Testing & Validation of Databases of Numerical Simulations using the EFEHR eGSIM Webservice

Graeme Weatherill

GFZ German Research Centre for Geosciences, Potsdam, Germany
graeme.weatherill@gfz-potsdam.de

DT-GEO

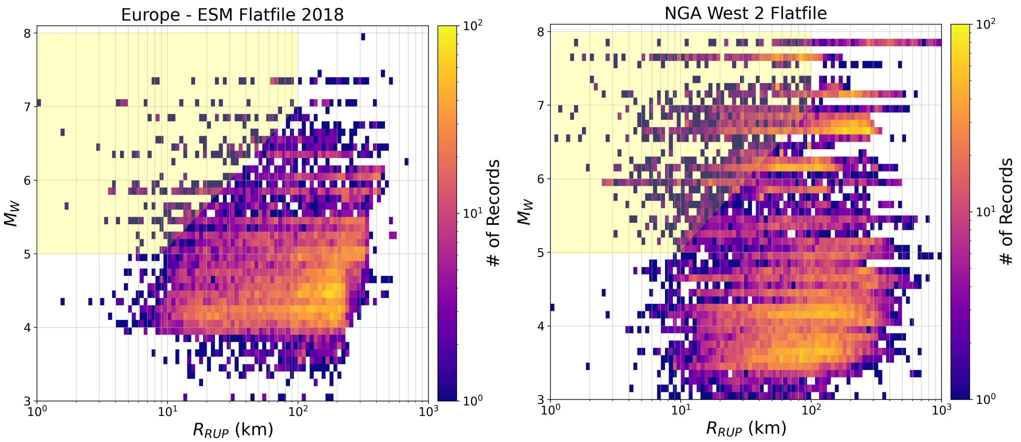
EFEHR

This project has received funding from the European Union's Horizon research and innovation programme under the grant agreement No 101058129

Context & Motivation

Latest generation European seismic hazard model (ESHM20) adopted GMMs calibrated on a large database of ground motion observed in Europe (ESM, Lanzano et al., 2018; Luzi et al., 2020)

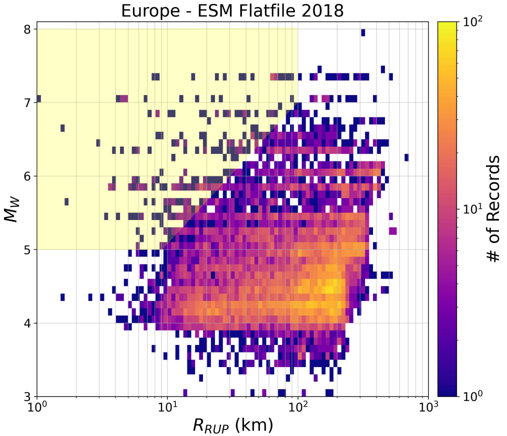
- ESM: 10,000's of records but still limited for large M, small R range
- Near-source scaling was a problem! (e.g. Kotha et al., 2022)
- New generations of observational databases will increase data, but with limited numbers of records from large events

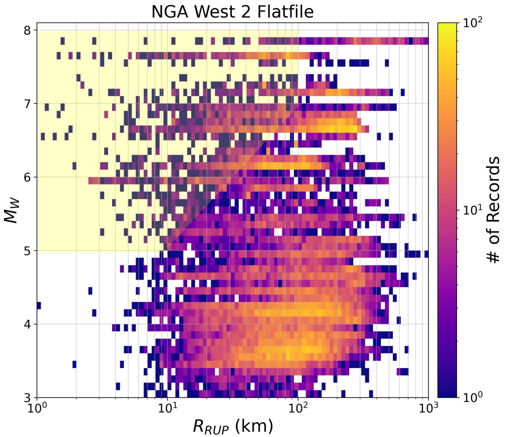


Context & Motivation

Latest generation European seismic hazard model (ESHM20) adopted GMMs calibrated on a large database of ground motion observed in Europe (ESM, Lanzano et al., 2018; Luzi et al., 2020)

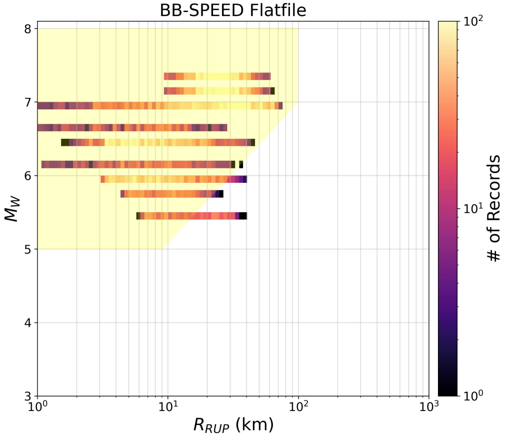
- ESM: 10,000's of records but still limited for large M, small R range
- Near-source scaling was a problem! (e.g. Kotha et al., 2022)
- New generations of observational databases will increase data, but with limited numbers of records from large events





Next Generation European GMMs:

- Aim for fully non-ergodic GMMs (if possible)
- Improved calibration of near-source scaling integrating data from simulations
- Dynamic: Regularly updatable

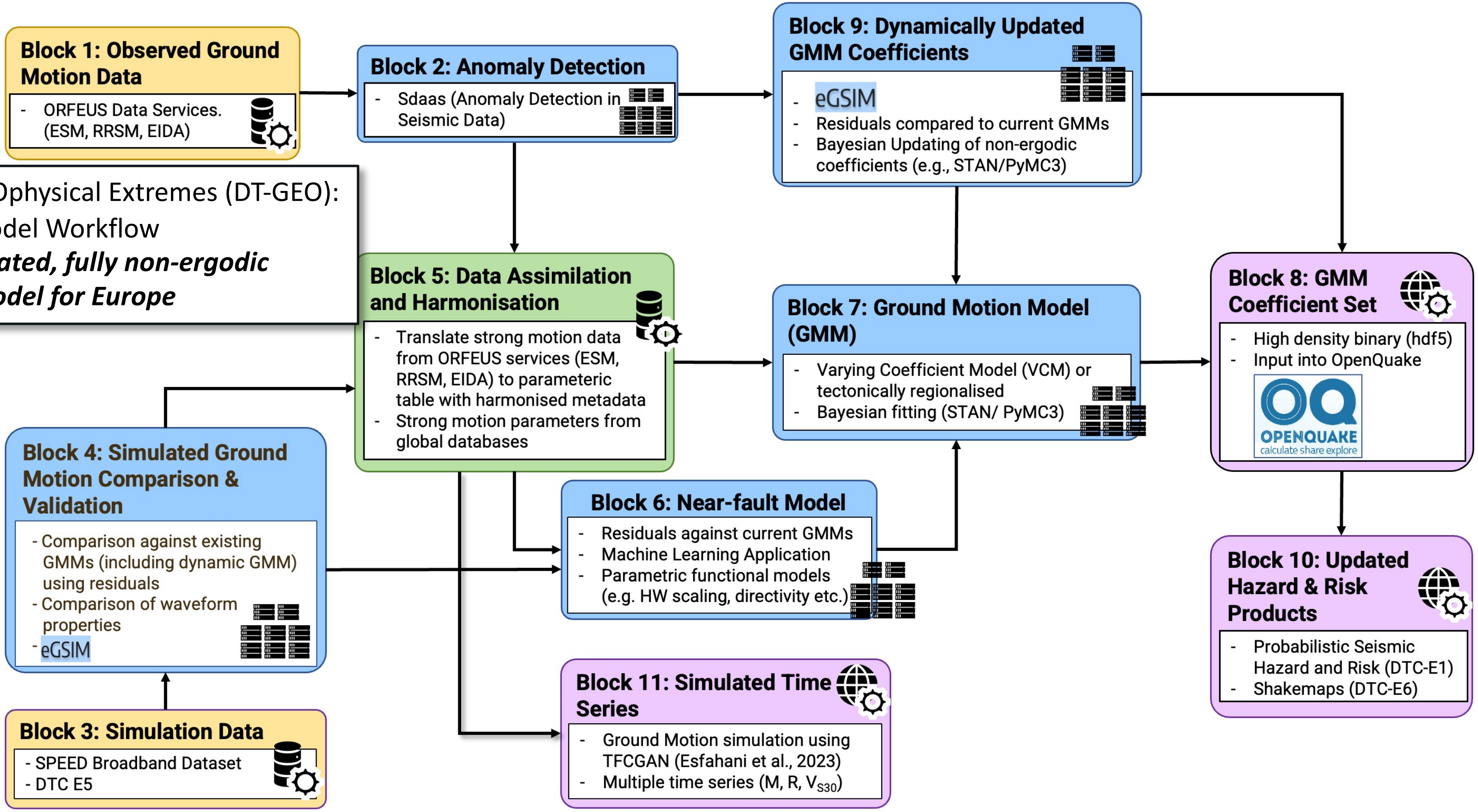


Need to understand if simulations are *usable* for this purpose

Context & Motivation

Digital Twin for GEOphysical Extremes (DT-GEO): Ground Motion Model Workflow

A dynamically updated, fully non-ergodic Ground Motion Model for Europe



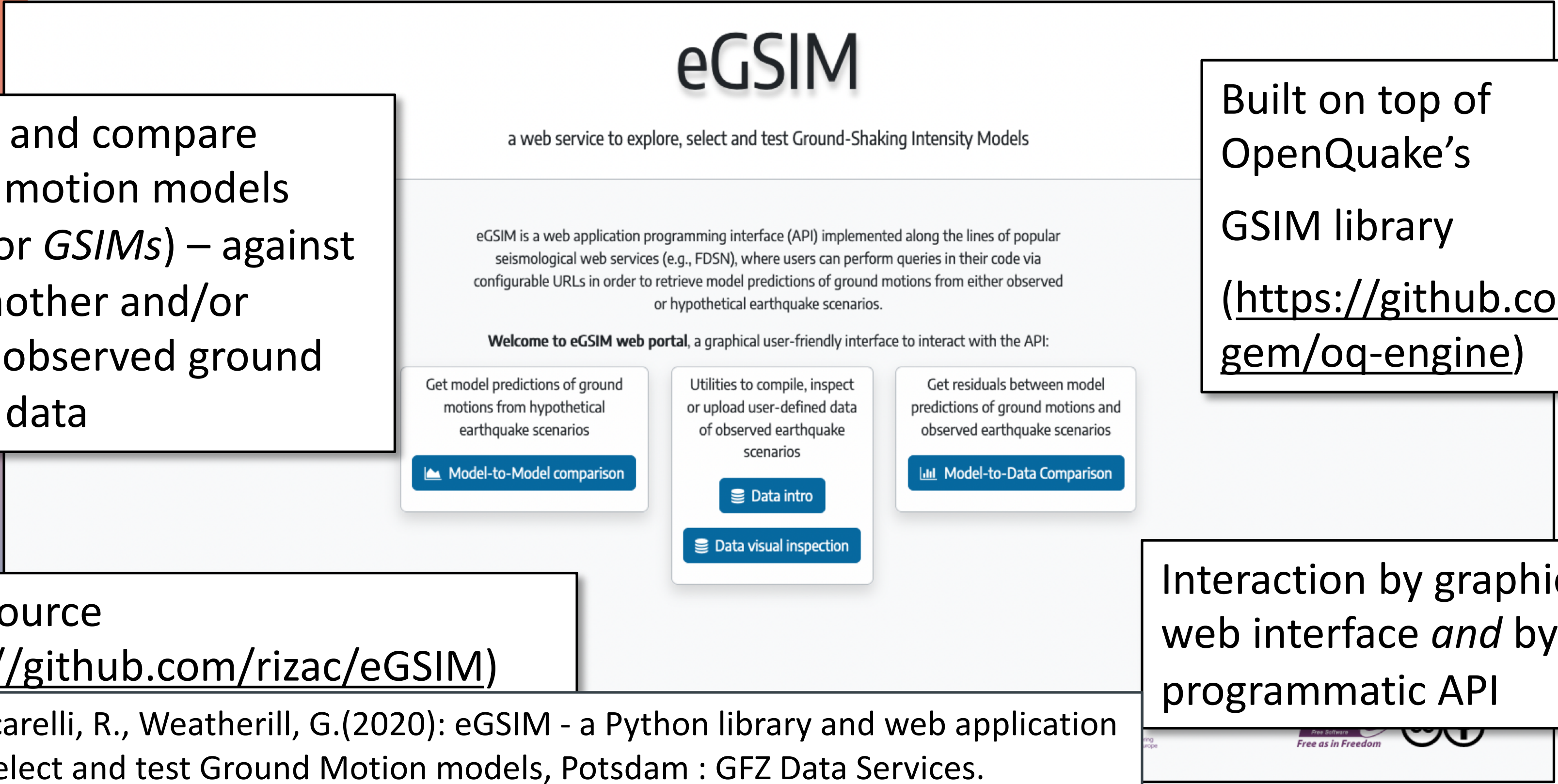
Context & Motivation

Rezaeian *et al.* (2024) Recommendations

The following steps can be taken for validating a simulation platform for use in GMM development or PSHA (i.e. for validation metrics related to Sa):

1. Simulate a wide range of events that span from the conditions for which ample observations are available (generally small to moderate magnitudes) to relatively rare, but hazard-controlling, situations.
2. Compute residuals relative to suitable GMMs and partition the residuals (mixed effects)
3. Evaluate trends of between-event residuals with respect to source parameters, which should include magnitude and potentially additional relevant source parameters (e.g. rupture depth).
4. Evaluate trends of within-event residuals with respect to path parameters (e.g. rupture distance).
5. As needed, address path misfits in the simulations to facilitate the next step.
6. Evaluate trends of within-event residuals (or site-terms derived from those residuals) against site parameters.
7. Compare standard deviation terms to empirical models.

eGSIM Webservice



Explore and compare ground motion models (GMM or *GSIMs*) – against each another and/or against observed ground motion data

Built on top of OpenQuake's

GSIM library

(https://github.com/gem/oq-engine)

Open source (https://github.com/rizac/eGSIM)

Interaction by graphical web interface *and* by programmatic API

Zaccarelli, R., Weatherill, G.(2020): eGSIM - a Python library and web application to select and test Ground Motion models, Potsdam : GFZ Data Services. https://doi.org/10.5880/GFZ.2.6.2023.007

https://egsim.gfz-potsdam.de/home

eGSIM Webservice

gmid	event_id	event_time	event_longitude	event_latitude	event_depth	magnitude	strike	dip	ra
INT-20230206_0000008|KO-TUNC|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0118|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0119|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0122|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0123|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
...	...	...	...	...	...	...	...	...	
INT-20230324_0000175|KO-TUNC|HN	INT-20230324_0000175	2023-03-24T13:10:34.000000Z	37.55	38.01	7.0	4.40	347.000000	87.00000	17
INT-20230521_0000127|KO-ARPRA|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8
INT-20230521_0000127|KO-CMRD|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8
INT-20230521_0000127|KO-KHMN|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8
INT-20230521_0000127|KO-KOZK|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8

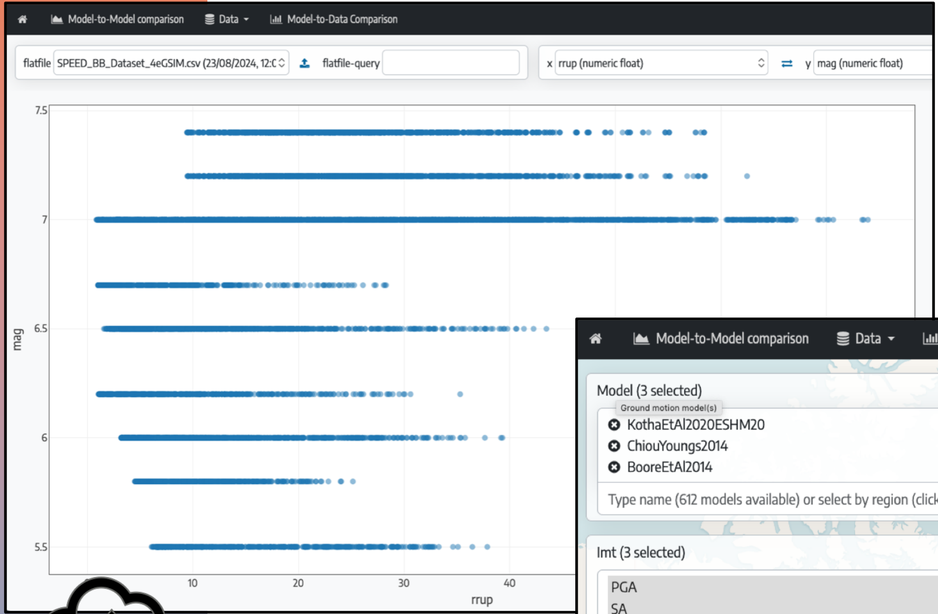
Flatfile includes metadata (source, path, site properties) and ground motion intensity data (PGA, PGV, Sa (T), etc.)

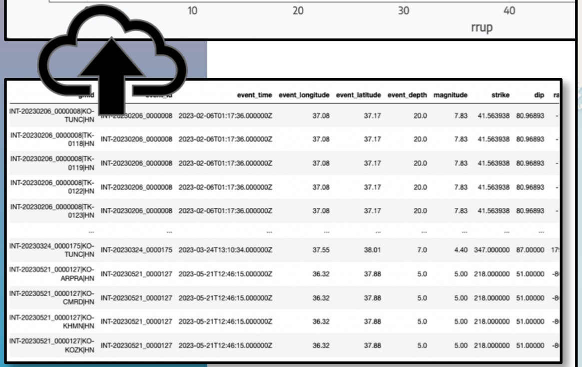
Different flatfiles needed for different horizontal component definitions (e.g. GM_{xy}, RotD50 etc.) or spectra definitions (e.g. EAS, Sa)

Ground motion databases can be organised into flatfiles *according to the eGSIM template* – which users can explore using eGSIM

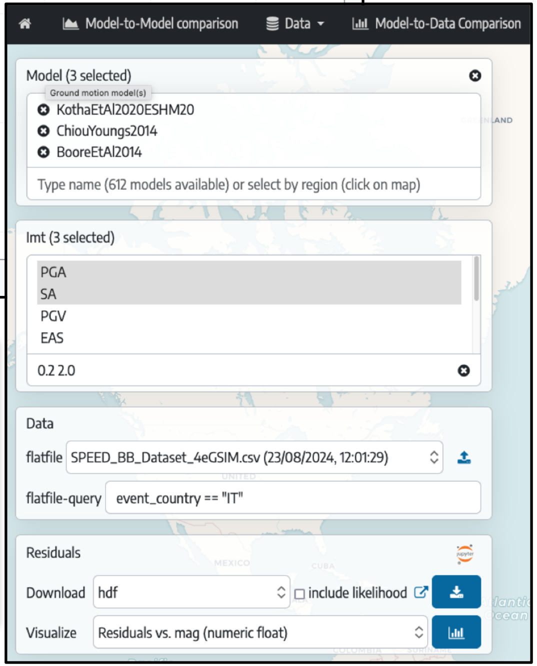
- Users can build their own databases and explore their fit to existing ground motion models
- Integration into rapid analysis (retrieve residuals for recent event, ongoing sequences etc.)
- Direct integration into automatic workflows

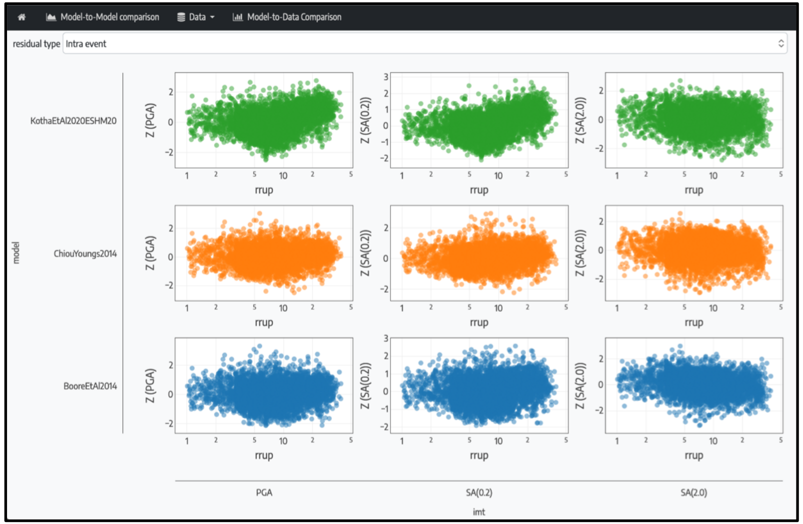
eGSIM Web Interface





[illegible]	[illegible]	event_time	event_longitude	event_latitude	event_depth	magnitude	strike	dip	ra
INT-20230206_0000008|KO-TUNC|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0118|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0119|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0122|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
INT-20230206_0000008|TK-0123|HN	INT-20230206_0000008	2023-02-06T01:17:36.000000Z	37.08	37.17	20.0	7.83	41.563938	80.96893	-
...	...	...	...	...	...	...	...	...	
INT-20230324_0000175|KO-TUNC|HN	INT-20230324_0000175	2023-03-24T13:10:34.000000Z	37.55	38.01	7.0	4.40	347.000000	87.00000	17
INT-20230521_0000127|KO-ARPRA|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8
INT-20230521_0000127|KO-CMRD|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8
INT-20230521_0000127|KO-KHMN|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8
INT-20230521_0000127|KO-KOZK|HN	INT-20230521_0000127	2023-05-21T12:46:15.000000Z	36.32	37.88	5.0	5.00	218.000000	51.00000	-8





eGSIM Webservice (API Usage)

jupyter Comparing Observations and Simulations for a Single Earthquak... Last Checkpoint: 6 minutes ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

Run Markdown

Can we use eGSIM to compare observations against physics-based simulations?

In [1]:
```
%load_ext autoreload
%autoreload 2
import warnings; warnings.filterwarnings("ignore")
```

In [2]:
```
%matplotlib inline

# Python standard library imports
import io
import os
import json
import requests
from typing import List, Dict, Tuple, Optional

# Numeric/scientific Python etc.
import numpy as np
import pandas as pd

# Visualization tools
import matplotlib.pyplot as plt
from matplotlib.colors import Normalize, LogNorm
import seaborn as sns

```

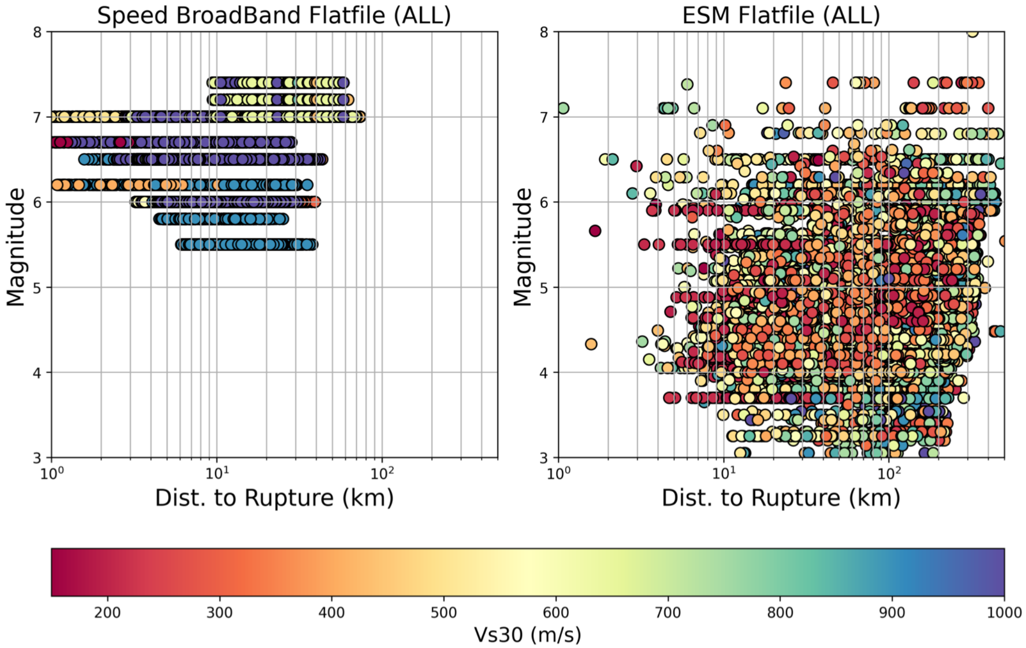
How to call eGSIM's API to get Ground Motion Model Residuals

In [ ]:
```
# Set the eGSIM URL to get the ground motion residuals
egsim_url_residuals = "https://egsim.gfz-potsdam.de/api/query/residuals"


def get_residuals_from_egsim(
    flatfile_path: str,
    gmms: List,
    imts: List,
    data_format: str = "hdf",
    query_string: str = ""
) -> Dict:
    """Retreive the residuals for the flatfile and the selected
    set of ground motion models and intensity measure types

    Args:
        flatfile_path: Local path to the selected flatfile
        gmms: List of ground motion models (OpenQuake class names)
        imts: List of intensity measure types (e.g. PGA, PGV, SA(0.1) etc.)
        plot_type: Column to return for x-values (e.g. mag, rrup, etc)
        query_string: Selection query to apply to the data
```

Comparing Simulations & Observations – Residual Analysis for the BB-SPEEDset



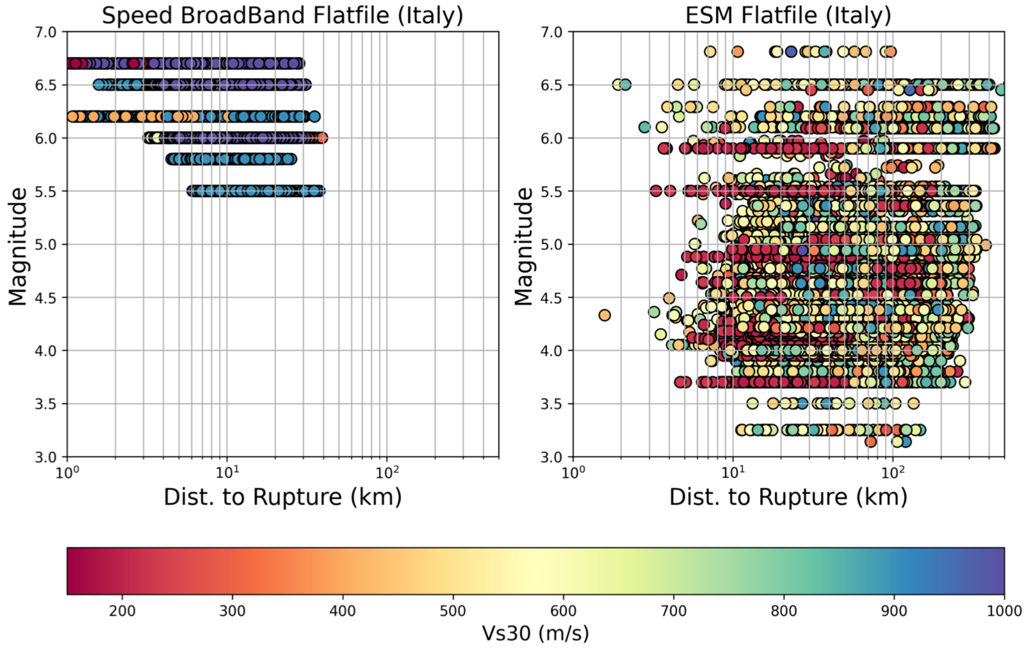
Broadband (BB) SPEED:

- Long period ground motion simulated by SPEED (Spectral Elements) code
- Broadband prediction from long period motion based on artificial neural network (ANN) (ANN2BB, Paolucci et al., 2018)

BB-SPEEDset: Harmonized flatfile of BB-SPEED simulations (Paolucci et al., 2021)

- 22 scenarios (≈ 500 – 650 records per scenario)
- 10 Italy (mix of real and hypothetical events)
- 2 from Greece, 9 Turkey (Mw 7.2 and Mw 7.4 Sea of Marmara scenarios), 1 Japan
- Most simulations on rock (some on reference Vs profile for stiff or soft soil)

Comparing Simulations & Observations – Residual Analysis for the BB-SPEEDset



Broadband (BB) SPEED:

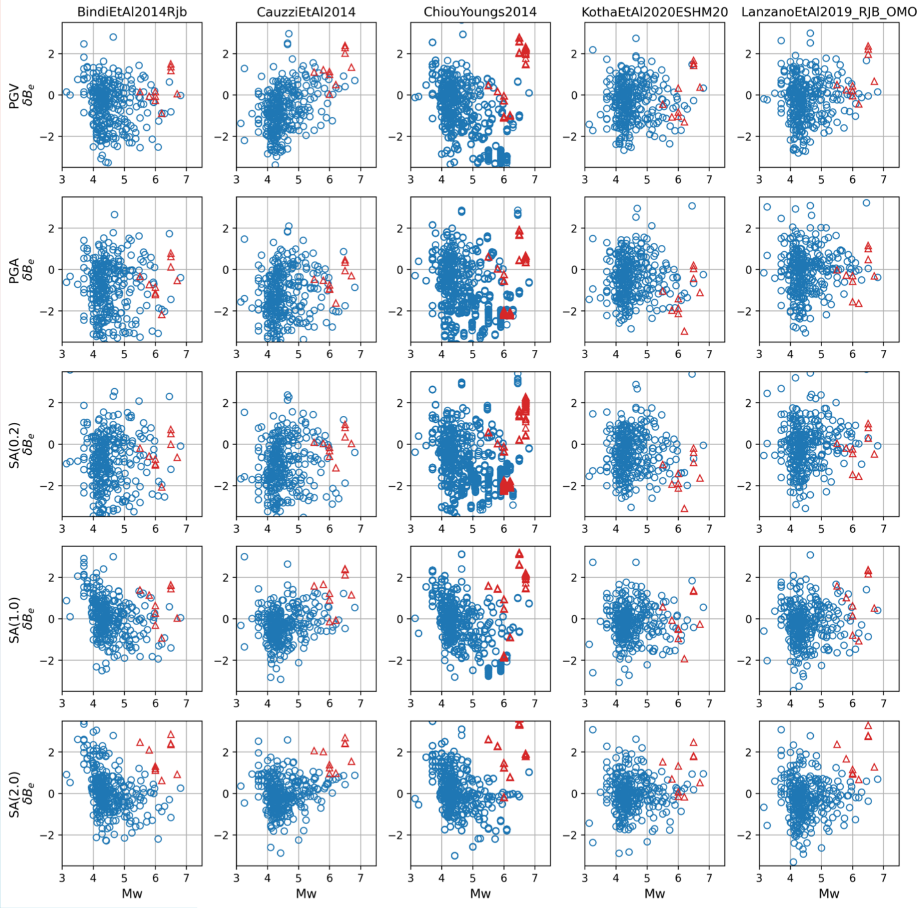
- Long period ground motion simulated by SPEED (Spectral Elements) code
- Broadband prediction from long period motion based on artificial neural network (ANN) (ANN2BB, Paolucci et al., 2018)

BB-SPEEDset: Harmonized flatfile of BB-SPEED simulations (Paolucci et al., 2021)

- 22 scenarios (≈ 500 – 650 records per scenario)
- 10 Italy (mix of real and hypothetical events)
- 2 from Greece, 9 Turkey (Mw 7.2 and Mw 7.4 Sea of Marmara scenarios), 1 Japan
- Most simulations on rock (some on reference Vs profile for stiff or soft soil)

Example from Italy – compare residual trends from simulations with observed data from ESM (limited only to Italy records)

Comparing Simulations & Observations – Residual Analysis for the BB-SPEEDset



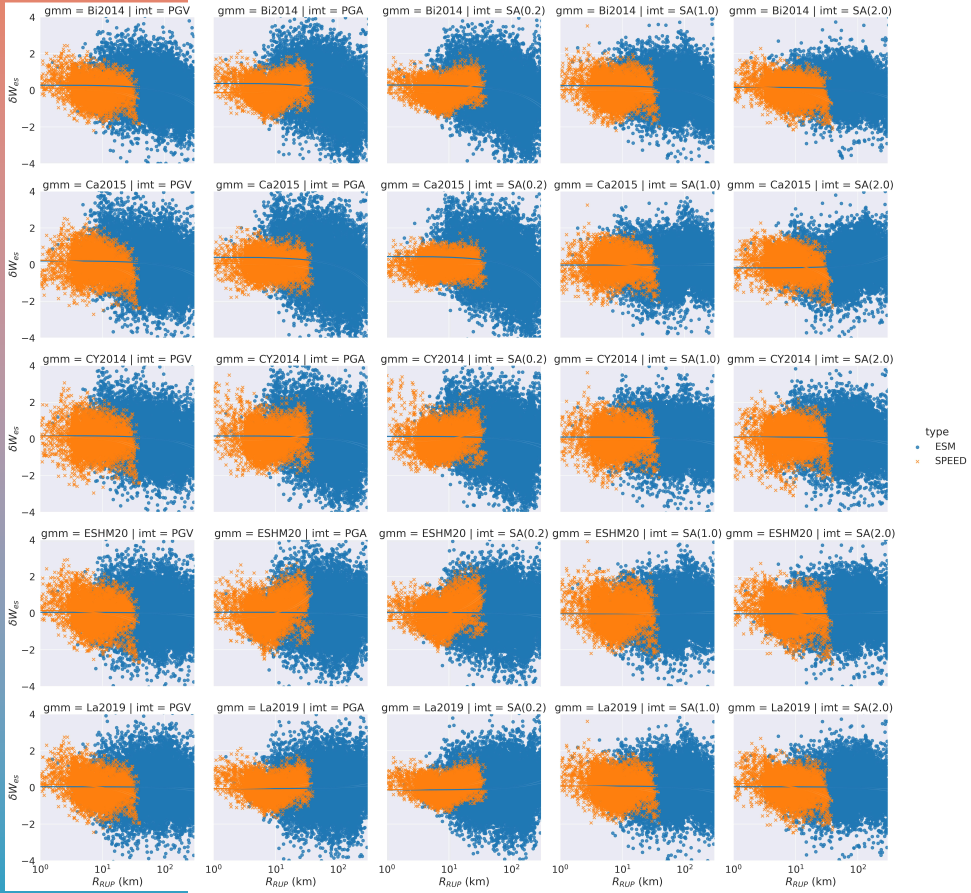
Reasonable agreement between data, simulations and Italian GMM (Lanzano et al., 2019)

BB-SPEED produces higher between-event residuals (δB_e) at long periods and PGV

Some important caveats – will bias high δB_e for BB-SPEEDset

- Simulations limited to short distances
- Simulation dataset <u>not</u> directivity centred (stations more likely to reflect forward directivity condition)
- Too few events to compare variability

Comparing Simulations & Observations – Residual Analysis for the BB-SPEEDset



Reasonable agreement between data, simulations and Italian GMM (Lanzano et al., 2019)

BB-SPEED produces higher between-event residuals (δB_e) at long periods and PGV

Some important caveats – will bias high δB_e for BB-SPEEDset

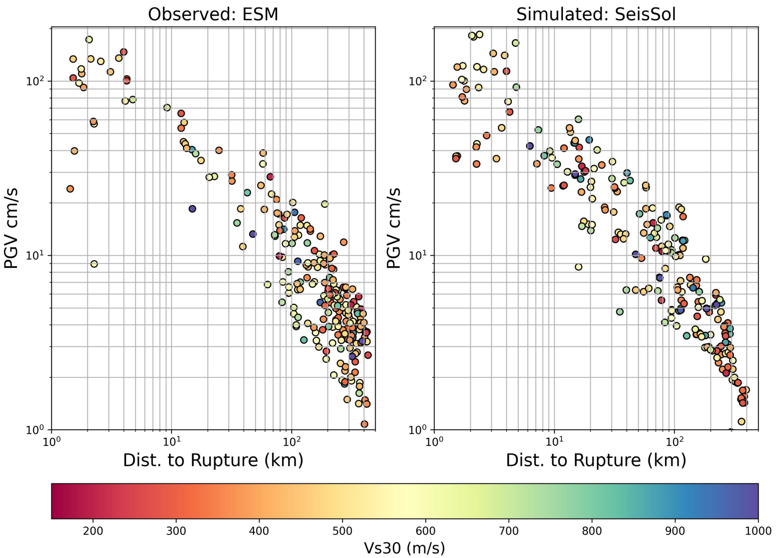
- Simulations limited to short distances
- Simulation dataset not directivity centred (stations more likely to reflect forward directivity condition)
- Too few events to compare variability

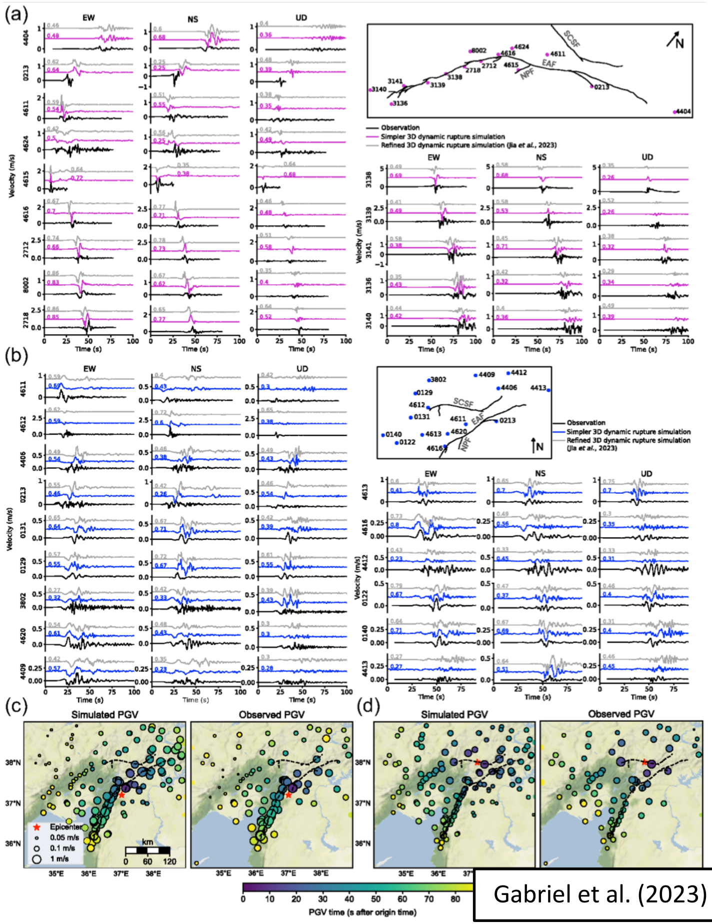
Attenuation trends captured well but with bias toward positive δW_{es} at short distances

Artificial Neural Network to predict short period motion is trained on global database of records

- Inherent global scaling with respect to M and R
- Irreducibly ergodic

Comparing Simulations & Observations – Residual Analysis for the 2023 Kahramanmaras EQ



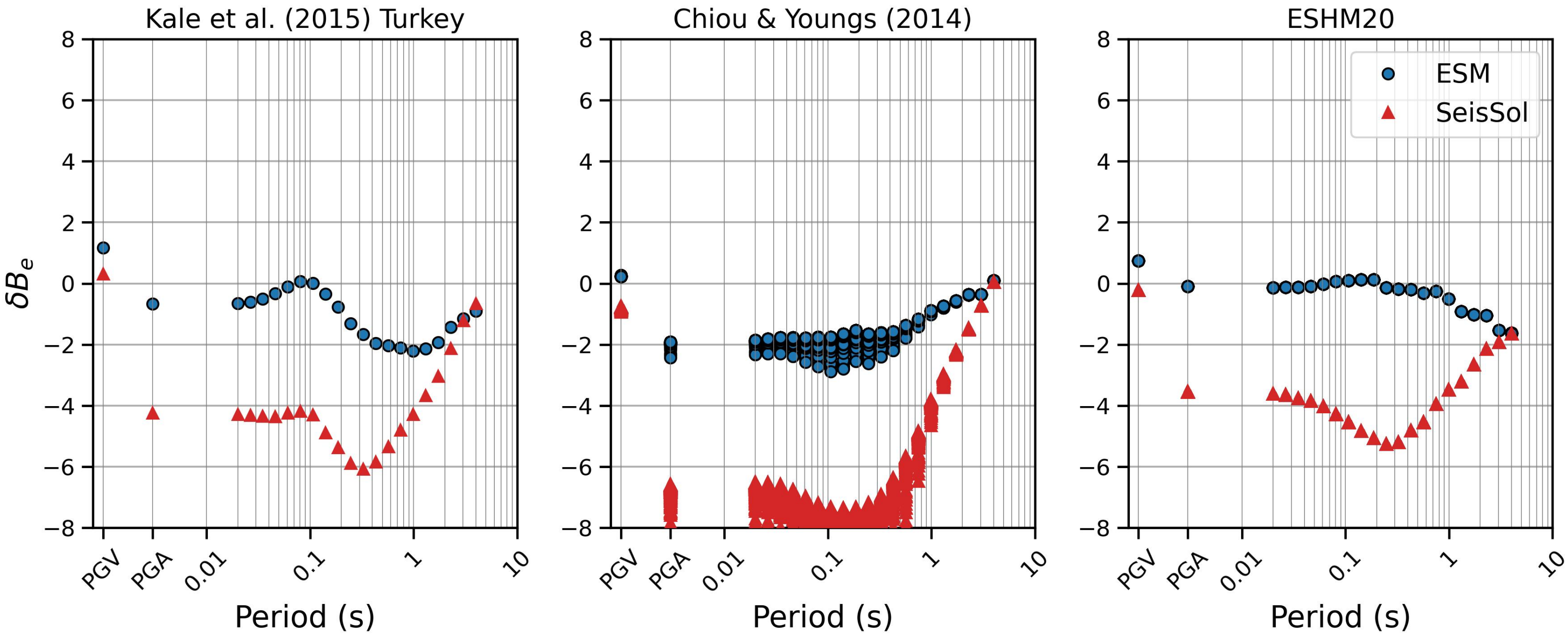


Gabriel et al. (2023)

Test case application comparing observations to simulations via eGSIM

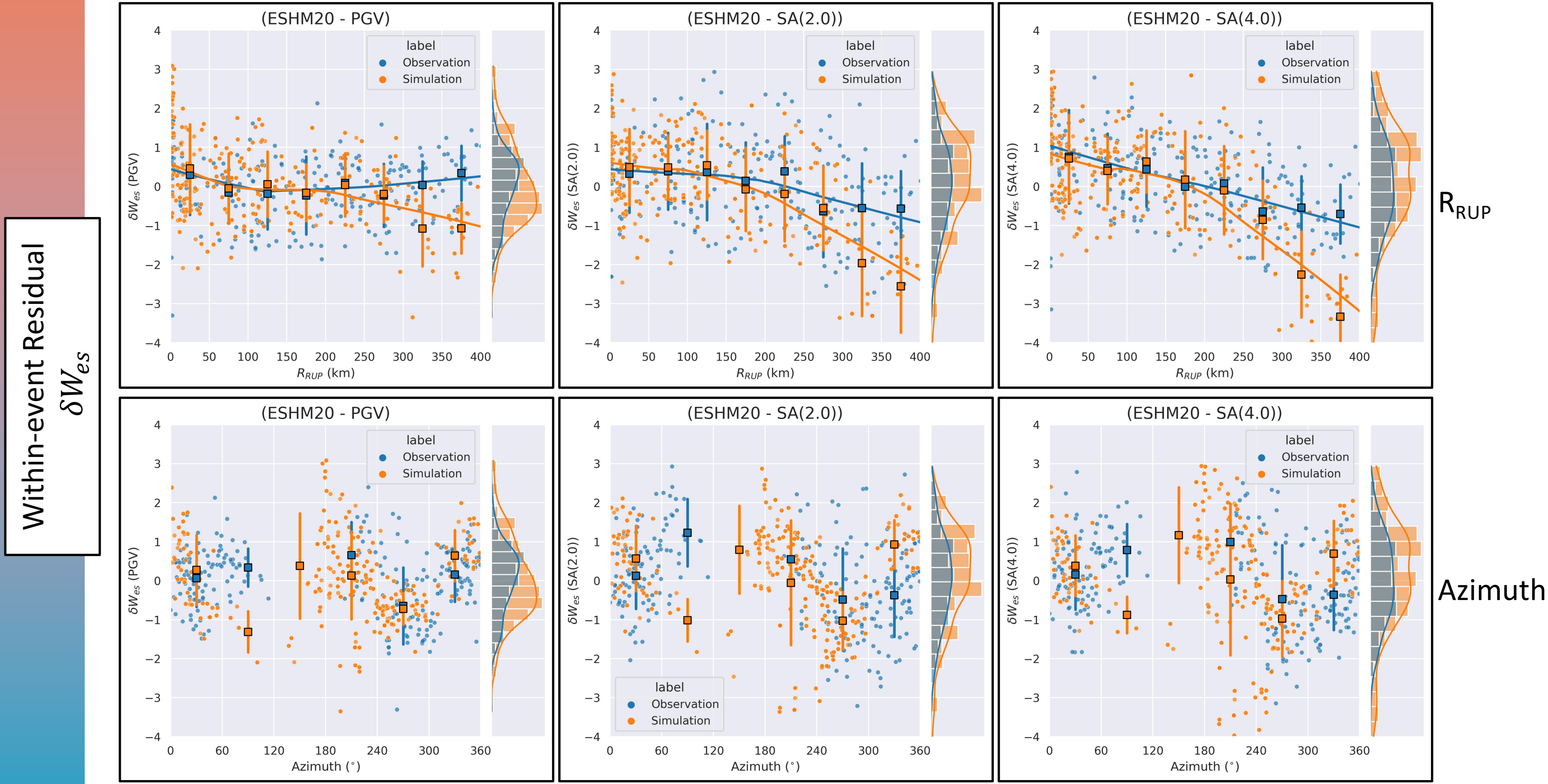
- Data from Kahramanmaras earthquake (1st event Mw 7.83) from ESM
- Simulations from Gabriel et al. (2023) using SeisSol
- Using only stations common to both observations and simulations *with measured* V_{S30}

Comparing Simulations & Observations – Residual Analysis for the 2023 Kahramanmaras EQ



- Maximum resolved frequency ≈ 1.5 Hz (0.67 s)
- Usable *response spectral periods* ≥ 2 - 3 s
- Without high frequency content – waveforms cannot be used directly for engineering seismology applications
- Can be used to explore and potentially calibrate long period phenomena (directivity pulse) and attenuation

Comparing Simulations & Observations – Residual Analysis for the 2023 Kahramanmaras EQ



A Vision for an Open[-Source] Ecosystem for Testing and Validating Numerical Simulations

Geo-INQUIRE DT-GEO

Physics-Based Simulations
- 3-component time series
- Software agnostic
- Standard data format?

Simulation Configuration Data
- Rupture parameters
- Fault surface & slip model
- Sensor locations & site properties (if relevant)
- General Attributes (ID, software, contact etc.)

Simulation Data Lake (or Data Warehouse/Lakehouse)
- Search by structured query of metadata and access via API
- Waveform download in ASDF, mSEED etc.
- Source/Site format TBD
- FAS, SA, other IMs etc.
  - *Store/generate none?*
  - *Calculate on-the-fly?*
  - *Persist?*
- IMs as metadata?

Observed Earthquake Time-Series
- IRIS, EIDA, other …
- Obspy

Get Intensity Measures
- *SCEC ts_process*
- *USGS groundmotion-processing*
- *Stream2segment*
- *Custom?*

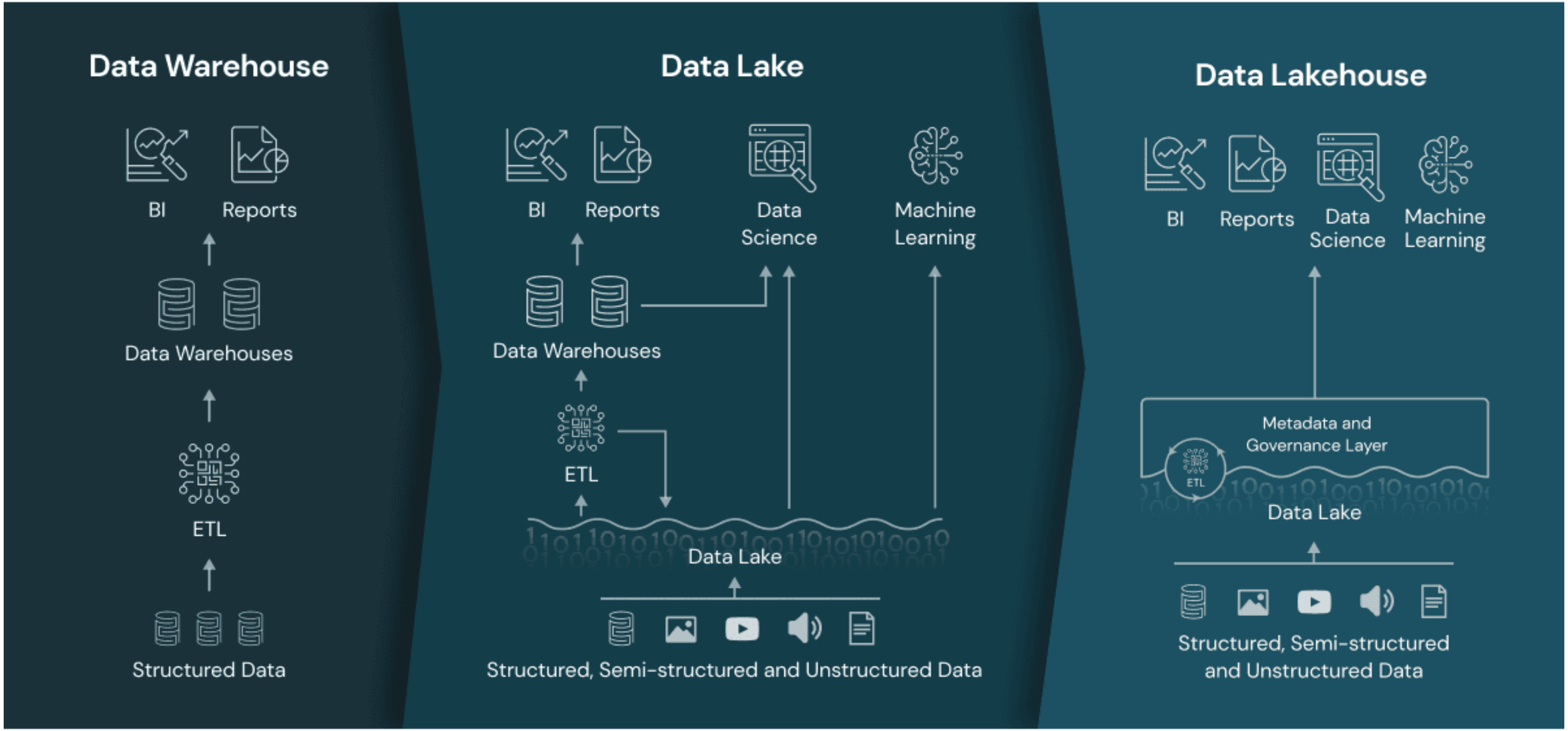
Build Flatfile
- OpenQuake (metadata)
- Other?

Compare Time-Series against Observations
- SCEC GMSVToolkit
- Custom?

Compare Databases against GMMs
- eGSIM

- All components using open-source software and services
- Common data standards would make it possible to build this into a single environment
- Should be containerised/hosted on HPC systems (perhaps a post DT-Geo objective)

Data Lakes, Data Warehouses or Data Lakehouses?



BI: Business Intelligence
ETL: Extract, Transform, Load

Ground motion simulation data *can be* structured – for many use cases this is not only desirable but critical

Metadata: Source

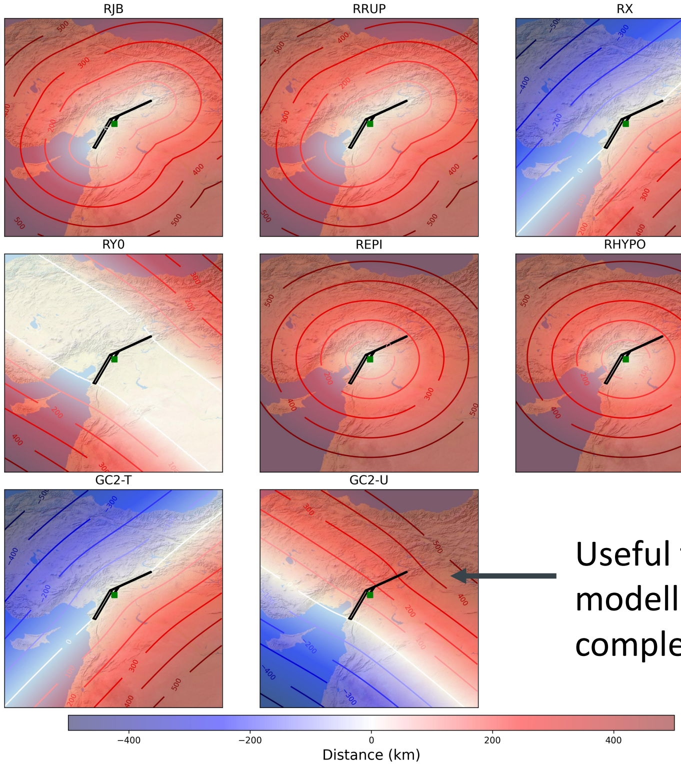
Most modern GMMs characterise the source as a finite fault and require attributes that reflect this:

Slip: Rake

Geometry: Dip, Width, Top of Rupture Depth, Bottom of Rupture Depth, Hypocentre position within fault

Source-to-Site Distance: R_{JB}, R_{RUP}, R_x, R_y (R_{y0}), Generalized Coordinate System 2 (GC2) T and U

My workflow: *Translate finite faults into OpenQuake surface objects (Python) to calculate distances and retrieve geometry information*



Useful for directivity modelling around complex ruptures

Metadata: Source

Most modern GMMs characterise the source as a finite fault and require attributes that reflect this:
Slip: Rake
Geometry: Dip, Width, Top of Rupture Depth, Bottom of Rupture Depth, Hypocentre position within fault
Source-to-Site Distance: R_{JB}, R_{RUP}, R_x, R_y (R_{y0}), Generalized Coordinate System 2 (GC2) T and U

eGSIM can tell you what parameters you need depending on the GMMs you want to compare against:
https://egsim.gfz-potsdam.de/flatfile-metadata-info

Nearly all physics-based simulation software characterises finite fault surfaces but:

1) These are rarely disseminated (in digital format) with the simulation results

2) Not an obvious prevailing standard file format (SRCMOD & USGS provide SLP, FSP, geojson; SCEC BBP genslip produces .GSF; other software produce custom formats)

3) ASDF *can* store information from QuakeML binary dump – but QuakeML is limited here!

4) Is the way we define the finite fault source extent from the slip distribution consistent in simulations and observations? What about in complex multi-segment rupture cases?

5) Could other information from the slip model be of interest for GMMs?

Perhaps we could 1) adopt a standard format? 2) Benchmark the source and distance calculations?

Metadata: Receiver, Velocity Model, Mesh etc.

Unique identifier and location of the receivers (including Coordinate Reference System) is essential!

For comparisons with GMMs (scalar quantities, easy to query, can be attributes in HDF5, ASDF etc.):

- (Nearly) All: V_{S30}
- NGA, state-of-the-art etc.: $Z_{1.0}$, $Z_{2.5}$, H_{800}, f_0, Z_{ref}

Should include descriptive information about source of the site data

Some broadband studies use different V_{ref} and velocity models for low and high frequency parts of the simulations (e.g. Lee et al., 2022; CyberShake 22.12 etc.) – not clear how to define GMM parameters consistently

“Large volume” metadata (e.g. full 3D velocity model, mesh, etc.):

- High cost to storage and access
- Is there a use-case that would need this information for GMMs?
- Is it sufficient to provide descriptive information to reproduce the velocity profile, mesh etc., rather than providing the data themselves?
- What about proprietary data/formats/software?

Simulations in European GMMs: Where Do We Go from Here?

Many research groups in Europe working on physics-based simulations:

- Large community but very diverse objectives and applications
- Different software and methods in use!
- Simulation results not easy to access, near impossible to benchmark

Not easy for ground motion modellers to integrate insights from simulations, understand uncertainties, limitations etc. *(unless they develop the expertise in-house!)*

Geo-INQUIRE

DT-GEO

Ongoing Horizon Europe projects building infrastructures to make simulation calculations and results accessible

Important opportunity but outcomes need to be used by the engineering seismology community! Need to establish if these are sufficient to address the scientific questions

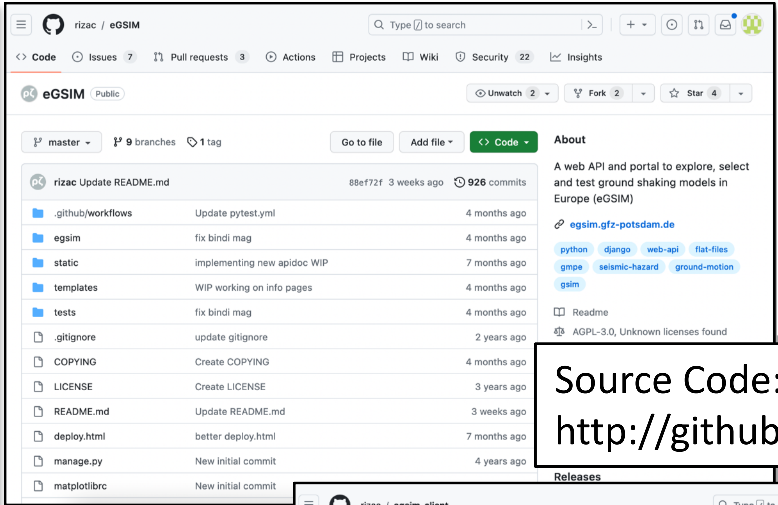
eGSIM (+ other V&V tools) positioned on the interface between simulations and GMMs – can become more useful if the simulation data lakes: 1) have suitable architecture for the purpose, 2) are filled with data

Next generation of GMMs in Europe (to end of 2020s) likely to remain largely empirically-based.

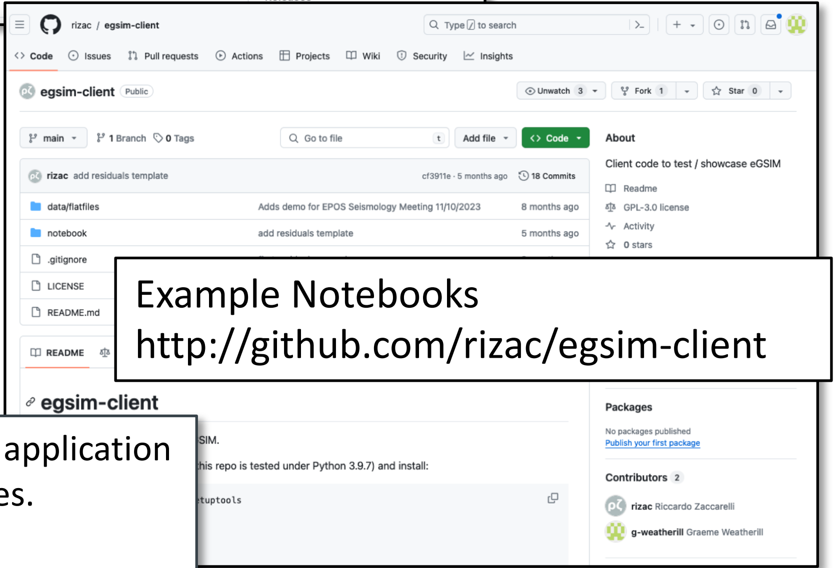
EFEHR

Could propose establishing working group on physics-based simulations and their integration in GMMs and PSHA?

http://egsim.gfz-potsdam.de



Source Code:
http://github.com/rizac/eGSIM



Example Notebooks
http://github.com/rizac/egsim-client

graeme.weatherill@gfz-potsdam.de

riccardo.zaccarelli@gfz-potsdam.de

Zaccarelli, R., Weatherill, G.(2020): eGSIM - a Python library and web application to select and test Ground Motion models, Potsdam : GFZ Data Services. https://doi.org/10.5880/GFZ.2.6.2023.007