COntrol SEISmicity and Manage Induced earthQuakes (COSEISMIQ)

Deliverable 6

Deliverable 6: Demonstration of the RISC decision support framework at the Hengill site in real time

Authors: ETH (Antonio P. Rinaldi, Vanille A. Ritz, Shyam Nandan, Francesco Grigoli, Stefan Wiemer)

Publication Date:

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Contents

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Summary

In this Deliverable 6 we summarize a potential use of the COSEISMIQ data for forecasting seismicity and comparing forecasting models. Such models are compared with state-of-the-art approach to calculate the information gain, with respect to a null hypothesis. The models implemented in here are tuned to forecast seismicity in the entire Hengill geothermal field and are the following: a basic ETAS model (the null hypothesis), a more advanced ETAS version (EM2, deliverable 5), and the EM1 model as implemented and calibrated in the past months (deliverable 5). For simplicity, we decided to test the models in a pseudo-forecast exercise, leading to a comparison of models via a Cumulative Information Gain (CIG). This represents a decision module to assign weights to the models based on the information gain to calculate hazard and risk forecasting (deliverable 4). Application in real-time of the current approach could be easily implemented, assuming the hydraulic and seismic data flow is correctly working.

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

RISC tool structure

The overall objective of the COSEISMIQ project was the development of an enhanced system to perform hazard risk calculation in real-time as new data are collected.

Such a high-performance system, the RISC tool, requires the assembly of several modules that handles the different operations, like processing of real-time data and forecasting the future seismicity rate (Figure 1).

A true real-time application of such a tool depends on several factors including not only scientific development, but also technological advancement and commercial value. During the COSEISMIQ period, we achieved the scientific development and demonstrated in steps on such a system may work. In previous deliverables, we have advanced our knowledge on how to perform real time detection and location of an induced seismicity field (Deliverables 1 and 2), how to calibrate complex models at both large and small scale (Deliverables 3 and 5), and how to use model results for hazard/risk calculation (Deliverable 4). With this Deliverable 6, we aim at presenting the last piece of the tool, which is the application of models for forecasting purposes and they ranking based on the probability of reproducing the data. Such a comparison can essentially occur in two ways. Given the data collected at some point in time *t*, a first option would be to run for each model a pseudo-forecasting test by running the model calibration and forecast recursively by increasing the amount of data up to *t*-1. A second option, would be to calibrate the models with all data up to time *t*, but use the past forecast up to time *t-1* to compare models with the newly collected data (form *t-1* to *t*).

In this Deliverable we present the first option by running a pseudo forecasting exercise and calculation the information gain of two advanced models with respect to a base case (the null-hypothesis).

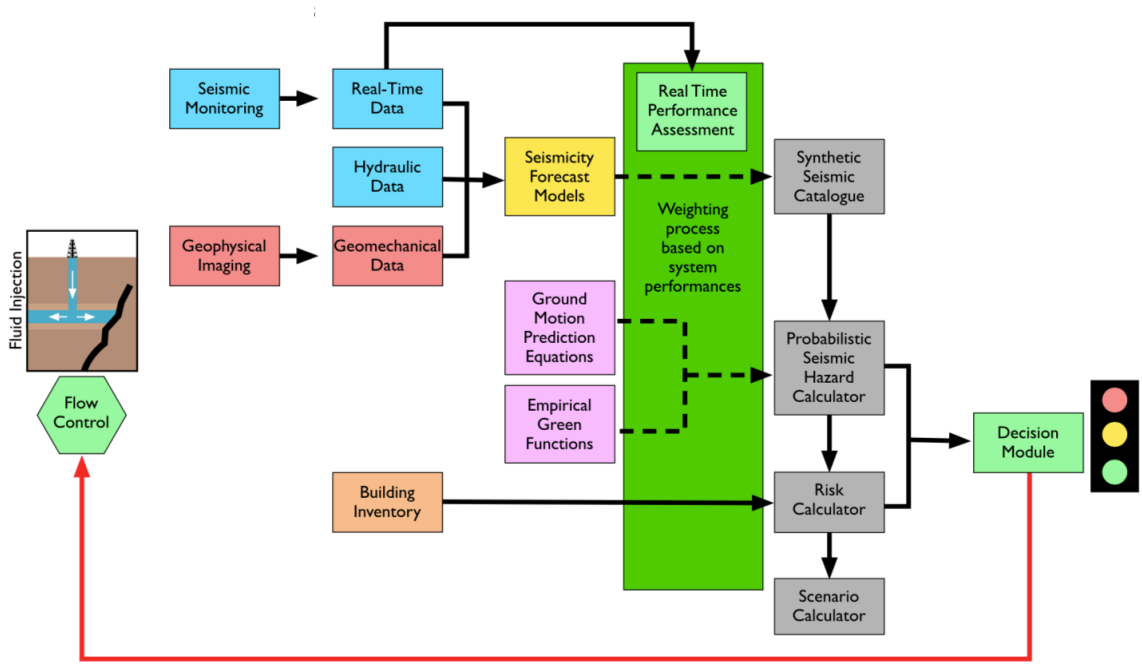

Figure 1: The RISC tool workflow (Grigoli et al., 2017). Once the data are collected in real-time and stored in easily accessible webservices (blue box see Deliverable 1/2/3/5), a centralized core is handling the retrieval of data from the webservice and running the modelling module (red box), who handles the calibration with the collected data. The modelling module may run a full pseudo-perspective test or a full calibration followed by a forecast for future times. Once model results are available, the RISC tool core can compare the models directly by using the pseudo-perspective run or just comparing the new data with past forecast. The core than assign the weights based on the information gain for hazard/risk calculation (see Deliverable 4).

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Data input

In this section we summarized the seismic and hydraulic data that have been collected during the COSEISMIQ project. The seismic data accounts for the events collected at the COSEISMIQ stations in the Hengill area, and are available via standardize FDSN webservice. The hydraulic data were manually provided by the Icelandic partner OR.

Seismic catalogue

The available seismic dataset and how it was produced is summarized in Deliverable 5. Here, we use the final catalog for the period December 2018 to January 2021. More data will be available in the future, up to August 2021 when the high-density seismic network was taken down. For the purpose of the deliverable, such addition to the dataset was deemed unnecessary, as not hydraulic data is available for the year 2021.

Figure 2 shows the whole seismic dataset, with events occurring in the entire region. Each event is classified according to a score (see Deliverable 2/5). For the current deliverable we decided to use an evscore > -5 to account for a more complete dataset (middle panel in Figure 2).

A statistical analysis of the whole dataset shows that the overall magnitude of completeness (Mc) is 0.3 (Figure 3a), with a b-value of 0.93 (Figure 3b). For all models, we assume the Mc=0.3 and filter the catalog above this magnitude.

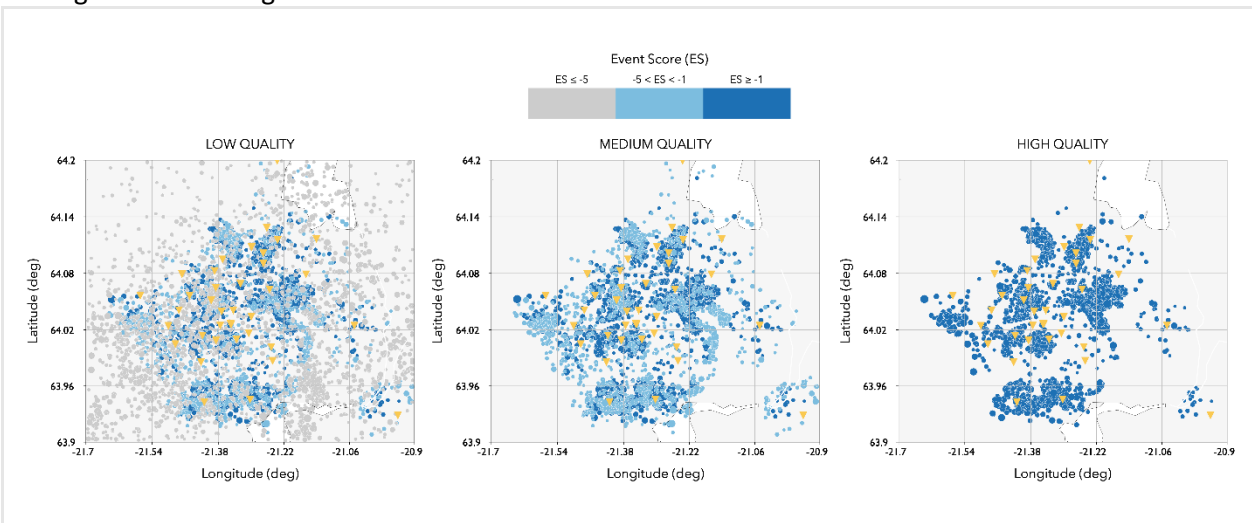

Figure 2: Seismic events in the Hengill region between December 1st 2018 and January 2021 sorted by event quality.

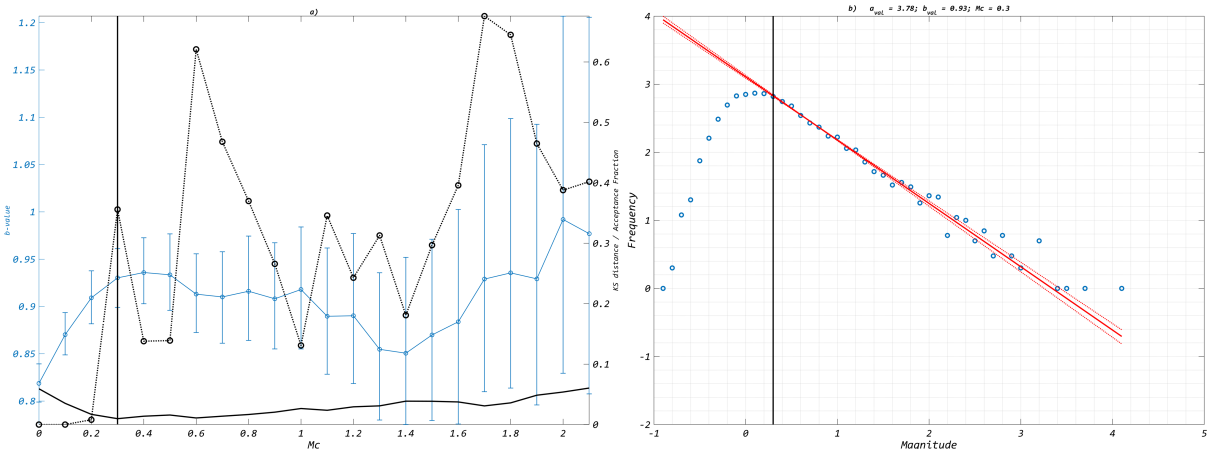

Figure 3: Statistical catalogue analysis for the determination of a-value, b-value and Mc

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Hydraulic data

Hydraulic data covering the COSEISMIQ period was obtained in December 2020. The data covers most of the field with the exception of wells in the Eastern part (Ölkelduháls region) and some wells in the Southern region of Hverahlíð. Data granularity varies from 10 minutes' intervals for the injection wells to monthly averages for production wells. Injection and production rates were assumed to remain at the level of the last recorded monthly average for forecasts in late 2020 and 2021.

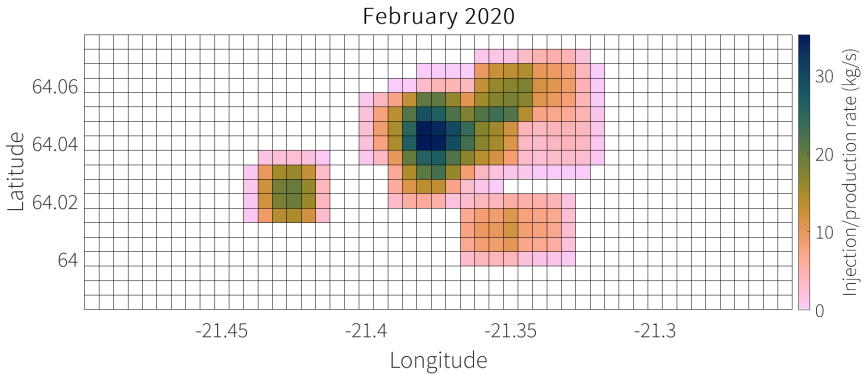

Figure 4: Injection and production rate of fluids in Hengill averages on the month of January 2020, spatially distributed with a bivariate Gaussian distribution.

Injection rates and volumes are distributed in a symmetrical bivariate Gaussian distribution around the well-head to account for fluid migration in the subsurface. The radius of the distribution (1 sigma) was assumed to be about 1 km. Both injected and produced volumes are counted positive and summed as they are to be used in models that need them. Figure 4 shows the distribution of the rates for the month of February 2020, first month for which a forecast is produced. Figure 5 shows the cumulative volume injected during the learning period spanning December 2018 to end of January 2020, i.e. the period used as learning phase for all models to perform the first forecast.

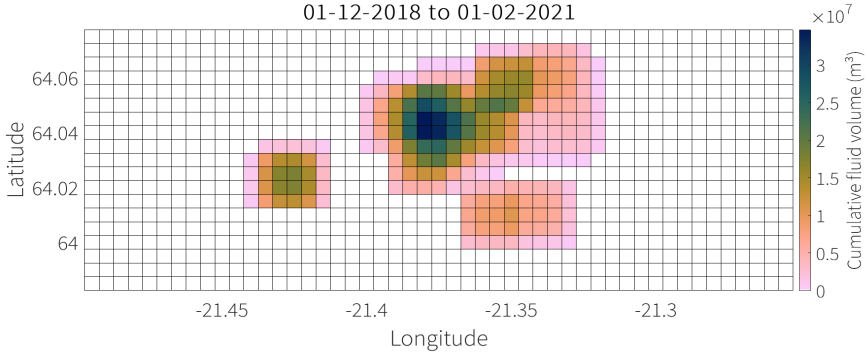

Figure 5: Cumulative volumes of injected and produced fluids in the Hengill area from December 1st 2018 to February 1st 2020.

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Seismicity Forecasting Models

EM0 – The null hypothesis

In this deliverable we implemented a new model (EM0) to act as a basis for comparison. EM0 is the simplest version of the Epidemic Type Aftershock Sequence (ETAS) model. This model views the seismicity as composed of background earthquakes and triggered earthquakes. The background earthquakes are driven by the forces of plate tectonics and anthropogenic factors such fluid injection and extraction. These earthquakes trigger a cohort of aftershocks which then trigger their own aftershocks and so on. In its simplest from the ETAS model describes the conditional seismicity rate of magnitude m events, $\lambda(t, x, y, m|\mathcal{H}_t)$, at any location (x, y) and time t as:

$$\lambda(t, x, y, m|\mathcal{H}_t) = \left[\mu + \sum_{i:t_i<t} g(t - t_i, x - x_i, y - y_i, m_i)\right] \beta e^{-\beta(m-M_0)} \quad (1)$$

where μ is the background intensity function, assumed to be independent of time, while $\mathcal{H}_t = \{(t_i, x_i, y_i, m_i): t_i < t\}$ stands for the history of the process up to time t. The variables (t_i, x_i, y_i, m_i) respectively correspond to the time, *x*-coordinate, *y*-coordinate, and magnitude of the i^{th} earthquake in the catalog, while $g(t - t_i, x - x_i, y - y_i, m_i)$ is the triggering kernel, defined in Eq. (2), quantifying the temporal and spatial influence of past events onto future events:

$$g(x, y, t - t_i, x - x_i, y - y_i, m_i) = Ke^{a(m_i-M_0)} \frac{T_{norm}\, e^{-\frac{t-t_i}{\tau}}}{\{t - t_i + c_0\}^{p_0}} \frac{S_{norm}}{\{(x - x_i)^2 + (y - y_i)^2 + de^{\gamma(m_i-M_0)}\}^{1+\rho}} \quad (2)$$

M_0 is the magnitude of the smallest event able to trigger some other ones, while T_{norm} and S_{norm} ensure normalization of the temporal and spatial components of the triggering kernel, respectively. Eq.(2) combines in a standard way the fertility law $Ke^{a(m_i-M_0)}$ giving the number of events directly triggered by an earthquake of magnitude m_i, a time kernel based on the exponentially tapered Omori-Utsu law for aftershocks and an isotropic power-law spatial kernel.

ETAS models have been quite successful in describing and forecasting the evolution of natural seismicity. As a result of their robustness, they are also being considered a strong candidate model for modelling natural seismicity. However, it is important to make a distinction between the natural and induced seismicity in the context of ETAS models. While in the former case the assumption that μ is independent of time is a reasonable assumption, the same cannot be said for the case of induced seismicity, where the time dependent factors such as the rate of fluid injection and extraction can have strong impact on μ, thereby making it time dependent itself. However, for the simplicity the null model is considered to feature a space and time dependent ETAS parameters.

EM1 – Seismogenic index model

This model is based on a purely empirical relationship between the injected volume and the seismicity (e.g. Shapiro, 2018). Other forms accounts, for example for the injection rate rather than the injected volume, as well as for an exponential decay of the seismicity after shut-in (Mignan et al., 2017; Broccardo et al., 2017). A similar model was described in Deliverables 4/5.

The proposed model provides a deterministic forecast with no flow model, and the expected number of events follow the relationship:

$$N_{expected,\geq M_c} = V(t)_{inj/prod} 10^{a_{fb}-bM_c} \quad (3)$$

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Here for simplicity do not account for the seismicity decay after shut-in (as the injection is still current) and consider the volume as injected and or produced (Figure 4-5). Given its simplicity, this model is ideal for real-time application for large scale region like the Hengill area where both injection and production activities occur.
Compared to what implemented in Deliverable 5, we improved here the minimization function, switching from a Least-Square objective function to a more complex Maximum Likelihood Estimate method (Broccardo et al., 2017), defined as:

$$\ln \mathcal{L}(\mathcal{D}|\theta) = \frac{N(a_{fb} - bM_c)}{\log e} + \sum_{n=1}^{N} \ln \dot{V}(t_n) - V(t)10^{a_{fb}-bM_c} + N\ln(b) +$$
$$N\ln(\ln 10) - b\ln(10) + \sum_{n=1}^{N} m_n - N\ln\left(10^{-bM_c} - 10^{-bM_{max}}\right) \quad (4)$$

The three model parameters are here estimated as follow: M_c is estimated from the entire seismic catalogue (Figure 2), while the productivity a_{fb} and b-value for each space bin are estimated by minimizing the log-likelihood function in the equation above with a standard Matlab algorithm(Broccardo et al. 2017):

This log-likelihood optimisation function is adapted to Hengill, where both injection and production occur continuously without shut-in.

EM2 – Advanced ETAS model

The EM2 is a variant of ETAS model which features spatially variable background rate. All other aspects of EM2 are like EM0. However, even though the background rate is spatially heterogeneous, the model still assumes that it is invariant in time.

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Pseudo-forecasting: results and consistency tests

In this section, we summarized the results of the pseudo-forecasting exercise. We take as a base Learning Period (LP) all the data recorded up to Feb. 1st, 2020. Using this data, we calibrate all the competing models. Then we perform a forecast for the next 30 days and update the LP with the additional information. We do this process of training and forecasting repeatedly until the end of the dataset, which for this deliverable is set to end of January 2021. Below the results for each individual model and we also computed for each model the so-call "N-test" by comparing the number of simulated events with the observation.

EM0 – The null hypothesis

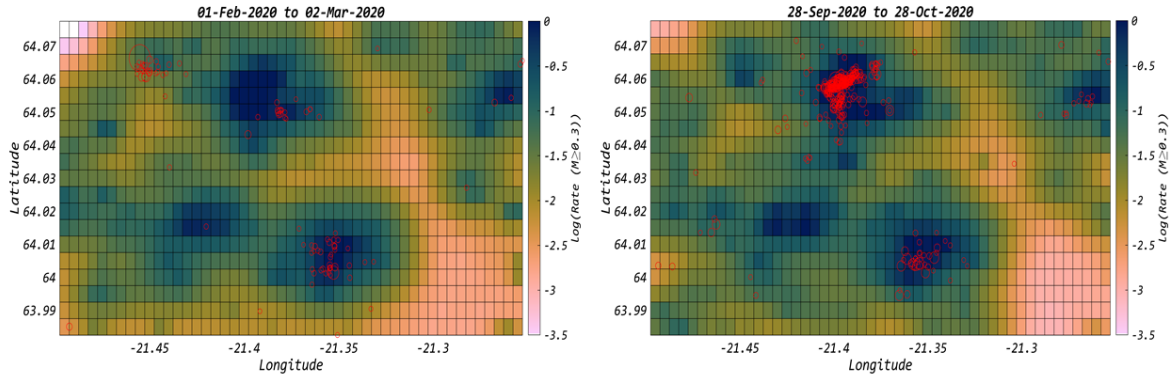

Figure 6: Mean rate forecasted by the EM0 for the first testing period [1 Feb, 2020 to 02 March, 2020] (left panel) and during [28 Sep, 2020 to 28 Oct, 2020] (right panel); Red circles, whose size scale with magnitude show the location of $M \geq 0.3$ earthquakes that occurred during the forecasting period

Figure 6 shows the spatial distribution of forecasted rates for two arbitrarily chosen forecasting periods (1st of February to 2nd of March and 28th of September to 28th of October 2020). For both periods, the forecast highlights regions of reinjection (Húsmúli in the North and Gráuhnúkar in the South-West) as well as the very active production region of Hverahlíð in the South. EM0 also forecasts seismicity in the Eastern part of the study region (Ölkelduháls) where a mix of natural seismicity and injection/production is known to occur.

The N-tests for the corresponding forecast periods is shown in Figure 7, where the distribution of forecasted number of events is compared to the actual recorded number of events in the period. We see that in the first chosen period, the forecast and record are in agreement (black line within the 95% confidence interval of the forecasted distribution). However, the second period in September-October 2020 fails as many more events are recorded compared to the forecast.

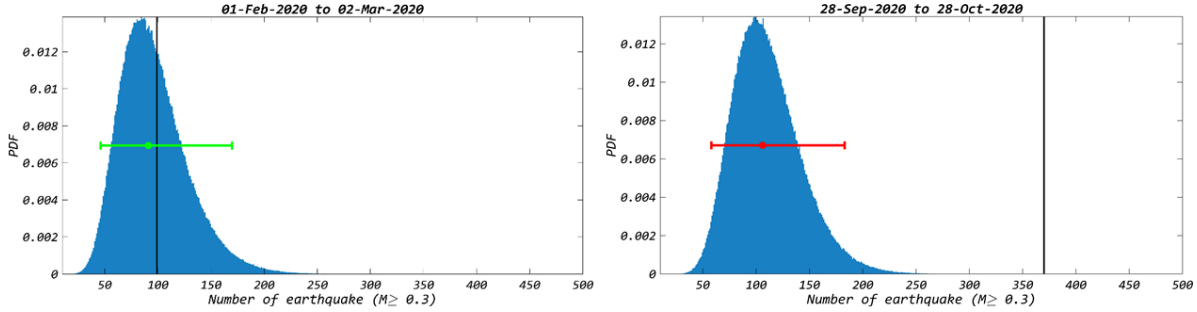

Figure 7: Distribution of total number of earthquakes forecasted in the study region by the EM0 model for the first testing period [1 Feb, 2020 to 02 March, 2020] (left panel) and during [28 Sep, 2020 to 28 Oct, 2020] (right panel); Horizontal error-bar indicate the 95%ile of the number distribution; Vertical black line shows the total number of earthquakes observed during the testing period; if the black line is with in the 95% interval of the forecasted numbers, the forecast is considered to be consistent (green error-bar) otherwise it is considered to be inconsistent (red error-bar).

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

EM1 – Seismogenic index model

Figure 8 shows the spatial distribution of forecasted rates for the same two arbitrarily chosen forecasts as EM0. The forecast for EM1 is only provided in cells where volume is injected or produced, leading to an incomplete map. Comparison with EM0, and later with EM2, is only valid in those grid cells where EM1 forecasts a rate.

Similar to EM0, EM1 forecasts high seismicity rates in regions of reinjection (Húsmúli in the North and Gráuhnúkar in the South-West) as well as the very active production region of Hverahlíð in the South but fails at assigning rates to the Ölkelduháls region as we do not have injection/production data for the wells located there. The forecast for the Southern production region (Hverahlíð) is quite interesting as the fit between data and model is satisfactory even though at least two active wells are missing from our modelling. This suggests the flexibility of the Seismogenic Index model to adapt and optimize parameters even with incomplete hydraulic data, assuming such incompleteness is occurring for the entire period and that the missing information is for well in the same region.

The N-tests for the corresponding forecast periods is shown in Figure 9. Worth mentioning that EM1 is a frequentist model, hence we assume a Poisson's distribution with a mean same as the forecasted number of events. The number of recorded events per period is calculated by summing records only in the cells where EM1 provides a forecast leading to a slightly different number of occurred events for the same forecast between the ETAS models and EM1. Similar to EM0, EM1 passes the N-test for the first forecasting period and fails the September-October forecast. In this case we use a further information, which is the injection rates. The high number of recorded events in this period (clustered around the Húsmúli reinjection area) is not explained by the hydraulic data at our disposal but could correspond to the drilling of a new well, a surge in injection rate in the HN wells or changes in the injection protocol in Húsmúli.

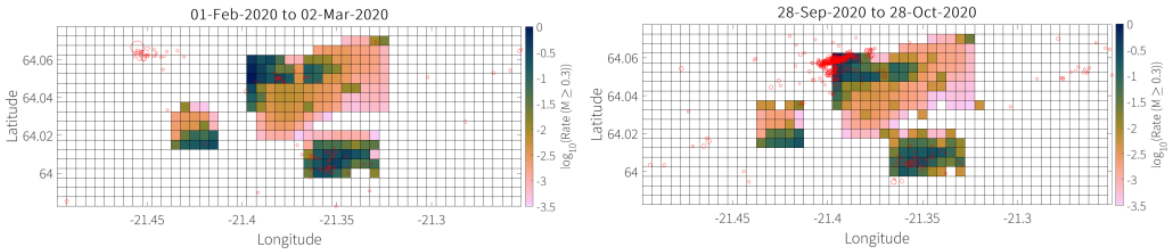

Figure 8: Mean rate forecasted by the EM1 for the first testing period [1 Feb, 2020 to 02 March, 2020] (left panel) and during [28 Sep, 2020 to 28 Oct, 2020] (right panel);

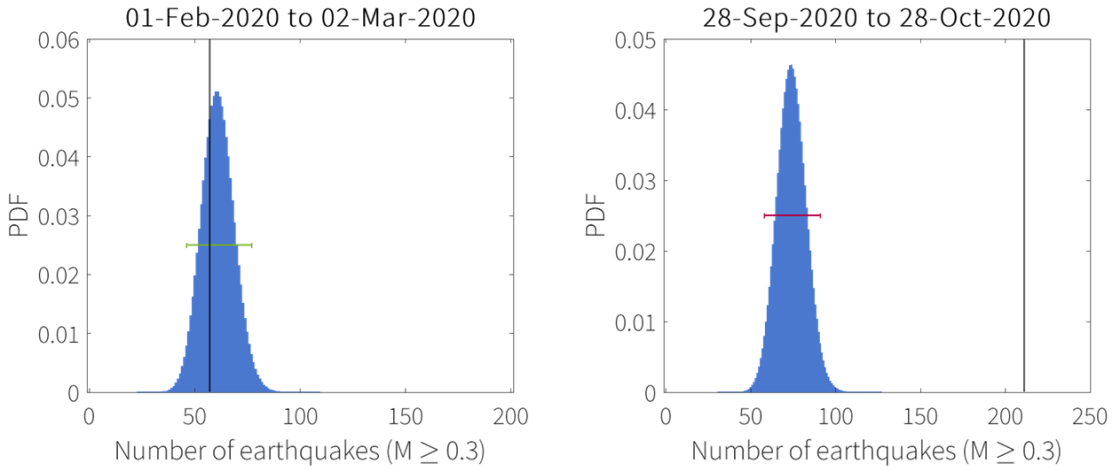

Figure 9: Distribution of total number of earthquakes forecasted in the study region by the EM1 model for the first testing period [1 Feb, 2020 to 02 March, 2020] (left panel) and during [28 Sep, 2020 to 28 Oct, 2020] (right panel)

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

EM2 – Advanced ETAS model

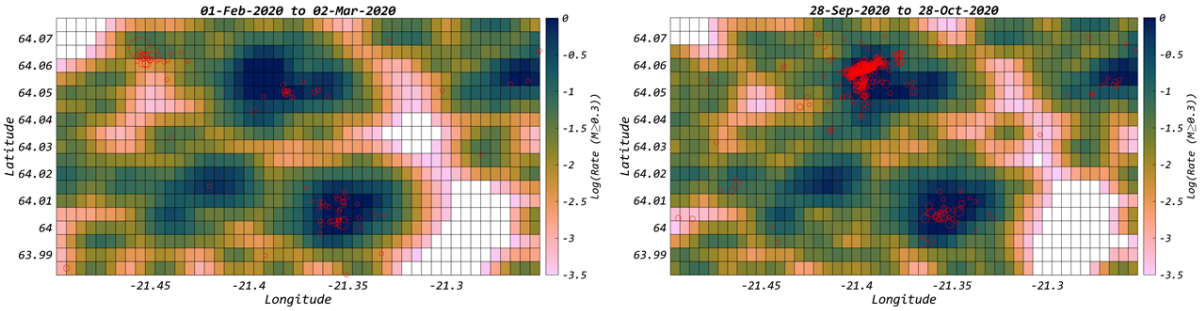

Figure 10: Mean rate forecasted by the EM2 for the first testing period [1 Feb, 2020 to 02 March, 2020] (left panel) and during [28 Sep, 2020 to 28 Oct, 2020] (right panel);

Figure 10 shows the spatial distribution of simulated rates for the same two forecasts. Similarl to EM0, EM2's forecast highlights regions of reinjection (Húsmúli in the North and Gráuhnúkar in the South-West) as well as the very active production region of Hverahlíð in the South and the Ölkelduháls region. With the addition of the spatially variable background rate, EM2's forecasts differ slightly from EM0, especially in the areas with lower seismicity rates.
The N-tests for the corresponding forecast periods is shown in Figure 11. As the other models, EM2 passes the N-test for the first forecasting period and fails the September-October forecast.

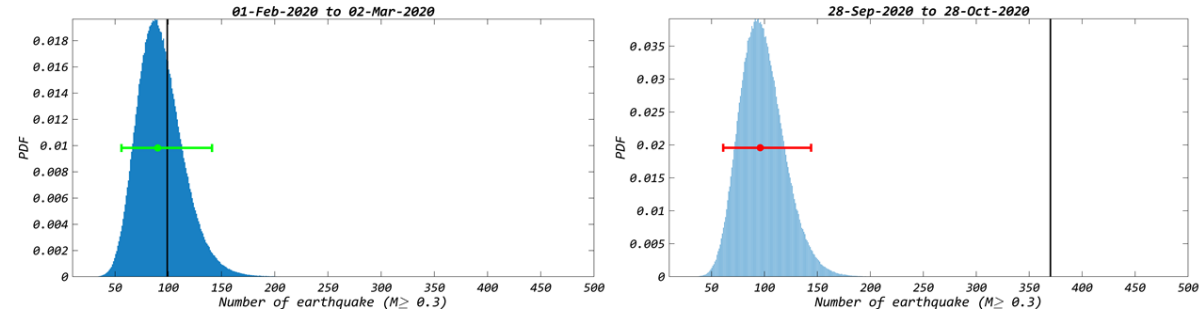

Figure 11: Distribution of total number of earthquakes forecasted in the study region by the EM2 model for the first testing period [1 Feb, 2020 to 02 March, 2020] (left panel) and during [28 Sep, 2020 to 28 Oct, 2020] (right panel)

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

Model Comparison and information gain

In order to compare the models and compute the information gain, we calculate for each forecast (12 in total) the Log-Likelihood for each space bin in our computational domain. Such a Log-Likelihood is simply assumed to be the Poissonian for EM1, while exploit the full probability distribution for the ETAS models (EM0 and EM2). Figure 12 shows the map of the log-likelihood calculated for each model for all cells of the grid. For EM1, the log-likelihood is only computed in the cells where an event rent can be forecasted, excluding the regions counting only natural seismicity. The colormap is tapered to white close to zero.

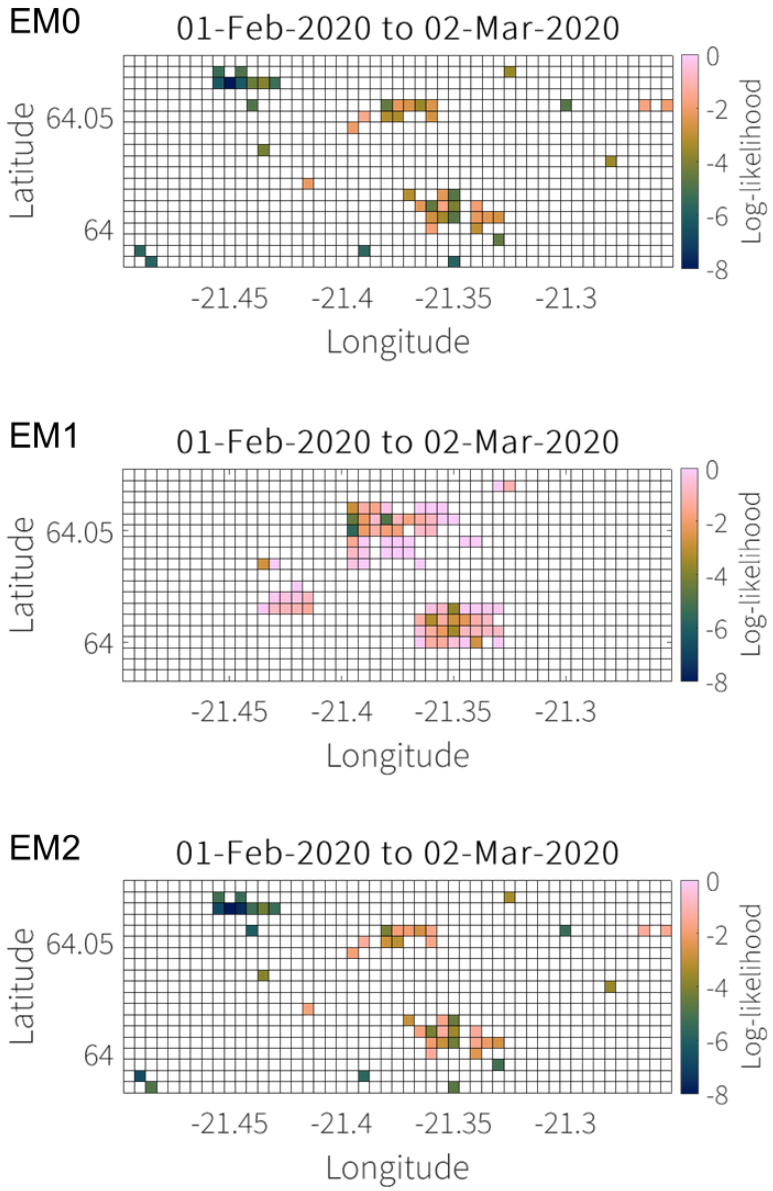

Figure 12: Log-likelihood map for each model calculated for the first forecasting period. For model EM1, the Log-likelihood is computed only for regions with injection/production wells.

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

The information gain is calculated for both EM1 and EM2 with respect to the null hypothesis (EM0), and for each forecast period (12 in total, each of 30 days). We have for each model a time-series illustrating how the information gain may change as new data are included in the LP. Figure 13 allows us to directly compare the models and providing then the basis for a weighing scheme.
The information gain for EM1 is calculated with the log-likelihood of EM0 summed only in the regions where EM1 forecasts a rate whereas the information gain for EM2 takes advantage of the whole study region. Both models show limited information gain compared to the null hypothesis, with the advanced ETAS model being slightly better for the first forecasts, and the EM1 model sometimes performing better. Both EM1 and EM2 perform significantly worse in the forecasts corresponding to the periods with high seismic rates and failed N-tests. This means that an the models are less efficient than EM0 in handling sudden changes in the seismicity rate, although it must be mentioned that also EM0 fails the N-tests for the forecasts in the late period of the dataset (see Figure 7).

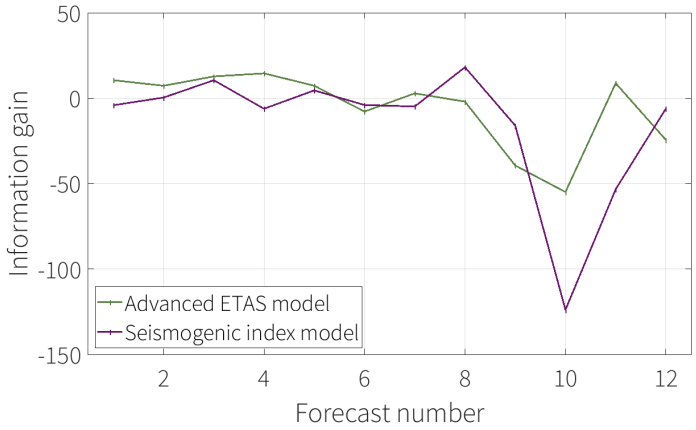

Figure 13: Information gain for EM1 (purple) and EM2 (green) - calculated relative to EM0 the null hypothesis. For the comparison between EM1 and EM0 we account for the log-likelihood computed only for the cells with active injection/production wells.

Future improvements and applications

Here we summarize and discuss the problems with running real-time systems and the future improvement in model comparison.

- The models are strictly linked to the availability of having data flowing in real-time. In the context of COSEISMIQ, while the seismicity data were promptly available in real-time, the hydraulic data real-time service was not implemented due to limitation in accessing such a dataset. In this sense, models like EM0 and EM2 are easier to be implemented as they rely only on the recorded seismicity.
- At the time of writing this deliverable, even assuming the hydraulic data were available, the seismic network has been taken down, hence even the models relying only on seismicity data cannot be operated in real time (too much different quality of the seismic catalogue at the present compared to when the dense Hengill temporary 2C network was available).
- The inclusion of more sophisticated models as the ones described in Deliverable 3 and Deliverable 5 is the goal of future development. Regarding the entire Hengill field, it is impossible to calibrate a full model given the geothermal area size. However, applications at smaller scale are possible (e.g., comparing the models in the reinjection area of Húsmúli).
- The weighting of models will be performed based on the model performance compared to a null hypothesis.
- The integration of the seismicity rates outputs of empirical models into the hazard modelling follows the probabilistic framework laid out in Deliverable 4 and Broccardo et al., 2019.

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117

- Further empirical models will also be tested and developed in future projects. For example, the inclusion of a forcing term in ETAS models that could account for the collected hydraulic data. Potentially even the implementation of new hybrid models is possible, for example coupling well know seismological modelling techniques (ETAS, Rate-and-State) to deterministic calculation (i.e., numerical modelling of fluid flow).
- Potential empirical model could even be simpler, and accounts for b-value variability as linked to the fluid injection/production rate.
- The lessons learned in developing the RISC system will be of extreme importance for future project. One example is the DEEP project, where all the current developments will be re-evaluated and applied to the FORGE EGS site.

References

Broccardo, M., Mignan, A., Wiemer, S., Stojadinovic, B., & Giardini, D. (2017). Hierarchical Bayesian modeling of fluid-induced seismicity. *Geophysical Research Letters*, *44*(22), 11-357.

Broccardo, M., A. Mignan, F. Grigoli, D. Karvounis, A. P. Rinaldi, L. Danciu, H. Hofmann, C. Milkereit, T. Dahm, G. Zimmermann, V. Hjörleifsdóttir, S. Wiemer, (2020), Induced seismicity risk analysis of the hydraulic stimulation of a geothermal well on Geldinganes, Iceland, *Natural Hazards and Earth System Sciences*, 20, 1573-1593.

Grigoli, F., S. Cesca, E. Priolo, A. P. Rinaldi, J. F. Clinton, T. A. Stabile, B. Dost, M. Garcia Fernandez, S. Wiemer, T. Dahm, (2017), Current challenges in monitoring, discrimination and management of induced seismicity related to underground industrial activities: a European perspective. *Reviews of Geophysics*, 55, 310-340.

Mignan, A., Broccardo, M., Wiemer, S., & Giardini, D. (2017). Induced seismicity closed-form traffic light system for actuarial decision-making during deep fluid injections. *Scientific reports*, *7*(1), 1-10.

Shapiro, S. A. (2018). Seismogenic index of underground fluid injections and productions. *Journal of Geophysical Research: Solid Earth*, *123*(9), 7983-7997.

The GEOTHERMICA is supported by the European Union's HORIZON 2020 programme for research, technological development and demonstration under grant agreement No 731117