Calculation Method for Large Seismic Hazard Models and the 2022 Revision of the New Zealand NSHM

Chris DiCaprio, Chris Chamberlain, et. al. GNS Science, New Zealand

Te Tauira Matapae Pūmate Rū i Aotearoa
NSHM The New Zealand National Seismic Hazard Model
A GNS Science Led Research Programme

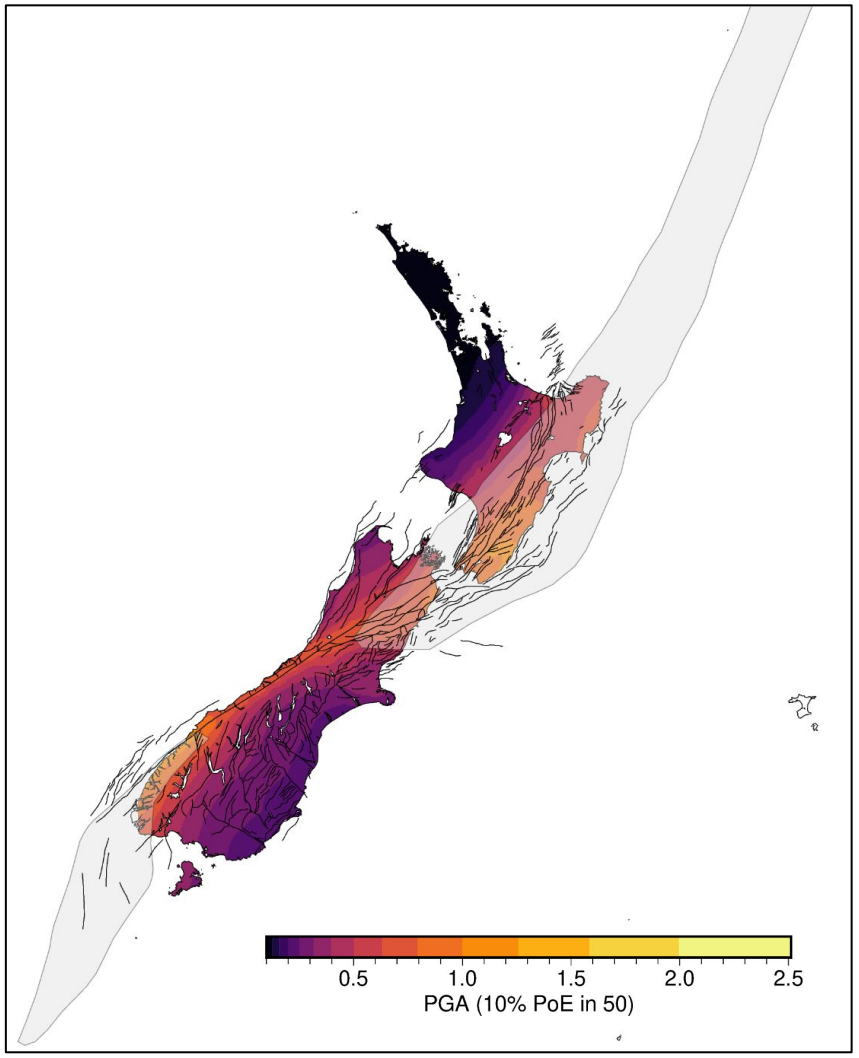
E mahi ana me
In collaboration with

Ngā hoa tuku pūtea
Funding partners
MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI
Toka Tū Ake EQC

Outline / Summary

- **The NZ NSHM 2022 model is quite large for an NSHM and was initially not possible to calculate**
  - Capturing epistemic uncertainty was an important feature of NZ NSHM 2022
- **Our approach:**
  - decompose the logic tree into independent parts (component branches)
  - Store those realizations in a database (available in the cloud or on local disk)
  - *a posteriori* assemble realizations for all branches of logic tree and calculate aggregate statistics
- **Use of cloud computing and database technologies**
  - New portable DB format
- **Brief examples of application**

The New Zealand NSHM Revision 2022: Size of the Task



- 324 (SRM) x 3024 (GMCM) = 979,776 total branches
- 3774 sites (0.1 deg calculation grid)
- 16 spectral periods
- 18 site conditions (Vs30)

Initial tests using OQ v3.17 were hitting time and memory bounds. **We needed an alternate approach to calculating the hazard curves.**

Assuming we were able to calculate the model, we needed a way to easily and rapidly do so

Break the Problem Down . . .

Break problem into smaller independent parts and calculate the solution in 2 stages

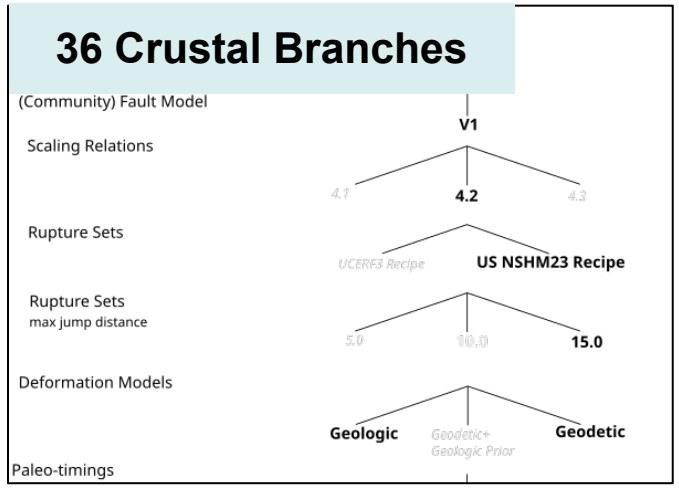
- Manageable sized parts
- Parallelization

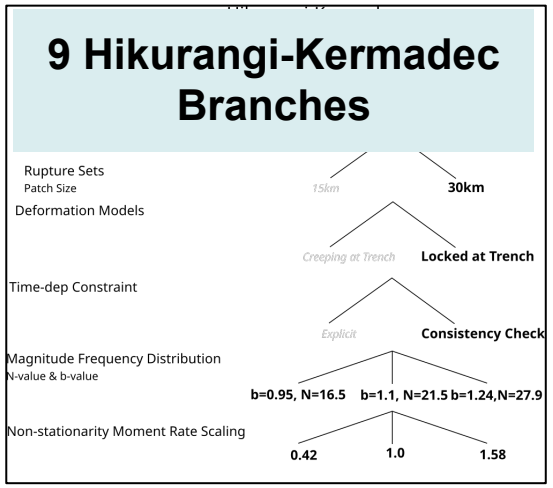
Stage 1:

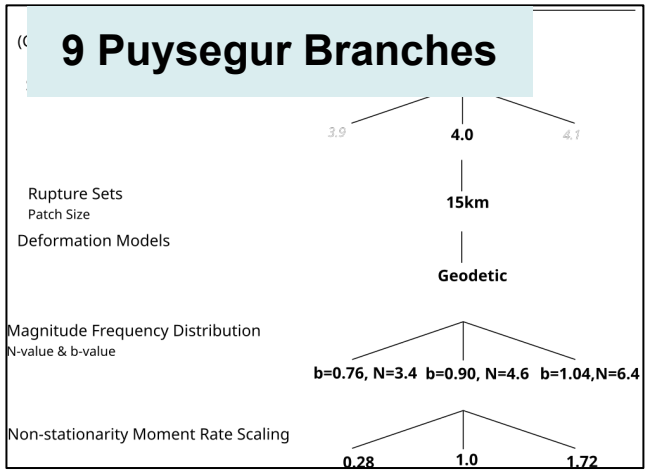
- Treat each branch of the SRM tree separately
  - 49 OpenQuake runs each producing 12 - 21 realisations
- Store independent realisations in cloud database (referenced by unique ID)

Stage 2:

- Calculate complete ~1M branch realizations as combinations of component realisations
- Calculate weighted mean, std, and quantiles from the full set of realisations



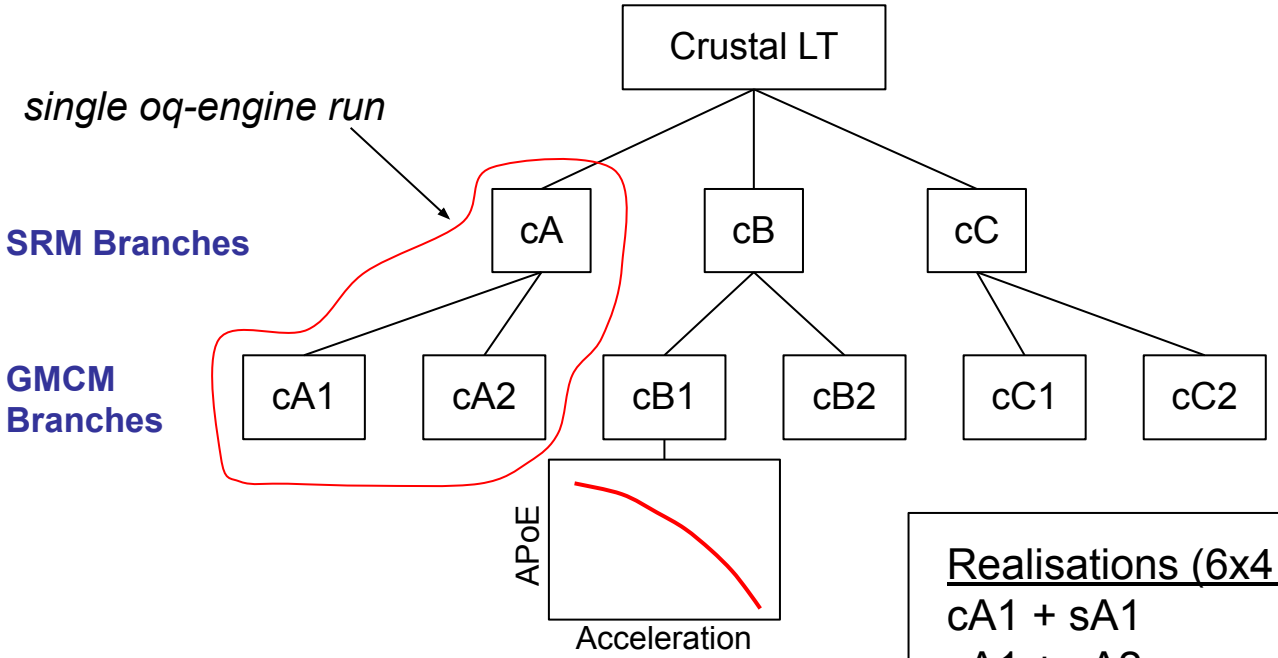


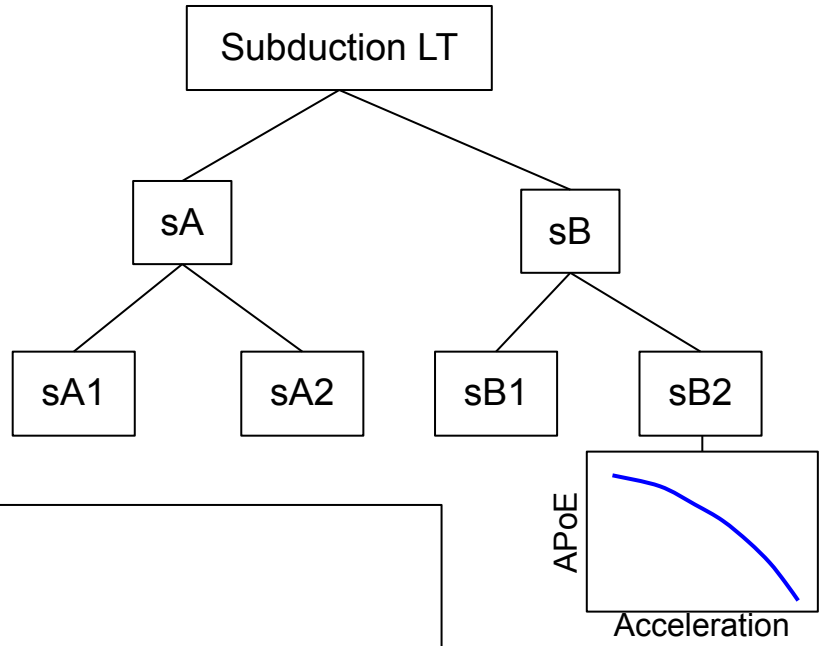


1 Intra-Slab Branch

<u>324</u> SRM Logic Tree Branches

Toy Example - 6 SRM branches, 4 GMCM branches





- 5 oq-engine runs
- 2 realisations per run
- 24 total combinations

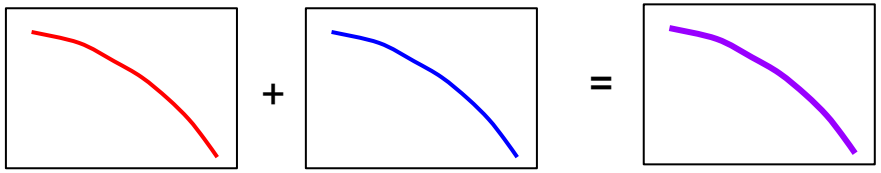
Realisations (6x4 = 24 total)

cA1 + sA1

cA1 + sA2

…

cB1 + sB2 =



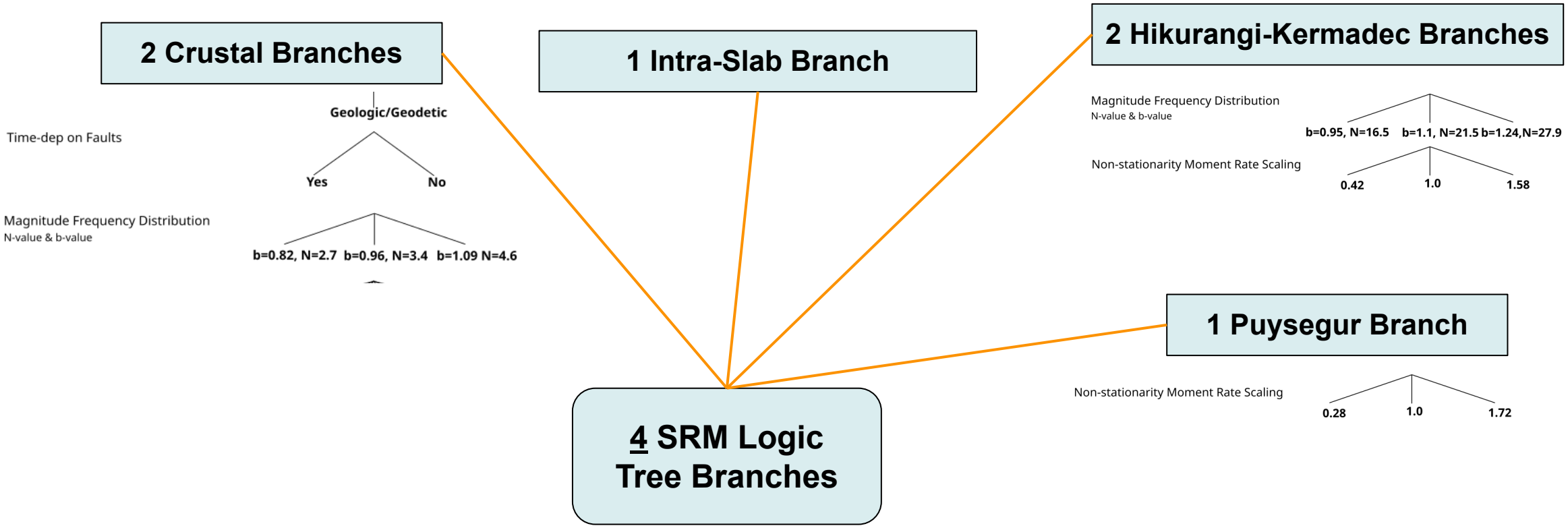
sum hazard curves in rate space

…

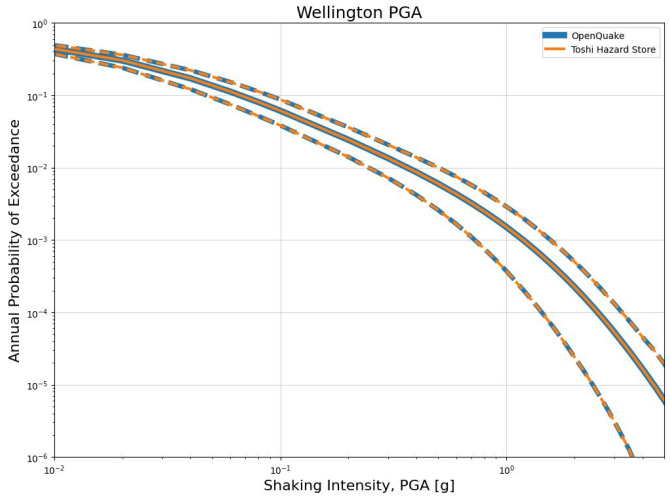
…

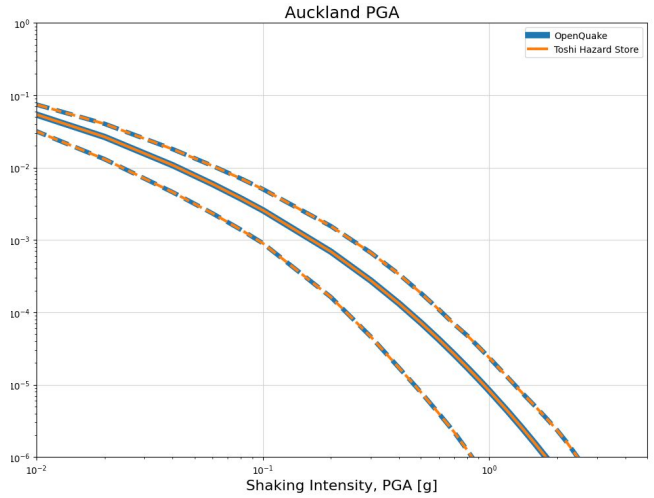
cC2 + sB2

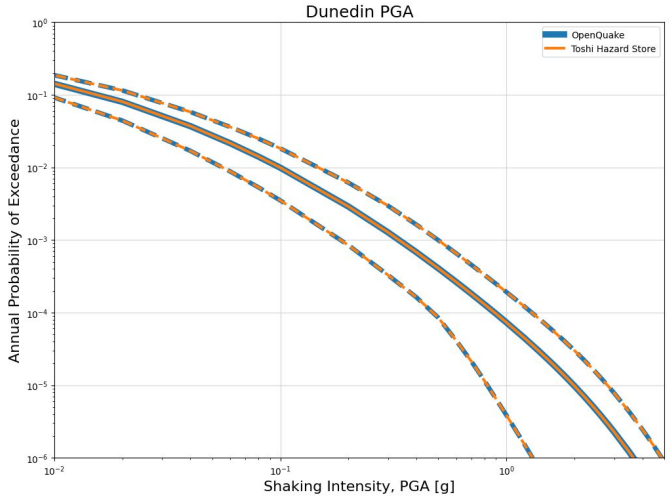
Verification

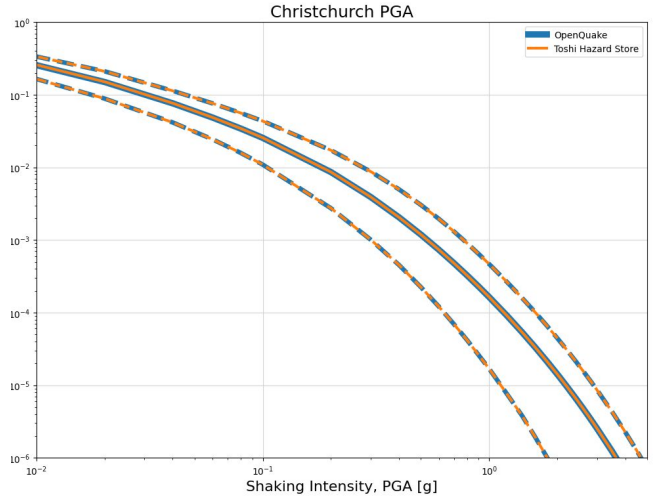


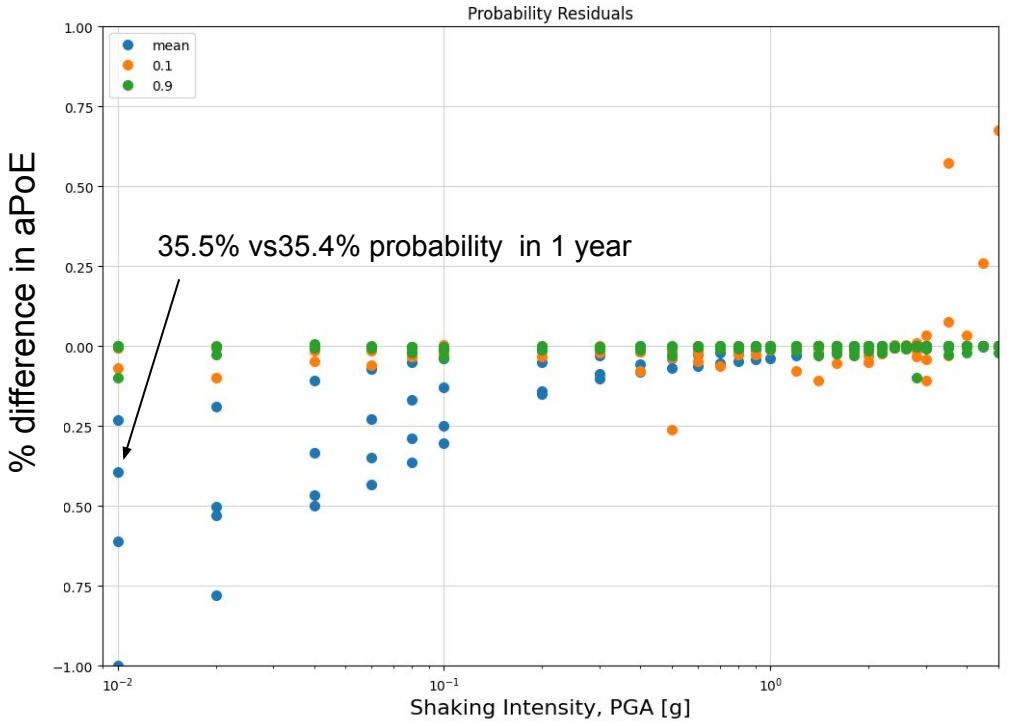
Verification





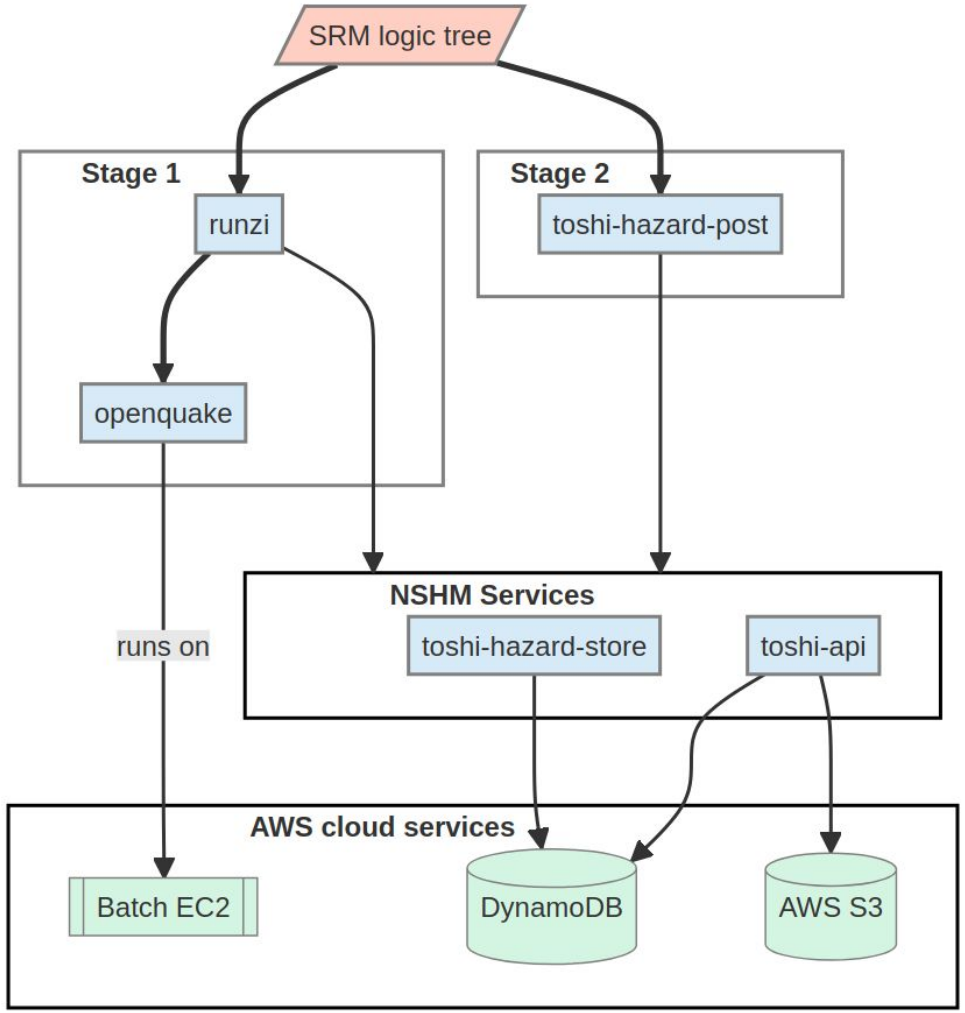






IT Stack and Deployment of Hazard Calc

- Allowed us to **parallel process hazard curves on modest compute nodes ... sometimes up to 750 concurrent jobs**, each between 1 to 24 hours duration.
- Recombination and calculation of aggregate statistics on workstation or parallel in cloud
  - **Embarrassingly parallel**
- Both stages of the process **easily parallelize and scale** across multiple dimensions of the model
- Breaking the calculation into stages and storing intermediate products provide benefits for experimentation
  - New logic tree configurations, easy sensitivity testing, sub-sampling to find minimum-viable logic trees, minimise calculation overheads, etc.



Parquet: a Portable Database

Initial model bilt using AWS DynamoDB cloud database

- Difficult to share
- Difficult for users to create their own database: requires cloud expertise and sophisticated IT stack

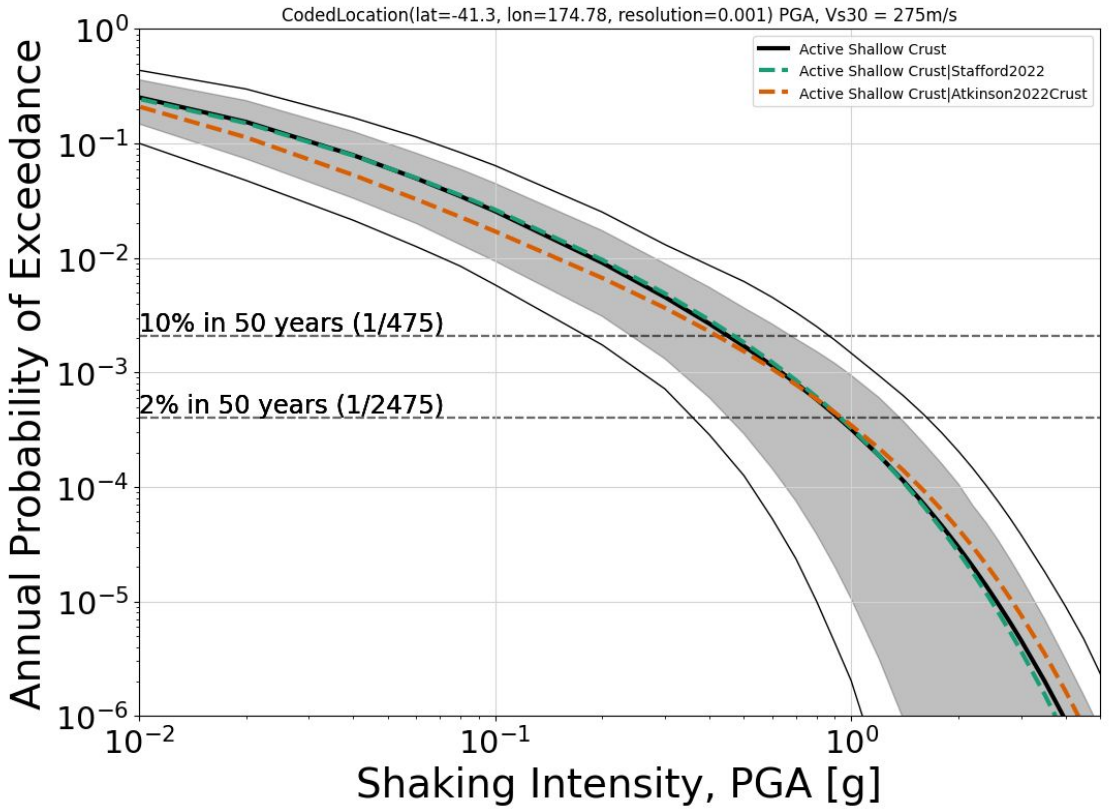
Moving to Parquet file format for component realizations

- Cloud (S3) or local storage
- Portable / sharable
- Easy to use
- Powered by PyArrow python library

Aggregate curves for dissemination to users still stored using DynamoDB

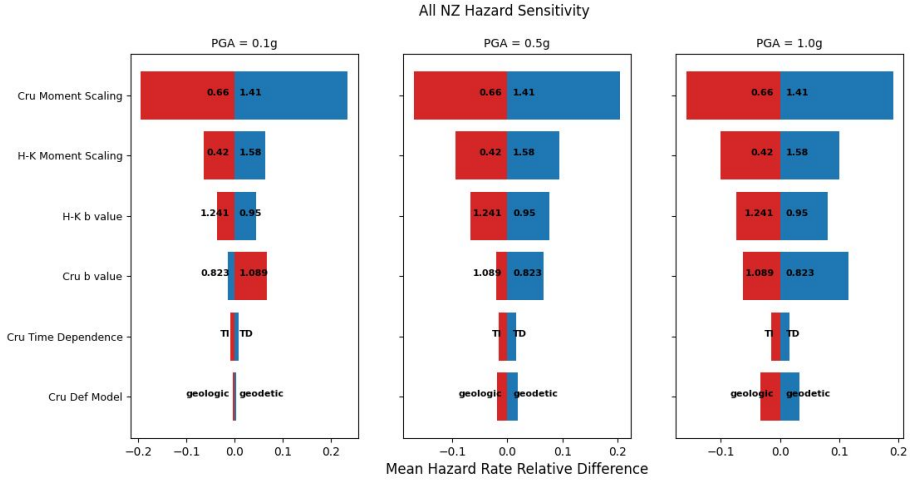
Component Re-Use Example: Sensitivity Testing

Every logic tree branch (both SRM And GMCM) has a unique identifier in the DB making it easy to select the components of interest.




```
args.srm_logic_tree = srm
args.gmcm_logic_tree = gmcm
args.hazard_model_id = trt
run_aggregation(args)

gsim_names = ['Stafford2022', 'Atkinson2022Crust']
for gsim_name in gsim_names:
    print(gsim_name)
    gmcm_1gsim = copy.deepcopy(gmcm)
    gmcm_1gsim.branch_sets[0].branches =
select_gmm(gmcm.branch_sets[0], gsim_name)
    args.gmcm_logic_tree = gmcm_1gsim
    args.hazard_model_id = f'{trt}|{gsim_name}'
    run_aggregation(args)
```



Summary

- **2 stage hazard calculation process allows for calculation of very large logic trees, circumventing software and hardware limitations**
- **Uses:**
  - Easy parallelization using modest computer hardware (e.g. cloud compute)
  - Full logic tree enumeration
  - Re-use of common components of model (e.g. sensitivity testing, logic tree re-weighting)
  - Sharing of common components of model
- **Move to new database technology makes it easier for others to:**
  - Use our model components
  - Build their own large hazard model

NB: Recent performance improvements in OpenQuake Engine (v3.20+) able to handle NZ NSHM 2022

DiCaprio, et. al., Calculation of National Seismic Hazard Models with Large Logic Trees: Application to the NZ NSHM 2022. Seismological Research Letters 2023 https://doi.org/10.1785/0220230226