecast hours 12–18. For both classes, we evaluate the performance of the models in 6-hr bins.

In addition, we consider one model, R0, which is non-causal in the sense that it represents the best fit of a R&J model to the sequence. For this model, we divide the sequence in two parts, a coinjection and a post-injection period and fit a set of parameters to each period.

Table 1 summarizes all eight models. For all models, we have both fixed and free parameters. The fixed parameters are based on generic values determined for other sequences or values found in the literature. For all approaches we fix the *b*-value at 1 and the *c*-value at 0.01 which is somewhat arbitrary; both values are often found in literature. We use neither the *b*-value fitted to the sequence nor the *b*-value of the seismic source zone as we cannot justify using either of them in the forecasting mode. Using a *c*-value of 0.01 is consistent with the literature (Reasenberg & Jones 1989). For the R&J models we use generic parameters defined by Reasenberg & Jones (1989), for model R1 we use them for the whole sequence, for R2 we start with those parameters. For the ETAS models, we use the generic parameters of $p = 1.2$ and $\alpha = 0.8$, which are typical values for tectonic events (Ogata 1992). We apply them to the whole sequence for models E1 and E2 and use them as starting values for E3–E5.

In addition to the fixed parameters, we list the free parameters for each model in Table 1. The more free parameters a model includes, the better its fit to past data should be; yet more parameters come at the cost of less robust models. To comparatively evaluate the model performances one would then have to resort to measures, such as the Bayesian Information Criterion or Akaike Information Criterion, which penalize models with more degrees of freedom. In our case we choose a different strategy that mimics prospective testing: Because each model forecasts the seismicity of the period [t, $t + 6$ hr] only with information obtained until time t, the forecast has zero degrees of freedom; therefore comparing the model likelihoods is sufficient.

Modelling results

We compute forecasts of the seismicity for the eight models summarized in Table 1. For each model, we determine seismicity rates within 6 hr for a magnitude bin from $M_W = 0.9$–3.5 and compare them with observed events above $M_W \geq 0.9$. We start to forecast with the start of the fluid injection at 2.12.2006, 6 p.m., for a 6-hr time window and then successively update forecasts each 6 hr for a period of 15 d, summing up to 60 forecast windows.

The rate forecasts for all models in Fig. 6 are shown compared to the observed seismicity indicated by the solid black line with circle markers. The top panel shows the three different Reasenberg and Jones models R0 to R2, the middle panel shows the ETAS model where flow rate was not included (E1 and E3) and the lowest panel shows the three ETAS models where flow rate was included in the modelling (E2, E4 and E5).

From a first visual inspection of Fig. 6, we observe that model R1 underpredicts the rate of earthquakes during the intense injection period by more than a factor of 10, suggesting that an R&J model with generic parameters is not a suitable model to explain the induced sequence well. Model R0, which uses parameter values estimated retrospectively for the entire sequence for the co- and post-injection periods, demonstrates that if the 'right' set of parameters were known beforehand, an R&J model would be able to explain the seismicity well. Model R2, which updates the generic parameters every 6 hr, matches the seismicity rate much better than Model R1 in the coinjection period of the sequence; however, the model consistently overpredicts the post-injection seismicity of the sequence. This is similarly observed for model E1, an ETAS model using pre-defined parameters that are not updated during the forecasting experiment. Neither R1 nor E1 are able to respond adequately to the change of boundary conditions that occurs when the injection stops. Models E2–E5 fit the seismicity rate well based on a visual inspection.

Performance evaluation

To quantitatively test the model forecasts in a pseudo-prospective approach, we use the N(umber)-test (Schorlemmer *et al.* 2007, 2010; Lombardi & Marzocchi 2010; Werner *et al.* 2010; Woessner *et al.* 2011). This test compares the total forecast rates with the total number of observed earthquakes in the entire volume and indicates whether the too few or too many events are forecast or if the forecast is consistent with the observation. For example, if the model forecasts 0.5 events and 1 is observed, the cumulative Poisson

Table 1. Summary of the models and updating strategies used in the study. Model names are used in the text, the type indicates the base model. Main differing assumptions are indicates together with fixed/initial parameter values and free parameter values estimated from the sequence.

Name	Type	Assumptions	Fixed par.	# Free par.
R0	R & J	Retrospective Two fitted periods No update	$b = 1, c = 0.01$ d	2 p, a
R1	R & J	Generic parameters No update	$p = 0.91, a = -1.67$ $b = 1, c = 0.01$ d	0
R2	R & J	Start with generic par. Update after 6 hr Increasing time window	$b = 1, c = 0.01$ d	2 p, a
E1	ETAS	Generic parameters	$p = 1.2, \alpha = 0.8, c_f = 0$ $b = 1, c = 0.01$ d	2 K, μ
E2	ETAS	Generic Parameters with flow rate	$p = 1.2, \alpha = 0.8$ $b = 1, c = 0.01$ d	3 K, μ, c_f
E3	ETAS	Start with generic par. Update after 6 hr	$b = 1, c_f = 0$	5 p, α, c, K, μ
E4	ETAS	Start with generic par. with flow rate	$b = 1$	6 $p, \alpha, c, K, \mu, c_f$
E5	ETAS	Start with generic par. flow rate with info from forecast bin	$b = 1$	6 $p, \alpha, c, K, \mu, c_f$

distribution (PCDF) results in a quantile score of δ = PCDF (1, 0.5) = 0.910. We reject the forecasts at the 0.05 significance level, thus for δ-values smaller than 0.025 and larger than 0.975. We determine the N-test for each 6-hr bin; the rejection ratio R_N denotes the percentage of test bins that are rejected.

In addition, we perform the L(ikelihood)-test (Schorlemmer *et al.* 2007, 2010). This test evaluates whether the forecast number of events and the distribution in the magnitude bins is consistent with the observation, again assuming the entire volume as one spatial bin. For each magnitude bin we compute the log-likelihoods and sum this to a joint log-likelihood of the forecast. To verify that the joint log-likelihood is consistent with what is expected if the model is correct, we simulate 10 000 synthetic catalogues consistent with the forecast model and compute their log-likelihood values. This distribution of likelihood values is then compared with the observed log-likelihood. The quantile score γ then measures the amount of simulated log-likelihood values that are smaller than the observed log-likelihood. This test is one-sided and we reject a model if $\gamma < 0.025$ which implies that the observed log-likelihood is much smaller than expected if the model is true. According to the N-Test, we define the rejection ratio R_L that denotes the percentage of test bins that do not pass the L-test.

In Fig. 7 we show the log-likelihoods for each 6-hr bin of model E5. Bins, in which the observed log-likelihood falls within the 95 per cent confidence interval of the simulated values (grey error bars) are indicated by black squares; light grey squares denote bins in which the observed log-likelihood score falls outside the confidence limits. The rejection rate R_L for E5 is 0.15 (Table 2), therefore nine bins out of 60 are rejected.

In addition, we calculate the joint log-likelihood of each model as a sum of over all time bins. Less negative joint log-likelihood indicates a better fit between model and data.

In Table 2 we summarize the scores of both tests and the joint log-likelihoods for each model. The joint log-likelihood confirms the visual inspection: The poorest model in terms of likelihood is R1, the best model in terms of the joint log-likelihood is model E5, closely followed by E4. However, none of the pseudo-prospective models reach the same likelihood as model R0, which is a retrospective fit to the entire sequence.

The models allow us to forecast the rate, or probability, of larger and potentially felt or damaging events. Fig. 8 shows such a forecast as the probability of a $M_W \geq 2.0$, 3.0 and 4.0 events as a function of time for the next 6-hr period based on model E5. Probabilities of magnitudes above 2 are high from the start and decay to probability values of less than 0.5 only after the injection is finished. Probabilities for higher magnitudes such as 3 and 4 are smaller, but reach 0.51 and 0.07 nonetheless.

Translating forecast rates to time-varying seismic hazard

The forecast rates for each magnitude bin of each model are the basic input needed to produce a time-dependent hazard model. We convert rates into probabilities of a given ground motion intensity, using standard procedures introduced originally by Cornell (1968). Hazard is the result of a combination of seismic rates, their frequency–size distribution and the Ground Motion Prediction Equation (GMPE) and its uncertainty. In contrast to the standard hazard assessment, which is computed for recurrence periods of hundreds to thousands of years, we are here interested in short-term hazard in the order of hours to days. This is identical to the time-dependent hazard assessment introduced for aftershocks sequences by Wiemer (2000) and used in California (Gerstenberger *et al.* 2005). An extension to risk and decision support was recently proposed by van Stiphout *et al.* (2010).

To adapt the ground motion forecast to for Switzerland, we adjust the attenuation relation according to Fäh *et al.* (2003), expressed in the European Macroseismic Scale intensity (EMS). We integrate the hazard from $M_W \geq 2.5$ upward. The maximum magnitude, M_{max}, that is used in the induced hazard is another critical factor; it determines the roll-off at higher intensities. As an initial estimate we choose a somewhat arbitrary value of $M_W = 5.0$; however, we show the effect of a different maximum magnitude later.

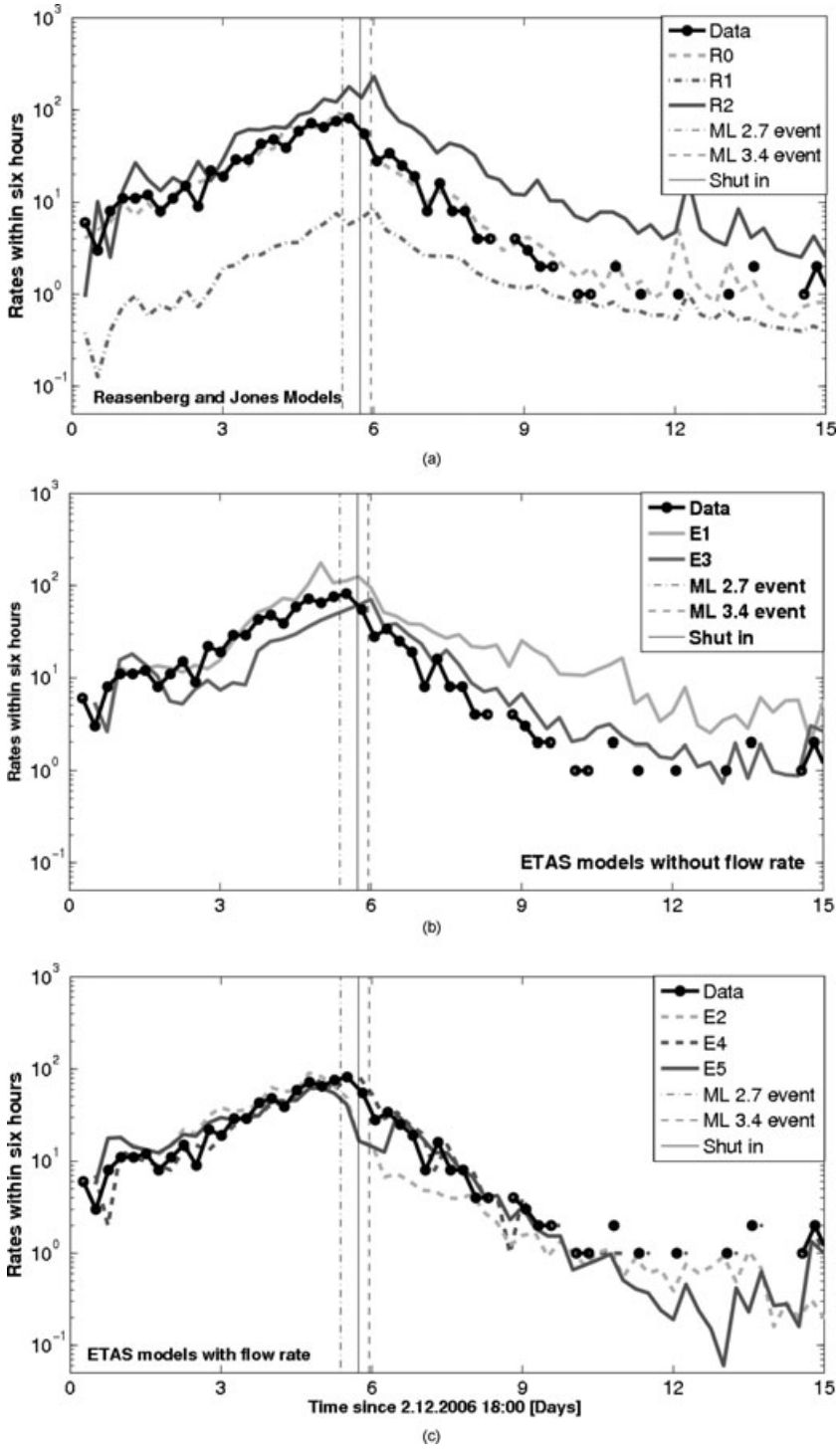

Figure 6. Summary of all eight models; (a) three models based on the Reasenberg & Jones approach, (b) two models based on ETAS approach without flow rate and (c) three models based on ETAS approach with flow rate. In all three panels, the observed rates within the next 6 hr is indicated with a bold black line and circles. The time of the shut in and the two largest events that led to actions within the traffic-light system are indicated.

Each model can then be translated into a hazard curve for the next 6-hr period. Fig. 9(a) shows three typical examples of hazard curves, at three different times (day 3, day 6 and day 12), the background hazard curve is also indicated. The hazard during the intense induced sequence exceeds the background by a factor of more than 100. To visualize the evolution of the hazard as a function of time, we show the probability of exceeding EMS intensities III, IV and V for model E5 as a function of time in Fig. 9(b). EMS intensity III is the level at which a few people start feeling a light shaking indoors, an event with intensity IV is already felt by many indoors and intensity V is felt by most indoors and the vibration is strong (Gruenthal 1998). The exceedance probabilities reach levels of 0.2 for EMS III and IV already on day 1. A probability level of 0.1 is reached on day 4 for EMS intensity V. The maximum probability levels are 0.99, 0.55 and 0.12 for EMS intensity III, IV and V, respectively. All three maximum levels are reached after 5 d and 6 hr, that is,

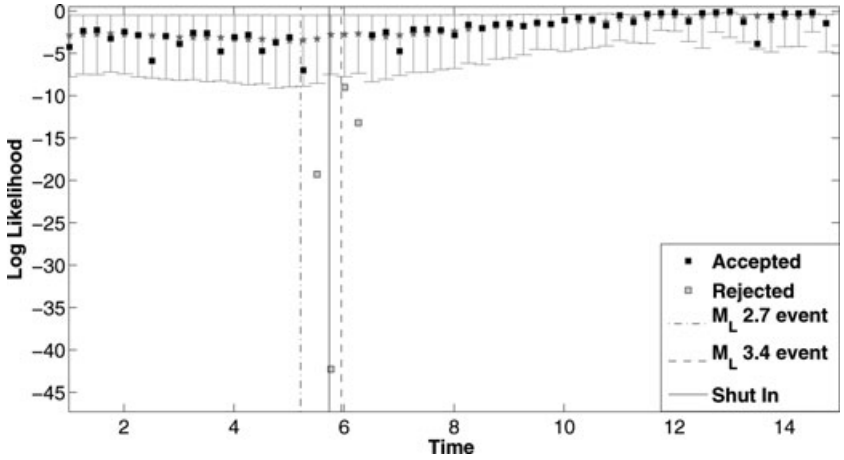

Figure 7. Log-likelihood values of the L-test for model E5 as a function of time. The panels display the mean and the 97.5 and 2.5 percentiles (grey dot and bars); days on which a log-likelihood value LLS(t) is accepted are indicated with black squares; if the model is rejected, we indicate this with grey squares.

Table 2. Quantitative evaluation of model forecasts: model name, model type, fraction of rejected time bin in N-Test, fraction of rejected time bins in L-Test and joint log-likelihood of the each model.

Name	Type	R_N	R_L	Joint log-likelihood
R0	Reasenberg & Jones	0.33	0.17	−170.85
R1	Reasenberg & Jones	0.73	0.60	−1433.77
R2	Reasenberg & Jones	0.54	0.54	−846.79
E1	ETAS	0.65	0.63	−478.11
E2	ETAS	0.48	0.2	−266.67
E3	ETAS	0.43	0.35	−276.19
E4	ETAS	0.42	0.20	−204.35
E5	ETAS	0.40	0.15	−204.33

before water injection was stopped and before the largest event occurred.

DISCUSSION AND CONCLUSION

The ultimate termination of the Basel EGS pilot study due to the public outcry as a consequence of the sequence of strongly felt earthquakes poses a substantial challenge for future EGS systems. In hindsight, it is obvious that the potential for triggering felt earthquakes was underestimated by GEL, as well as by the regulatory bodies. However, even today, more than 3 yr after the main injection, scientists are neither able to accurately forecast the response of a rock volume to the injection of water at high pressures, nor are there universally accepted guidelines for hazard and risk assessment prior, during and after a stimulation. We see our study as a contribution towards an improved ability to forecast induced seismicity as it unfolds.

The seismicity recorded during and after the stimulation of the Basel EGS is one of the best monitored sequences of its kind. Our analysis of the monitoring completeness as a function of time (Fig. 3) shows that $M_c(t)$ varies between 0.5 and 0.9. Higher values of M_c are typical during the first hours of intense aftershocks sequences, when the coda of larger events mask the smaller ones (e.g. Woessner & Wiemer 2005). We suspect that the temporal changes in M_c (Fig. 3) are caused by the same mechanism. In addition, it is possible that $M_c(t)$ changes as a results of temporal changes in the activity areas: If areas of a higher M_c (i.e. further away from the sensors) are more active, the overall M_c will appear to increase.

Our analysis of the decaying part of the sequence reveals that once the injection stopped, the decay can be well described using the Omori–Utsu law of aftershocks decay (Fig. 5). The comparison between model and data passes the Kolmogorov–Smirnov test, which in our experience is a rather strict test for compliance to aftershock seismicity (Conover 1972; Woessner *et al.* 2004). From a statistical viewpoint, there is nothing special about the sequence, which is consistent with the conclusion by Deichmann & Giardini (2009) that the earthquakes induced by the reservoir stimulation below Basel, rather than representing a case of hydrofracturing, occurred mainly as shear dislocation on pre-existing faults that were triggered by the increase in pore pressure due to the injected water, but driven by the ambient tectonic stress. In addition, we suggest that an unknown fraction of the events are not directly triggered by the change in pore pressure but rather indirectly as 'aftershocks' to other events, or as 'daughter events' in a cascading model.

Our assessment that the seismicity will take about 31+29/−14 years to decay to the background is consistent with the observations of aftershock sequences (Stein & Mian Liu 2009) as well as with predictions of laboratory studies (Dieterich 1994). In aftershock sequences, the duration of the sequences, defined as the time when the rate returns to that before the event, has been proposed to be inversely correlated with the tectonic loading rate. The rate and state model of fault friction, which predicts changes in fault properties after earthquakes, and which is commonly used for aftershock studies (Dieterich 1994), predicts an aftershock duration of

$$t_a = \frac{A\sigma}{\dot{\tau}}, \tag{11}$$

where $\dot{\tau}$ is the rate of shear stressing on the fault, σ is the normal stress and A is a constitutive parameter. Although the stressing rate is is hard to measure, it is roughly proportional to the loading rate. The loading rate in Switzerland, and also in the Basel region is known to be low ($<1\,\mathrm{mm\,yr^{-1}}$) and aftershock sequences are therefore expected to last longer than in tectonically more active regions.

The major unknown in the estimation of the aftershock duration is the local background rate. Ideally, monitoring at a comparable completeness level of the same rock volume for several years would establish the background seismicity rate. This approach is prohibitively expensive. We chose the regional background rate, extrapolated from the microseismicity record, as a proxy, but the local variability of the background at the scale of 1 km is unknown.

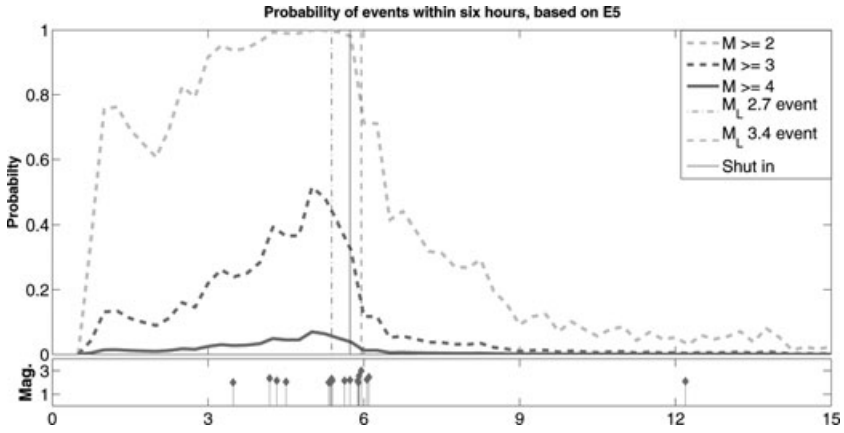

Figure 8. Probabilities for events within the next 6 hr for magnitudes ≥2, 3 and 4 with time, based on model E5. Observed events above magnitude 2 are indicated in the lower panel.

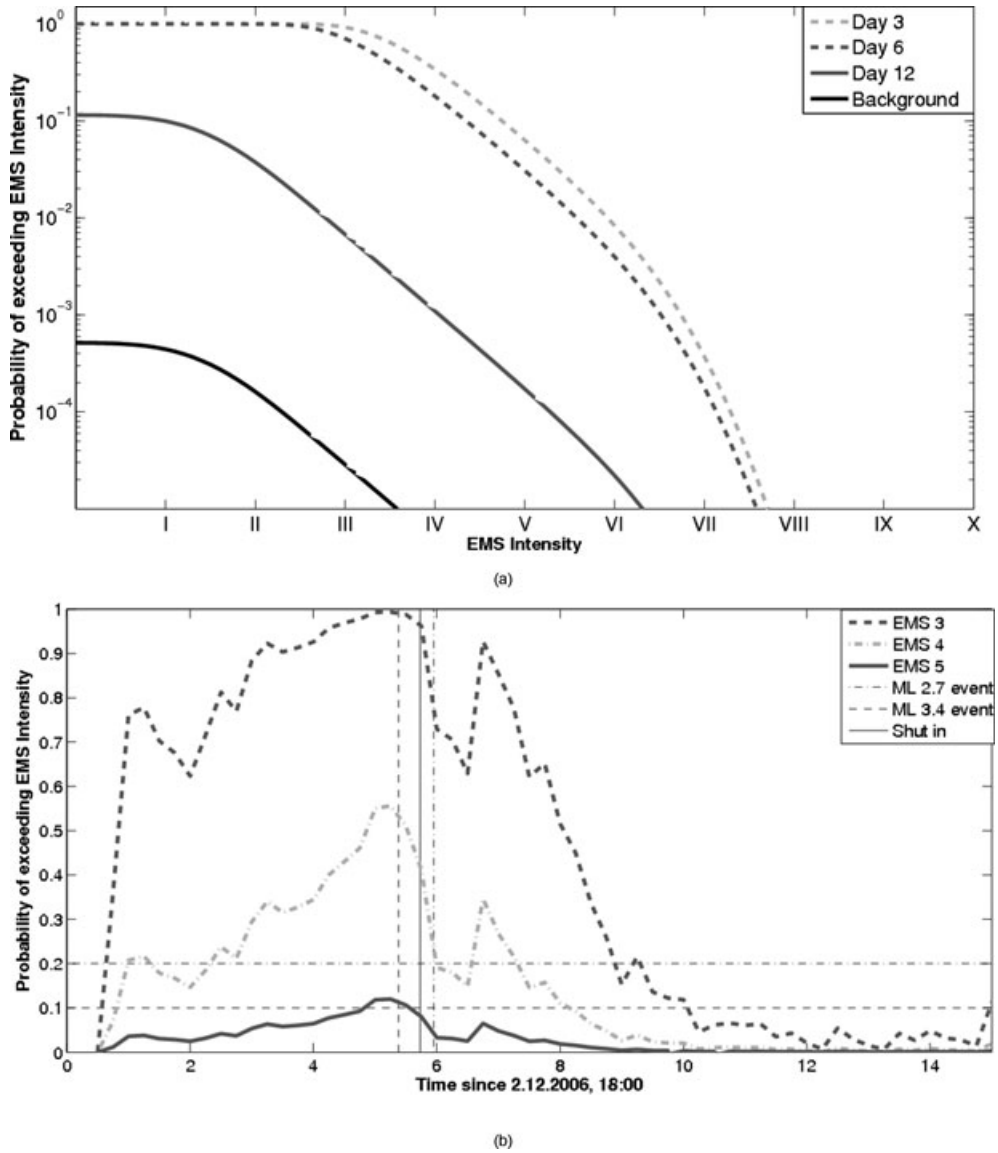

Figure 9. Hazard curves based on model E5. (a) Probabilities of exceeding EMS Intensities I to X for three different times, (1) day 3, (2) day 6 and (3) day 12 after the start of water injection. All three curves are based on $M_{max} = 5$. (b) Probabilities of exceeding EMS intensities III, IV and V within the next 6 hr for the first 15 d. Indicated are also the times of the two largest events during the first 15 d and the time of the termination of water injection.

Supporting evidence for a low background rate comes from the observation that micro-earthquakes in the past 3 yr have been confined to the induced volume, no events have been detected outside the well-defined volume shown also in Fig. 2, although the network of borehole sensors in principle would be able to detect them.

Our focus of modelling the rates of the microseismicity is on two largely statistical models, the Reasenberg and Jones approach (Reasenberg & Jones 1989) and the ETAS model (Ogata 1988; Hainzl & Ogata 2005). There are a number of reasons why we consider these models suitable alternatives to more physics-based modelling approaches (Kohl & Megel 2007; Shapiro *et al.* 2007, 2010; Baisch *et al.* 2009b).

(i) Catalogue data of the microseismicity used for forecasting is, at least in principle, readily available in near real time, a requirement for building a real-time hazard assessment system.

(ii) Statistical models are well understood and well tested, because they are used commonly to model the behaviour of aftershocks and swarms (Ogata 1988; Reasenberg & Jones 1989; Hainzl & Ogata 2005; Woessner *et al.* 2010).

(iii) The statistical models are comparatively simple, and make no assumptions about the underlying physics or rock properties. They can therefore be considered as starting models, while more refined models should be used for forecasting only if shown to be superior in their ability to forecast the seismicity.

(iv) The models output a seismicity rate forecast which can be converted to time-dependent hazard estimates: both of these outputs can readily be implemented as a measure for decisions on continuing an EGS experiment. This is not easily performed with physics-based models without introducing additional stochasticity (Hainzl *et al.* 2009; Cocco *et al.* 2010; Woessner *et al.* 2010).

One of our major objectives was to define a more quantitative alternative to the traffic-light system used so far in the monitoring and regulation of EGS creation (Bommer *et al.* 2006). Based on our findings here we propose that such a system would be built using a model such as E5 applied in real time, and translating the rates into hazard. This implies that first of all, a suitable real-time motoring system must exist that, within minutes, can deliver reliable locations and magnitudes for the dozens to hundreds of events that can occur during stimulation every hour. It also implies that beforehand, a plan of action must be agreed upon that specifies the actions to take once a certain hazard threshold is reached. Models such as E5, which integrates the flow rate, can be used to forecast the expected hazard if a certain flow rate is applied in the hours and days to come; they can thus not only be used from a regulatory point of view, but also assist the operators in their decision making. The approach we propose here is, in our assessment, a substantial advance when compared to the traffic-light system, because:

(i) Not only a single magnitude of one larger events counts, but also the many small ones that occur. Including these small events for forecasting and decision making increases the robustness.

(ii) We consider real-time information and update the forecast as new information arrives.

(iii) Our model has been tested in a pseudo-prospective sense and also works well at regional and local scale for natural seismicity.

(iv) The model is hazard/risk based, and is able to consider uncertainties in all parameters and thus allows for informed decisions making.

(v) Our model can be used to simulate alternative stimulation strategies and their implications for future seismicity; by assuming different parameter values in these simulations, adequate modelling of different tectonic environments in terms of their productivity can be achieved.

Translating the forecast rates into seismic hazard is a straightforward calculation, and has been already implemented successfully at local and regional scale (e.g. Gerstenberger *et al.* 2005, 2007). It was also applied in the Basel risk study (Baisch *et al.* 2009a). We believe that this translation should be an integral part of a future monitoring and regulatory framework, because it allows to set thresholds that are hazard/risk based and fully probabilistic by definition. We anticipate that the underlying forecast models will become increasingly complex, and ensemble forecasting or logic trees will be considered just like they are in weather forecasting or probabilistic seismic hazard assessment. By translating all forecast rates, and their uncertainties, into hazard, taking into account the uncertainties in the expected ground motions, seismologists have combined the best of their knowledge into (comparatively) simple numbers. The hazard integration then takes care of integrating all magnitude bins, as well as the fact that sometimes smaller events can create larger than to be expected ground motions.

Further adding knowledge on the fragility of buildings, and translating into time-dependent risk, is then the next logical step, also already a part of the SERIANEX study (Baisch *et al.* 2009a). van Stiphout *et al.* (2010) introduced time-dependent risk assessment at a local scale by analysing the seismicity in the L'Aquila region. It is entirely possible to compute a threshold, for example, of a 1 per cent probability of exceeding damages of CHF 10 million or more, or not exceeding a 10 per cent annual probability of causing a casualty. Therefore, such a threshold will be an integral part of any insurance scheme implemented for EGS systems. We believe that EGS technology and the associated clean energy does not come without a risk, as it is the case for almost any other technology. As scientists we strive to describe and reduce the risk. The decision on how much risk to take—essentially where to set a threshold in our Fig. 9(b)—is ultimately a political process. We see our paper in this context as a contribution towards building robust, and community-accepted, forecast models.

One of the major 'free' parameters of such a model is the maximum possible magnitude, $M_{\max}$, used in the hazard calculation. There is very little knowledge how to set this number. Contemporary PSHA studies tend to set a larger $M_{\max}$ than applied in both our study and SERIANEX. The underlying question is wide open: Does the intense activity during stimulation also increase the probability of triggering an event with source dimensions significantly larger than the stimulated volume? The conservative answer to this question taken in regional time-dependent hazard assessment is that all earthquakes are treated equally, such that triggering is possible all the way to the regional $M_{\max}$, which in the case of Switzerland is set between around 6 and 7.5 (Wiemer *et al.* 2009). Choosing a higher $M_{\max}$ will increase the hazard at higher ground motions, but may not be the critical factor in reaching an action threshold because already moderate events of M 3–4, with source dimensions well within the induced volume, can cause large enough ground motions to create non-structural damage. We show the effect of three different $M_{\max}$: (1) 3.7, which was used by Baisch *et al.* (2009a) for the Basel risk study, (2) 5, which was used for this study and (3) 7, which is the maximum magnitude of the seismic zone of Basel in Fig. 10(a). We show the time evolution for these three $M_{\max}$ for two different EMS intensities III and V. While the probabilities for the lower intensity and higher probabilities practically overlap, we see a higher differentiation for the higher intensity. The inset in Fig. 10(a) shows a snapshot of the probabilities for different $M_{\max}$

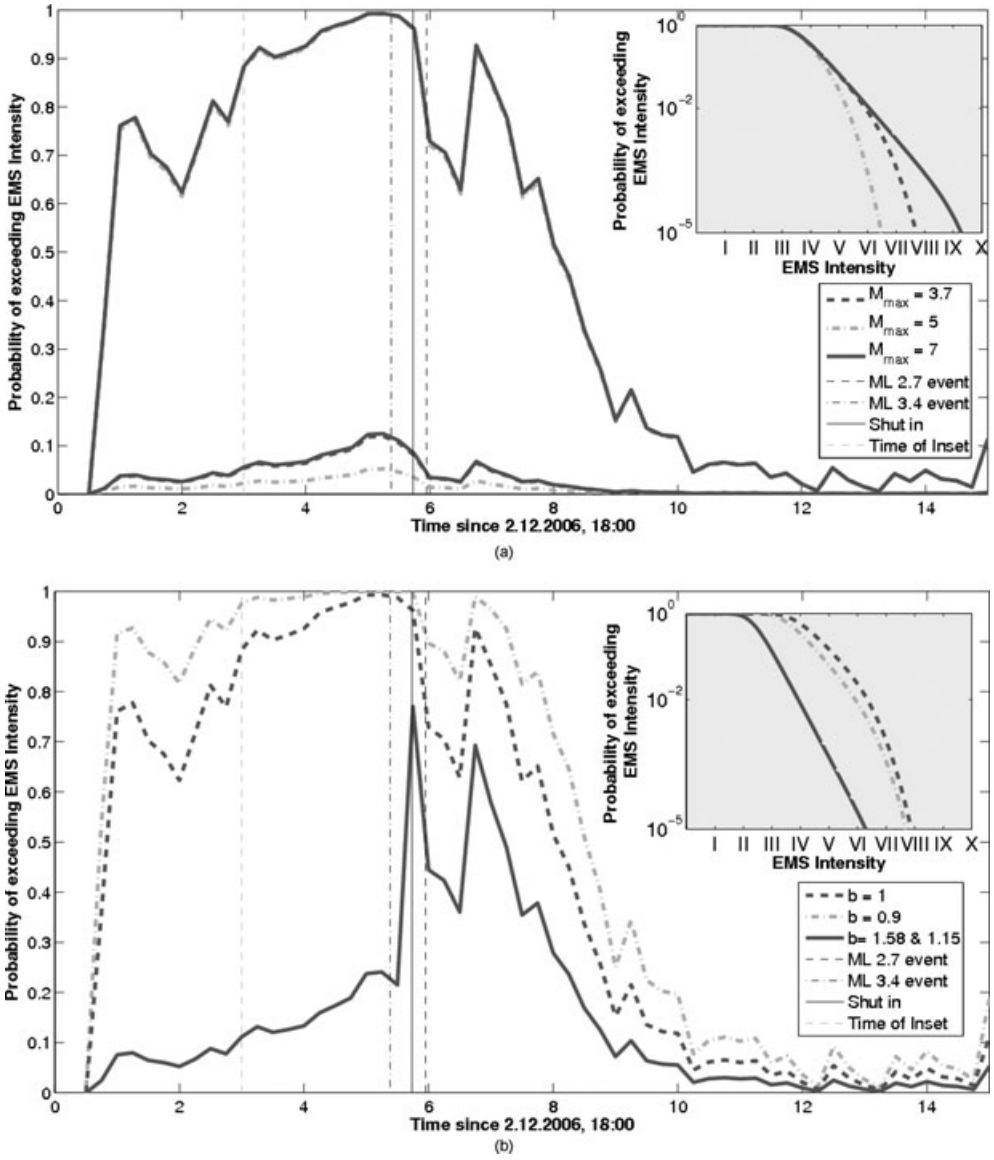

Figure 10. (a) Effect of different maximum magnitudes with (1) M_{max} = 3.7, (2) M_{max} = 5 and (3) M_{max} = 7. Higher probabilities show the probabilities of exceeding EMS intensity III, lower probabilities the exceedance probabilities for EMS V. The inset marks a hazard snapshot for day 3 (indicated by a light grey dashed line). (b) Effect of different *b*-values on the hazard integration with (1) b = 0.9, (2) b = 1 and (3) b = 1.58 for coinjection and 1.15 for post-injection events. The inset shows again a hazard snapshot for day 3.

at day three (indicated by a dashed line). The curve with the smallest M_{max} drops off the fastest while the other two only begin to drop off at EMS intensities VI and VIII, respectively.

Another factor that we have not yet discussed is the influence of the *b*-value. We assume a constant *b*-value of 1 for all of our models. As we do not test for different magnitude bins, it has no influence there. However, the *b*-value has an influence on the hazard integration. In Fig. 10(b) we show the time evolution of the hazard for three different assumptions: (1) *b*-value of 1, as in this study, (2) *b*-value of 0.9 as the seismogenic source zone of Basel (Giardini *et al.* 2004; Wiemer *et al.* 2009) and (3) *b*-values of 1.59 for coinjection and 1.15 for post-injection events, respectively, as we find for this data set (compare Fig. 4). The inset shows, similar as in Fig. 10(a), a snapshot of the hazard for different EMS intensities at day three. The hazard is strongly reduced with higher *b* values and increases as this value decreases. However, for this study we fixed the *b*-value at 1, as the values of the whole sequence cannot be justified for a real-time approach as they are only known *a posteriori*. For all curves in Fig. 10(a) we use a *b*-value of 1 and for all curves in Fig. 10(b) we use a M_{max} of 5.

We show here the importance of quantitative testing of a model performance. We apply tests defined by the international Collaboratory for the Study of Earthquake Predictability (CSEP, www.cseptesting.org). Work on induced seismicity, can in our opinion, benefit substantially from CSEP, for example, by using the community-accepted testing algorithms, such as the N- and L-test employed here, and by exploiting the models tested within CSEP. Both the R&J and ETAS models are currently being tested in fully prospective tests in a variety of testing regions (Schorlemmer *et al.* 2010). In addition, pseudo-prospective tests of models at the more local scale of aftershock sequences have been performed recently as an extension of the CSEP concept (Hainzl *et al.* 2009; Cocco *et al.* 2010; Woessner *et al.* 2010), which again offer highly relevant insights for modelling the induced seismicity. Reciprocally,

the CSEP process can in our opinion also benefit from the work on induced sequences, because it offers the possibility to evaluate and improve modelling and testing within a reasonably well-constrained environment.

The testing applied in our study reveals that a fully retrospective matching of a sequence and a model forecast is misleadingly successful. Model R0, a rather simple model, offers the best fit to the data both in terms of N- and L-test. It passes as an accepted model and is not rejected by the tests at the 5 per cent significance levels. However, the same model, when applied in a pseudo-prospective sense (R2), performs quite poorly when compared to other models. This illustrates that the pseudo-prospective approach applied by us, which recreates a real-time setup, is needed to achieve a less biased assessment of the forecasting ability of models. We suggest the rigorous testing approach applied in this study has been lacking so far in the evaluation of models for induced seismicity, and that future studies that propose methods for modelling induced seismicity should choose similar quantitative approaches.

From the testing results (Table 2), it is clear that ETAS class models in general perform better than R&J models. This finding is consistent with the current assessment of these models in various testing regions and pseudo-respective testing on aftershock sequences. The performance difference in our case may in fact be more substantial, because of the extreme conditions of very high rates and the dramatic change of process at the termination of the injection. The R&J model R2 is not able to adjust its forecast sufficiently at the end of the injection, and subsequently overpredicts the rates. The ETAS models, which differentiate between a background rate and a term of induced events, are much better able to adjust to this change. The ETAS formulation allows us also to specifically consider the induced flow rate, which is not possible for R&J. Including the flow rate does indeed substantially improve the model performance.

More work is needed to define a community-accepted real-time system alternative to the traffic-light system. For example, our findings here should be applied consistently to a number of induced sequences to check if the results are robust. Does model E5 always provide the best fit? In addition, it is possible that the forecasting ability can be further improved if the spatial variability of microseismicity is better understood. Of course we also need to test more physics-based models in comparison to the statistical ones—yet these models need to provide pathways to generate seismicity rates.

One question that remains unsolved is whether the Basel EGS experiment would have been successful if our model E5 were applied for the real-time decision making. Could we have prevented an event felt by the public and continued the experiment until reservoir creation? We see in Fig. 9(b) that probabilities of exceeding EMS intensity 3 already reached 20 per cent after 1 d and 50 per cent after 2.5 d. Probabilities of exceeding EMS 5 also reached a level of 0.1 after less than 4.5 d. So by choosing any of these arbitrary thresholds, action would have been taken more than 1 d before the actual termination of the water injection on 2006 December 6 11:33 a.m. These questions remains unsolved until a future application and the *a priori* agreed schedule of decisions.

ACKNOWLEDGMENTS

We would like to thank Geothermal Explorers Ltd. that provided seismicity and injection data without which this work would not have been possible. This study is part of the project GEOTHERM, funded by the Competence Center of Environment and Sustainability (CCES) of ETH (http://www.cces.ethz.ch/projects/nature/geotherm/). We thank T. Kraft, G. de Souza and J. Zechar for valuable comments on the manuscript.

REFERENCES

Baisch, S. *et al.*, 2009a. Deep heat mining basel: seismic risk analysis, Tech. Rep., Serianex.

Baisch, S., Voeroes, R., Rothert, E., Stang, H., Jung, R. & Schellschmidt, R., 2009b. A numerical model for fluid injection induced seismicity at soultz-sous-forets, *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.,* **47**(3), 405–413.

Baisch, S., Voeroes, R., Weidler, R. & Wyborn, D., 2009c. Investigation of fault Mechanisms during Geothermal Reservoir Stimulation Experiments in the Cooper Basin, Australia, *Bull. seism. Soc. Am.,* **99**(1), 148–158.

Bethmann, F., Deichmann, N. & Mai, M., 2007. Moment magnitude, in Evaluation of the Induced Seismicity in Basel 2006/2007: locations, Magnitudes, Focal Mechanisms, Statistical Forecasts and Earthquake Scenarios, Tech. rep., Report of the Swiss Seismological Service to Geopower Basel AG, Basel.

Bommer, J.J., Oates, S., Cepeda, J.M., Lindholm, C., Bird, J., Torres, R., Marroquin, G. & Rivas, J., 2006. Control of hazard due to seismicity induced by a hot fractured rock geothermal project, *Eng. Geol.,* **83**(4), 287–306.

Cocco, M., Hainzl, S., Catalli, F., Enescu, B., Lombardi, A.M. & Woessner, J., 2010. Sensitivity study of forecasts based on Coulomb stress calculation and rate-state frictional response, *J. geophys. Res.,* **115,** B05307, doi:10.1029/2009JB006838.

Conover, W.J., 1972. A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions, *J. Am. Stat. Assoc.,* **67**(339), 591–596.

Cornell, C., 1968. Enigineering seismic risk analysis, *Bull. seism. Soc. Am.,* **59**(5), 1583–1606.

Deichmann, N. & Ernst, J., 2009. Earthquake focal mechanisms of the induced seismicity in 2006 and 2007 below Basel (Switzerland), *Swiss J. Geosci,* **102,** 457–466.

Deichmann, N. & Giardini, D., 2009. Earthquakes Induced by the Stimulation of an Enhanced Geothermal System below Basel (Switzerland), *Seism. Res. Lett.,* **80**(5), 784–798.

Dieterich, J., 1994. A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. geophys. Res.,* **99**(B2), 2601–2618.

Dyer, B., Schanz, U., Ladner, F., Haering, M. & Spillman, T., 2008. Microseismic imaging of a geothermal reservoir stimulation, *Leading Edge,* **27,** 856–869.

Fäh, D. *et al.*, 2003. Earthquake catalog of Switzerland (ECOS) and the related macroseismic database, *Eclogae Geologicae Helvetiae,* **96**(2), 219–236.

Felzer, K., Becker, T., Abercrombie, R., Ekstrom, G. & Rice, J., 2002. Triggering of the 1999 M-W 7.1 Hector Mine earthquake by aftershocks of the 1992 M-W 7.3 Landers earthquake, *J. geophys. Res.,* **107,** doi:10.1029/2001JB000911.

Field, E.H., 2007. Overview of the Working Group for the Development of Regional Earthquake Likelihood Models (RELM), *Seism. Res. Lett.,* **78**(1), 7–15.

Gerstenberger, M., Wiemer, S., Jones, L. & Reasenberg, P., 2005. Real-time forecasts of tomorrow's earthquakes in California, *Nature,* **435,** 328–331.

Gerstenberger, M.C., Jones, L.M. & Wiemer, S., 2007. Short-term aftershock probabilities: case studies in California, *Seism. Res. Letts.,* **78**(1), 66–77.

Giardini, D., 2009. Geothermal quake risks must be faced, *Nature,* **461,** 848–849.

Giardini, D., Wiemer, S., Fah, D. & Deichmann, N., 2004. *Seismic hazard assessment of Switzerland, 2004, Tech. rep.,* Swiss Seismological Service, ETH Zurich.

Gruenthal, G., ed., 1998. European Macroseismic Scale 1998 (EMS-98), Cahiers du Centre Europeen de Geodynamique et de Seismologie 15.

Gutenberg, B. & Richter, C.F., 1942. Earthquake magnitude intensity, energy, and acceleration, *Bull. seism. Soc. Am.,* **32,** 163–191.

Haering, M.O., Schanz, U., Ladner, F. & Dyer, B.C., 2008. Characterisation of the Basel 1 enhanced geothermal system, *Geothermics,* **37**(5), 469–495.

Hainzl, S. & Ogata, Y., 2005. Detecting fluid signals in seismicity data through statistical earthquake modeling, *J. geophys. Res.,* **110,** B05S07, doi:10.1029/2004JB03247

Hainzl, S., Enescu, B., Catalli, F., Cocco, M., Wang, R., Roth, F. & Woessner, J., 2009. Aftershock modeling based on uncertain stress calculations, *J. geophys. Res.,* **114,** B05309, doi:10.1029/2008JB006011.

Kohl, T. & Megel, T., 2007. Predictive modeling of reservoir response to hydraulic stimulations at the European EGS site Soultz-sous-Forets, *Int. J. Rock Mech. Min. Sci. Geomech. Abstr,* **44**(8), 1118–1131.

Kraft, T. *et al.*, 2009. Enhanced geothermal systems in urban areas: Lessons learned from the 2006 Basel ml 3.4 earthquake, *EOS,* **32**(90), 273–274.

Lombardi, A.M. & Marzocchi, W., 2010. The Assumption of Poisson Seismic-Rate Variability in CSEP/RELM Experiments, *Bull. seism. Soc. Am.,* **100**(5A), 2293–2300.

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

Ogata, Y., 1992. Detection of precursory relative quiescence before great earthquakes through a statistical-model, *J. geophys. Res.,* **97**(B13), 19 845–19 871.

Ogata, Y., 1999. Seismicity analysis through point-process modeling: a review, *Pure appl. Geophys.,* **155,** 471–507.

Reasenberg, P.A. & Jones, L.M., 1989. Earthquake hazard after a mainshock in California, *Science,* **243**(4895), 1173–1176.

Reasenberg, P. & Jones, L., 1990. California aftershock hazard forecasts, *Science,* **247**(4940), 345–346.

Reasenberg, P. & Jones, L., 1994. Earthquake aftershocks: update, *Science,* **265**(5176), 1251–1252.

Ripperger, J., Kaestli, P., Fah, D. & Giardini, D., 2009. Ground motion and macroseismic intensities of a seismic event related to geothermal reservoir stimulation below the city of Basel: observations and modeling, *Geophys. J. Int.,* **3**(179), 1757–1771.

Schorlemmer, D., Gerstenberger, M.C., Wiemer, S., Jackson, D. & Rhoades, D.A., 2007. Earthquake likelihood model testing, *Seism. Res. Lett.,* **87,** 17–29.

Schorlemmer, D., Zechar, J.D., Werner, M.J., Field, E.H., Jackson, D.D., Jordan, T.H. & Group, T.R.W., 2010. First results of the regional earthquake likelihood models experiment, *Pure appl. Geophys.,* **167**(8-9), 859–876.

Shapiro, S.A. & Dinske, C., 2009. Fluid-induced seismicity: pressure diffusion and hydraulic fracturing, *Geophys. Prospect.,* **57**(2), 301–310.

Shapiro, S.A., Dinske, C. & Kummerow, J., 2007. Probability of a given-magnitude earthquake induced by a fluid injection, *Geophys. Res. Lett.,* **34**(22), doi:10.1029/2007GL031615.

Shapiro, S., Dinske, C., Langenbruch, C. & Wenzel, F., 2010. Seismogenic index and magnitude probability of earthquakes induced during reservoir fluid stimulations, *Leading Edge,* **29**(3), 304–309.

Smith, M.C., 1983. A history of hot dry rock geothermal energy systems, *J. Volcanol. Geotherm. Res.,* **15**(1–3), 1–20.

Stein, S. & Mian, L., 2009. Long aftershock sequences within continents and implications for earthquake hazard assessment, *Nature,* **461**(7269), 87–89.

van Stiphout, T., Wiemer, S. & Marzocchi, W., 2010. Are short-term evacuations warranted? Case of the 2009 L'Aquila earthquake, *Geophys. Res. Let.,* **37,** L06306, doi:10.1029/2009GL042352.

Tenzer, H., 2001. Development of hot dry rock technology, *Geo-Heat Cent. Quart. Bull.,* **22**(4), 14–22.

Utsu, T., 1961. A statistical study on the occurrence of aftershocks, *Geophys. Mag.,* **30,** 521–605.

Werner, M.J., Zechar, J.D., Marzocchi, W., Wiemer, S. & CSEP-Italy, Working Grp, 2010. Retrospective evaluation of the five-year and ten-year CSEP-Italy earthquake forecasts, *Ann. Geophys.,* **53**(3), 11–30.

Wiemer, S., 2000. Minimum magnitude of complete reporting in earthquake catalogs: examples from Alaska, the Western United States, and Japan, *Bull. seism. Soc. Am.,* **90,** 859–869.

Wiemer, S., 2000. Introducing probabilistic aftershock hazard mapping, *Geophys. Res. Let.,* **27**(20), 3405–3408.

Wiemer, S., 2001. A software package to analyze seismicity: Zmap, *Seism. Res. Lett.,* **72,** 373–382.

Wiemer, S., Giardin, D., Fäh, D., Deichmann, N. & Sellami, S., 2009. Probabilistic seismic hazard assessment for Switzerland: best estimates and uncertainties, *J. Seismol.,* **13**(4), 449–478.

Woessner, J., 2005. Correlating Statistical Properties of Aftershock Sequences to Earthquake Physics, *Ph.D. thesis, ETH Zurich*.

Woessner, J. & Wiemer, S., 2005. Assessing the quality of earthquake catalogues: estimating the magnitude of completeness and its uncertainty, *Bull. seism. Soc. Am.,* **95**(2), 684–698.

Woessner, J., Hauksson, E., Wiemer, S. & Neukomm, S., 2004. The 1997 Kagoshima (Japan) earthquake doublet: a quantitative analysis of aftershock rate changes, *Geophys. Res. Let.,* **31,** 3, doi:10.1029/2003GL018858.

Woessner, J., Christophersen, A., Zechar, J.D. & Monelli, D., 2010. Building self-consistent, short-term earthquake probability (STEP) models: improved strategies and calibration procedures, *Ann. Geophys.,* **53**(3), 141–154.

Woessner, J. *et al.*, 2011, A retrospective comparative forecast test on the Landers Sequence, *J. geophys. Res.,* **116**, B05305, doi:10.1029/2010JB007846.

Wohlenberg, J. & Keppler, H., 1987. Monitoring and interpretation of seismic observations in hot dry rock geothermal energy systems, *Geothermics,* **16**(4), 441–445.

Zechar, J.D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P.J. & Jordan, T.H., 2010. The Collaboratory for the Study of Earthquake Predictability perspective on computational earthquake science, *Concurrency and Computation: Practice and Experience,* **22,** 1836–1847.