Motivation:

- What *is* forecasting? (in this project)
  - NOT prediction/alarms
  - produce probabilistic description of eqk occurrence in a specific bin
    - poisson rate
  - There is predictability, particularly given clustering
    - Building towards more powerful models and strategies.

- What can we actually forecast?
  - Initial event ("mainshock") is probably beyond our reach at this point.
  - Series of events following a mainshock ("aftershocks").
    - Aftershocks can occur months/years after initial event
  - These are not insignificant
    - often aftershocks most deadly.
    - Even if not, still dangerous & damaging.
  - Probabilistic description therefore provides meaningful info…

- What can we do with this?
  - Information for immediate response
    - Is it safe to send ER workers into damaged buildings?
    - Evacuate weak buildings/bridges/etc…
  - Short-term planning for authorities and individuals
    - Stand by ER workers for x days
    - Large events – safe to continue?
    - Tourists – evaluate personal risk threshold, decide for themselves.
  - Long-term planning and risk management
    - Compute hazard/risk
    - Produce emergency plans
    - Set targets for emergency funds/investment in relevant resources.

- PROJECT AIMS

Our Data - Canterbury Eqk Sequence, NZ:

- Powerful eqk and aftershocks centred on/near Christchurch, NZ
  - Largest South island city, ~400K residents.

- Sept 2010 – Dec 2011

- Why are we using this sequence?
    - Complex – 4 distinct and significant events spread over more than 1 year
    - Well documented – wealth of data for studying eqk clustering and predictability.
    - Demonstrates the importance of aftershock forecasting

- Really demonstrates that eqks are not isolated events, and aftershocks do cluster and can follow predictable patterns.

- Our data begins immediately after Darfield M7.1

<u>Experiment Design:</u>

Suppose we have a model:

- Feed it data:
    - Real time vs best available
    - Outputs poisson rates

How do we test it?

- Compute likelihood against a catalogue of observed eqks.
    - Catalog gives number of observed earthquakes per spatial region, per magnitude bin (ABOVE M3.95), per month.
    - THIS ASSUMES POISSON AND INDEPENDENT!!
    - Our catalog is from CSEP

How do we compare models?

- Just compare their likelihoods!
- Use a metric – one good one is probability gain, just a measure of the difference in likelihood per earthquake

<u>Base Models:</u>

There are 3 types that we use:

- Physical – modelling strain and stress within the earth's surface & along fault lines
- Statistical
    - Statistical clustering
    - Smoothing
- Hybrid

Our portfolio consists of 15 total models

- 5 physical
- 6 statistical
- 4 hybrid

Ensembling:

Take several 'base models' and merge them – usually some kind of weighted avg, which may change dynamically with time.

- Advantages over selection
    - Best model might outperform ensemble but hard to choose reliably – might be catastrophically wrong, or the best model might be inconsistent.
    - Objective and transparently merging therefore better
    - Can even outperform best base model
        - Merging models of different kinds (eg physical with statistical) might strengthen their fortes and minimise their weaknesses

The challenge is then finding the right weights!

- We tend to use likelihoods to weight by past predictive skill.
    - Problematic – as with stocks/other time-dep systems, past behaviour is no guarantee of future behavious!
    - But it seems to work.

- Recent paper claims to show that a multiplicative approach to merging produces better results than an additive approach.
    - "information gains of the best multiplicative ensembles are greater than those of additive hybrids constructed from the same models."

- We try to exploit this finding to construct a better ensemble than existing ones, by means of log-linear pooling.

Log-Linear Pooling:

This is the model we developed during the project.

- Combination strategy – multiplicative

- How do we choose weights?
    - Which values of w_j would have been the best up until now?
        - Best = likelihood
        - Find best using non-linear optimisation

Existing Ensembles:

What are we going to compare our new ensemble to?

- Bayesian Model Averaging (BMA)
    - Weights proportional to posterior likelihood of each model
    - Classical and well understood model – good to test against.

- Score Model Averaging (SMA)
    - Weights proportional to inverse of LOG-likelihood

- Generalised SMA (gSMA)
    - Weights proportional to inverse of difference between the LOG-likelihood of one model and the likelihood of the model with the biggest LOG-likelihood

- Parimutuel Gambling SMA (PGSMA)
    - Rather different, “gambling” approach
    - Weights based on Parimutuel Gambling Score of each model (“mutual betting”)
        - All bets placed together in a pool, and payoff odds are calculated by sharing the pool among all winning bets.
    - Alpha = 1/maximum loss among all models
    - V_i = Parimutuel Gambling score of model i

<u>Results:</u>

- Quite active in promoting/supressing models.

- Changes quite significant – very different weightings by the end.

    - Reflecting base models changing skill?
    - Or just more info towards the end?

- Real-time vs Best-available is quite different – clearly the model is picking up on something in the BA.
    - Question – is BA any better?
    - If not, RT is preferable

<u>Comparison to Existing Models:</u>

- Quick comparison to show different behaviours
    - Only comparing 2 because others very similar

- gSMA similarly changeful – heavily emphasises past performance

- PGSMA much more cautious in changing weightings

Performance Ranking RT:

Likelihood comparison (SMALLER BAR IS BETTER)

- No ensemble beats best base model (ETAS)

- But our model is closest!

Probability Gain

Performance Ranking BA:

Likelihood comparison (SMALLER BAR IS BETTER)

- This time loglik does beat best base model

- Slightly worse than other ensembles

Probability Gain

- Significant improvement over most base models

- Very close to best ensemble

Discussion & Implications

- Wanted to see if multiplicative model was fruitful
  - Not head-and-shoulders better, but competitive.
  - First effort – further improvements could yield better results
  - Significantly slower than others (hours vs minutes)
    - Again not optimised much so maybe ok.

Machine Learning:

- Initially wanted OptimLogLinPool kind of side project.

- Machine learning = data + desired output → program
  - Can then give it new unseen data to work on.
- Brainstorming led to “prototype machine learning model”
  - OptimLogLinPool provides ‘optimal weights’
  - Give THESE + catalog to machine learner as data, allow IT to work out connection between optimal weights and observed earthquakes
- Ultimately, not enough data – weights are 15x20 dataset (300 datapoints). Nowhere near sufficient to get anything meaningful.
  - Abandoned it in favour of more promising above approach.