Draft Statistical Note: Model comparison for activity rate models

Matthew Jones, Rakesh Paleja, Tim Park (Shell Global Solutions UK)
Stijn Bierman (Shell Global Solutions NL)

October 28, 2015

1 Introduction

It is likely that over time several different models will be proposed to explain the spatio-temporal variability in seismic activity rate in the Groningen gasfield. Our definition of seismic activity rate is the rate of occurrence of earthquakes. In such a scenario it is desirable to be able to compare the ability of each of the possible models to explain the observed rate of occurrence of events as recorded in the event catalogue. The framework described in this note may be used to assess the relative ability of each models to predict data which has not been used to fit (estimate) the model.

2 Model Comparison

This section outlines the model comparison framework. We first describe a general likelihood for a proposed activity rate model. We then describe how a proposed model can be fitted to the available data and finally compared against other models.

2.1 General likelihood

All of the models fitted to the event data use a spatio-temporal Poisson point-process (SPP) likelihood, though different models will use different covariates and parameter specifications. Using $r\left(x,y,t\right)$ to represent a vector of covariates specified as a function of space and time, the likelihood of events $\{D=d\left(x_i,y_i,t_i\right)\}$ occurring at locations (x_i,y_i) at times t_i is as follows

$$L(D|\lambda\left(r\left(x,y,t\right),\theta\right))=\exp\Big[-\int_S\int_T\lambda\left(r\left(x,y,t\right),\theta\right)dtdS\Big]\prod_{i=1}^{n}\lambda\left(r\left(x_i,y_i,t_i\right),\theta\right). \tag{1}$$

Where S represents the surface area of the entire Groningen field, T is the time that has elapsed to date. The rate $\lambda\left(.\right)$ is assumed to be a function of a set of parameters θ as well as of the covariates; these parameters will be estimated from the data (current event catalogue), making use of any prior beliefs $g\left(\theta\right)$ that exist about their values.

2.2 Parameter estimation

Given an event catalogue D and a functional form for the relationship between the rate and the covariates and parameters, we first wish to estimate the model parameters from the data. We adopt a Bayesian approach, and first write down the posterior for the model parameters,

$$p(\theta|D,\lambda\left(.\right),r\left(.\right))\propto L(D|\theta,\lambda\left(.\right),r\left(.\right))g(\theta), \tag{2}$$

where $g(\theta)$ is the prior on the model parameters. To obtain parameter estimated from the posterior, we use Markov chain Monte-Carlo (MCMC) methods; specifically, the Metropolis-Hastings algorithm. Note that since we assume the SPP likelihood for all models, a model choice simply consists of a choice of covariates $r\left(.\right)$ and a choice of functional relationship between these and the activity rate $\lambda\left(.\right)$ (using model specific parameters θ); we therefore write $\mathcal{M}=\{\lambda\left(.\right),r\left(.\right)\}$ to denote a particular model choice.

2.3 Prediction of future events

The eventual goal of the analysis of seismic activity in the Groningen field is to find a model which gives good predictions for the locations of future events; therefore, we wish to compare models based on their predictive performance. We use $\mathcal{M}$ to denote a particular set of model assumptions (the specification of a rate function, set of parameters, and prior distributions for the parameters); under this model, the likelihood of a set of unobserved events Y at particular locations and times, given the current event catalogue, can be calculated as follows

$$L(Y|D,\mathcal{M}) = \int L(Y|\theta,\mathcal{M})p(\theta|D,\mathcal{M})d\theta. \tag{3}$$

where the first term under the integral is the likelihood (1) under a particular model $\mathcal{M}$, and the second term is the posterior distribution of the unknown parameters in that model, which can be calculated up to normalization by application of Bayes theorem using 2 Because of the form of the SPP likelihood, we cannot sample from either of these distributions directly; we must instead use MCMC methods to characterize them.

2.4 Cross-validation

In order to gain some idea about how the model may perform when predicting future events, we examine how well it would have been able to predict sub-sets of the observed events using information obtained from model assumptions and the remainder of the event catalogue. This process is known as cross-validation. We split the period of the event data into n_T spatio-temporal intervals, which we index by $\mathcal{T}_j$, $j = 1,\ldots,n_T$: this may involve blocking events in time (eg: into individual years), blocking in space only (eg: splitting the field into sub-regions), or blocking in both space and time (some combination of both). For each of these blocks, we then compute the predictive likelihood (3) of the omitted events given the remaining events, under a given set of model assumptions ($L(D_{\mathcal{T}_j}|D_{\mathcal{T}_{\neg j}},\mathcal{M})$). We repeat this for all blocks, and obtain the following quantity as our comparison measure between two models $\mathcal{M}_A$ and $\mathcal{M}_B$

$$C\left(\mathcal{M}_A,\mathcal{M}_B\right) = \prod_{j=1}^{n_T} \frac{L(D_{\mathcal{T}_j}|D_{\mathcal{T}_{\neg j}},\mathcal{M}_A)}{L(D_{\mathcal{T}_j}|D_{\mathcal{T}_{\neg j}},\mathcal{M}_B)} > 0. \tag{4}$$

Under this comparison criterion, a value of $C\left(\mathcal{M}_A,\mathcal{M}_B\right) < 1$ implies that, under the given blocking strategy, the predicted events are more likely under model $\mathcal{M}_B$ than under model $\mathcal{M}_A$, implying that model $\mathcal{M}_B$ gives a better fit to the data, whereas $C\left(\mathcal{M}_A,\mathcal{M}_B\right) > 1$ implies the converse. In practice, it is often preferable to work with with $\log C\left(\mathcal{M}_A,\mathcal{M}_B\right)$, since this turns the ratio (4) into a difference, centres the scale of the measure at 0 and gives comparable scales above and below zero.

For ease of interpretation, when comparing large numbers of models within this framework, rather than comparing all individual pairs of models, we will choose a particular model to be the baseline for the comparison, and compare all other models to this one; typically, our baseline model should be one from which we expect to see poor predictive performance. In section 3, we choose to use as the baseline a model which assumes a constant rate over all of space and time. When comparing models on this scale, we can easily assess the relative performance of any given pair of models by noting that

$$C\left(\mathcal{M}_A,\mathcal{M}_C\right) = \frac{C\left(\mathcal{M}_A,\mathcal{M}_B\right)}{C\left(\mathcal{M}_C,\mathcal{M}_B\right)}.$$

which means that on a log scale, $\log C\left(\mathcal{M}_A,\mathcal{M}_C\right) = \log C\left(\mathcal{M}_A,\mathcal{M}_B\right) - \log C\left(\mathcal{M}_C,\mathcal{M}_B\right)$.

3 Example

In this section, we illustrate the model comparison framework by comparing three models; a baseline model which assumes a constant activity rate; a model which assumes a log-linear relationship between time and activity rate and a model which assumes a log-linear relationship between compaction and activity rate. **We note that these models are hypothetical. The motivation is to show how the model comparison framework will work rather than proposing a particular model**. These hypothetical models are described in more detail below.

3.1 Constant (Baseline model)

We choose a very simple model as our baseline against which we compare the two other models. In this case we choose a constant activity rate model.

$$\lambda\left(c\left(x,y,t\right),\beta\right) = e^{\beta}$$

where $c(x,y,t)$ represents compaction and β is the parameter estimate.

3.2 Log-Linear Temporal Model

For the second model we assume that the activity rate only depends upon time and has a log-linear relationship,

$$\lambda\left(c\left(x,y,t\right),\{\beta_0,\beta_1\}\right) = e^{\beta_0+\beta_1 t}$$

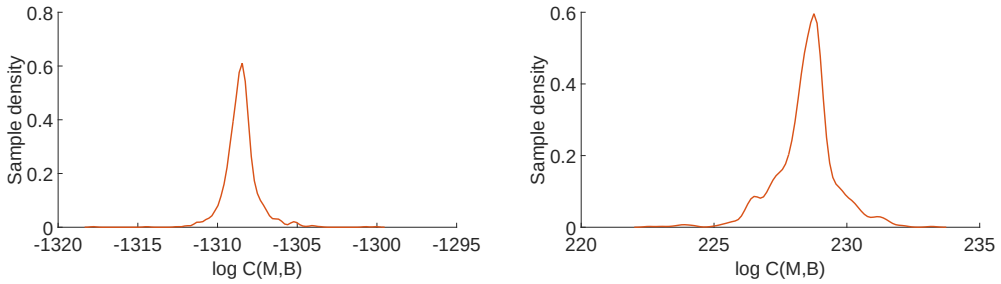
3.3 Log-Linear Compaction Model

For the final model we introduce compaction as a covariate. This allows the activity rate to vary both spatially and temporally,

$$\lambda\left(c\left(x,y,t\right),\{\beta_0,\beta_1\}\right) = e^{\beta_0+\beta_1 c(x,y,t)}$$

3.4 Results

For all models we estimate the parameters using the Metropolis-Hastings sampler outlined in section 2.2. We are then able to calculate the posterior density of the comparison criterion outlined in 2.4. The results of this for the two models are shown in Figure1.



(a) Temporal Model

(b) Compaction Model

Figure 1: Comparison of Models to Baseline

Figure 1(a) shows the comparison between the temporal model and the baseline. It is clear from this figure that the log of the comparison criterion is significantly below zero indicating that this model performs worse than the baseline. In contrast Figure 1(b) shows that the compaction model performs significantly better than the baseline as the log of the comparison criterion is significantly greater than zero. It then follows that the log-linear compaction model performs significantly better than the log-linear temporal model.

4 Prerequisites

The model comparison framework firstly requires a discussion amongst various stakeholders on the relevance and the key performance indicators to evaluate different models. Once a clear need is agreed and defined, it is also required that a set of rules are described for proposing models that need to be compared using this framework. Stakeholders need to agree on the data sets for model calibration and evaluation of model performance. The proposed framework is suitable for models which can be parameterised using part of the event catalogue and predict a likelihood for the remainder of the event catalogue. In this note we have illustrated how the predictive performance of any number of models can be compared if the key performance indicator is the ability to predict spatial and temporal variability in rates of event occurrence. In the obvious manner, comparisons of any

number of models using $C(\mathcal{M}_i, B)$ (for base model B and alternative models M_i, i=1,2,3,...) is straightforward. Moreover, comparison of sample densities for $C(\mathcal{M}_i, B)$ allows us to assess the extent of the difference in predictive performance of competing models.