THE SEISMIC HAZARD MODELLER'S TOOLKIT: AN OPEN-SOURCE LIBRARY FOR THE CONSTRUCTION OF PROBABILISTIC SEISMIC HAZARD MODELS

Graeme WEATHERILL[1], Luis RODRÍGUEZ[2], Marco PAGANI[3]

ABSTRACT

The methodology of probabilistic seismic hazard analysis (PSHA) is well established as a means of characterising the level of ground motion to which a location may be subjected to within a given time period. State-of-the-art PSHA calculation software is now freely available, permitting a wider implementation of PSHA across the globe. Yet whilst such developments may provide greater consistency in the implementation of the seismic hazard analysis calculations, the process of constructing the probabilistic seismic hazard input models is far less standardised. Few tools are available to assist the hazard modeller in ensuring consistency of implementation of the steps commonly used in the creation of a seismogenic source model for PSHA, or to understand the impact that modelling decisions, such as algorithm choice or configuration, made early in the model construction process may have upon the earthquake recurrence model, and ultimately the PSHA itself. As part of its commitment to ensure transparency and reproducibility in the whole hazard modelling process, the Global Earthquake Model (GEM) has produced an open-source toolkit for construction of seismogenic source models as input for PSHA. The Hazard Modeller's Toolkit ("hmtk") is a Python library of functions for creation of a PSHA input model, which is constructed following commonly adopted open development practices, and places an important emphasis on code testing to ensure correctness of implementation of the methods within. As the usage of the tools is focused upon PSHA model construction, the "hmtk" contains three categories of tools: seismicity, geology and geodesy. Each category may address different types of workflows, such as the creation of a magnitude frequency distribution within a set of uniform area sources using an observed earthquake catalogue, the model of recurrence on an active fault surface derived from the geological slip, or even the estimation of magnitude recurrence in a tectonically deforming block from the geodetic strain rate. The "hmtk" is also designed to be flexible in integrating different types of information, such as seismicity around an active fault source, or clusters of earthquakes in an aftershock sequence, to allow the hazard modeller a better understanding of the seismotectonics of their study region and to combine different streams of information into a hazard model. Additional facilities for visualisation are also provided. Finally, the "hmtk" takes as a dependency the OpenQuake hazard library, i.e. the python library inside OpenQuake-engine software for seismic hazard and risk assessment in which the scientific code of the hazard calculations is maintained. This link exposes core PSHA hazard calculation functions to the "hmtk", allowing modellers to understand the impact of the modelling decisions on the resulting hazard curves, greatly facilitating the possibility to undertake more detailed sensitivity studies during the model construction process.

[1] Hazard Scientist, Global Earthquake Model (GEM), Pavia, Italy, graeme.weatherill@globalquakemodel.org
[2] PhD Student, Understanding & Managing Extremes (UME) School, Pavia, Italy
[3] Hazard Coordinator, Global Earthquake Model (GEM), Pavia, Italy

INTRODUCTION

The mitigation of earthquake risk through the use of seismic resistant design codes and effective urban planning is ultimately contingent on a good understanding of the seismic hazard on a site-specific, urban and regional scale. The methodology of probabilistic seismic hazard analysis (PSHA) is well established as a means of characterising the level of ground motion to which a location may be subjected to within a given time period (Cornell, 1968; McGuire, 1976). The last decade has witnessed the emergence of several freely available software packages for the calculation of seismic hazard (e.g., CRISIS (Ordaz et al., 2013), OpenSHA (Field et al., 2003) and OpenQuake-engine (Pagani et al., 2014)), which have helped to remove one of the key barriers preventing wider adoption of state-of-the-art seismic hazard analysis practices in many areas of the world. Such software also allow for critical comparisons to be made between the different technical solutions implemented, in order to understand the degree of similarity between results provided by seismic hazard calculators, within controlled conditions (e.g. Thomas et al., 2010; Pagani et al., 2014; Monelli et al., 2014). Whilst the developments outlined previously may provide greater consistency in the implementation of seismic hazard analysis calculations, the process of constructing the probabilistic seismic hazard input models is far less standardised, and few, if any, tools are available that are specifically tailored to this purpose. The situation frequently emerges, therefore, that seismic hazard modellers may adopt tools that, although powerful in certain aspects, are not necessarily well suited to the construction of PSHA models. It then often becomes necessary for seismic hazard modellers to create their own implementations of "standard" methodologies, leading to an absence of implementation consistency and quality assurance, even if the methods selected in the model preparation process are the same. Alternatively, although there may exist available software code to apply particular methodologies, sometimes made available by means of electronic journal supplements, or the websites of individual authors and organisations, such implementations are heterogeneous in terms of coding language, data input and output format, and even in the actual definition of specific parameters themselves. This in itself creates a barrier to understanding the impact of modelling decisions on the resulting probabilistic seismic hazard curves, making the sensitivity studies, which should be integral to any such modelling process, more complicated and potentially prone to error. So whilst open and available software for implementing the seismic hazard calculation is allowing for greater consistency and reproducibility in seismic hazard analysis, far less progress is being made with respect to the model development process.

The Global Earthquake Model (GEM) has a strong commitment to ensuring reproducibility, transparency and accessibility in all aspects of the seismic hazard modelling process. This is in evidence in the manner of the construction of the OpenQuake-engine, a software for calculation of probabilistic seismic hazard and risk constructed following standards (e.g. public code repositories, open licenses) and development processes (e.g. agile, test-driven development, code review) that represent the best-practices in open-source software development (Pagani et al., 2014). There is, therefore, a clear motivation for extending this philosophy to the process of model preparation. This has lead us to the development of the Hazard Modeller's Toolkit, "hmtk" hereafter, an open-source library for the development of probabilistic seismic hazard input models.

The primary objective of the "hmtk" is to explore the full domain of the PSHA model preparation process. The tools are not limited to the analysis of seismicity but also to the characterisation of earthquake recurrence from geological and geodetic data, as well as the representation of uncertainty in the source model. Beginning with an initial overview of the development process, and to some extent the development philosophy from a design perspective, this paper will then outline, with some illustrative examples, the key features of the three main categories of tools:

1. **Seismicity**: For the statistical analysis and processing of an earthquake catalogue, e.g. descriptive statistics, declustering, completeness analysis and earthquake recurrence.
2. **Geology**: For the characterisation of earthquake recurrence on faults from geological slip, comparing different moment-balanced magnitude frequency distributions and modelling epistemic uncertainty

3. **Geodesy**: Experimental tools for the characterisation of activity rate and recurrence from geodetic strain of a seismogenic volume. This is based primarily on the "Seismic Hazard Inferred from Tectonics (SHIFT)" methodology proposed by Bird and Liu (2007) and Bird et al. (2010)

In addition to these tools, the "hmtk" also has built-in functionalities to assist in the interaction between the different data sets and seismogenic source typologies, as well as simple functions to assist in the visualisation of particular elements of each model.

Table 1. Overview of Functionalities of the Hazard Modeller's Toolkit

Tools categories	Data Inputs	Methods
Seismicity	Earthquake Catalogue Source Model Geometry (point, polygon, 3D fault geometry)	1. Descriptive Statistics of the Catalogue 2. Declustering 3. Completeness Analysis 4. Recurrence 5. Statistical Estimates of Maximum Magnitude 6. Smoothed Seismicity
Geology	Active Fault Sources (defined in YAML or GeoJSON format) Tectonic Regionalisation	1. Recurrence models including: Anderson & Luco (1983), Characterstic (Truncated Gaussian), Youngs and Coppersmith (1985) 2. Epistemic Uncertainty Analysis
Geodesy	Strain Rate Model Tectonic Regionalisation	1. Seismic Hazard Inferred from Tectonics (SHIFT) (Bird and Liu, 2007)

DEVELOPMENT PROCESS AND PHILOSPHY

The "hmtk" is intimately linked with the OpenQuake-engine hazard calculation software (Pagani et al., 2014), and this link offers several critical utilities that shall be expanded upon in due course. As with OpenQuake-engine, the "hmtk" is developed in the Python (2.7) programming language and released under the GNU Affero open-source license. The source code of the software is publically open and available in an online repository (https://github.com/GEMScienceTools/hmtk). A live open documentation and tutorial repository (https://github.com/GEMScienceTools/hmtk_docs) is also available to assist the modeller understand both the scientific basis of each methodology, alongside examples of usage. The "hmtk" itself does not necessarily operate as standalone software, but as a library of methods and functions that can be manipulated by the modeller to address the specific needs of a particular workflow. These methods can therefore be linked together by the modeller, typically using auxiliary functions constructed specifically for the seismic hazard context, which should provide the modeller with the flexibility to tailor their usage to the particular needs of the region or context in which they are working.

The architecture of library in its current form is created with the intention of addressing three key objectives:

i) *Adaptability*, i.e. cleaner separation of methods into self-contained components that can be implemented and tested independently of the remainder of the code.

ii) *Abstraction*, i.e. the specification of core behaviours of a class of methods, such that methods implemented within this class can be used interchangeably within the workflow. For example, the process of earthquake declustering can be described in terms of a set of common inputs (or, more realistically, sets of inputs) and outputs. The inputs in this case will be an earthquake catalogue and a set of parameters controlling the configuration of the algorithm that can be expanded in accordance with the requirements of the algorithm. The outputs will take the form of an index of the cluster number (i.e. the cluster to which each earthquake is assigned) and a marker to indicate whether the earthquake is a mainshock, foreshock or aftershock. Only by defining, for each class of methods, these common elements across the methods does it become possible to use methods interchangeably.

iii) *Usability and Portability*, i.e. the means to allow the library to be incorporated into larger applications or programmes for earthquake hazard and/or risk.

Within the scope of these objectives a strong emphasis is placed on developing a testing process that is integrated with the software. Adopting the strategies commonly followed for test-driven development, a concerted effort is made to ensure that methods implemented within the software are accompanied by unit-tests, i.e. tests that check for the correctness of behaviour of the code. The strong emphasis placed on testing presents important and often novel challenges in understanding the source modelling process. In cases such as seismicity, the tools themselves are investigative, and the behaviour of particular methods can heavily dependent on the data to which they are applied. Therefore, designing tests that will ensure correctness in the implementation of the method, especially under extreme cases such as those found in low seismicity regions, are complex to construct. For this purpose we often adopt synthetic catalogues with known properties such as b-value, maximum magnitude, completeness, manually inserted clusters. Tests of, for example, the tools to calculate earthquake recurrence will attempt to verify that the recurrence calculators correctly estimate the known recurrence properties of the synthetically generated catalogue.

The integration with OpenQuake-engine manifests itself in the shared dependency of the OpenQuake-hazardlib, the lightweight python library that forms the basis of the seismic hazard calculator in the OpenQuake-engine. The critical advantages this offers are: i) consistency in the definition of the seismogenic sources between model preparation and implementation process, ii) shared functionality, particularly with respect to the calculation of source geometry, and iii) direct implementation of low-level PSHA calculation tools to allow the modeller the means to generate simple seismic hazard curves in order to understand the real impact of the modelling decisions on the resulting hazard analysis. This unique functionality ensures that the modeller has the means to construct their model with a clearer perspective on the implications in a hazard context.

SEISMICITY TOOLS

The flexibility of the "hmtk" is well illustrated in the case of the seismicity tools. The process of analysing the earthquake catalogue for the characterisation of earthquake recurrence implies a well-established methodology, with clearly defined steps that a modeller will follow in the development of the seismogenic source. For example, in the estimation of the stationary rate of seismicity it is necessary to identify and purge the non-Poissonian events in the catalogue (declustering), and to provide estimates for the time periods over which particular magnitudes are assumed to be fully observed (completeness), before using methods widely adopted for estimating the recurrence properties of the source. The hazard modeller may also require the use of methods without necessarily wishing to integrate them into a complete workflow. In the same manner, when the objective is only directed towards the analysis of the catalogue itself, or a selected subset of it and not necessarily in the frame of PSHA source model construction, it can be preferable to use the methods in independently.

The seismicity tools are divided into six sections that can be integrated into a full workflow for the preparation of a seismic source model (Table 1): descriptive statistics, declustering, completeness, recurrence, maximum magnitude, smoothing. This category uses two inputs: the earthquake catalogue and the source model geometries. The catalogue is evidently needed, whereas the sources are mandatory only as part of the process to build an input model to be used on the OpenQuake-engine. When loaded into the toolkit, the catalogue can be analysed and explored with a set of basic statistics methods. These include the generation of one or two-dimensional histograms with bootstrapping to take into account errors on specific parameters (such as magnitude or depth), or simple utilities such as chronological sorting or filtering. Example of visualisations of the earthquake catalogue and some of the simple histograms that are associated with the catalogue are shown in Figure 1. More discussion of the visualisation tools will follow in due course.
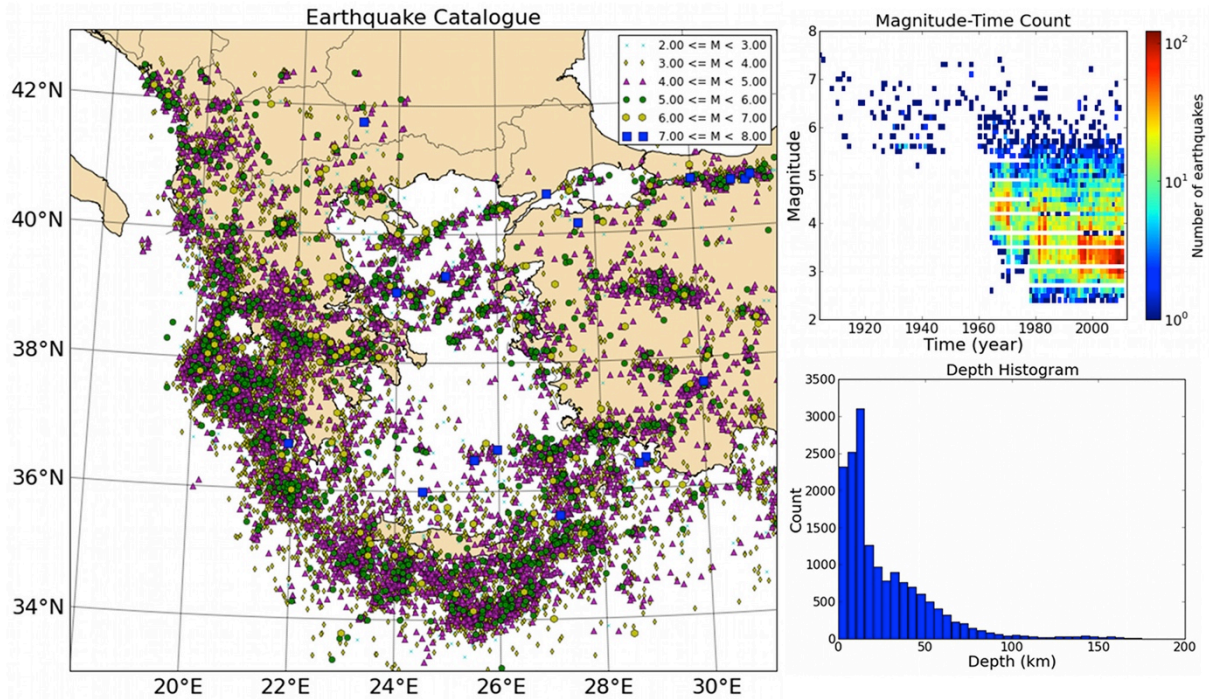
Figure 1. Example of catalogue visualisation completed with the hazard modeller's toolkit: catalogue mapping (*left*), magnitude-time density plot (*top right*) and simple histogram of depths (*bottom right*).

At the time of writing, the "hmtk" contains two common declustering algorithms: Gardner and Knopoff (1974) and AFTERAN (Musson, 1999). More declustering methods are intended to be implemented, either by the authors or with contributions from users. As outlined previously, the output of these algorithms are two arrays of indexes, the first index indicates the cluster number to which each event is assigned, and a second index describes whether the event is a foreshock, aftershock or mainshock. To construct a simple set of unit tests for the algorithms, we created small synthetic catalogues with clusters inserted by hand under heavily idealised conditions (i.e. within a very short time of the mainshock and a very close location, without "background" events). A test for the declustering algorithm was the recognition of the manually inserted cluster events. Statistical tests to check that the declustered catalogue can be considered Poissonian, such as the original test proposed by Gardner and Knopoff (1974), are planned for future versions.

Since each of the algorithms for estimating the parameters the a- and b-value of the Gutenberg-Richter distribution implemented in the hmtk (currently those of Weichert (1980) and Kijko and Smit (2012), in addition to a time-averaged maximum likelihood estimator) assumes time-varying completeness in the earthquake catalogue, the completeness algorithms must be able to produce an estimate of the completeness magnitude (M_C) with time. Currently, the two algorithms supported (Stepp (1971) and cumulative rate analysis) produce as output a completeness table indicating the magnitude of completeness in each time interval and the earliest time in the catalogue to which it corresponds. The architecture of this process is designed to ensure that the temporal completeness properties of the catalogue are well accounted for in the recurrence estimation process. Using the selection tools, the spatial variation in completeness in the earthquake catalogue can also be explored. As indicated previously, tests for the correctness of the implementation of the seismicity recurrence estimators use synthetic catalogues with known properties (including b-value and temporal variation in completeness), ensuring that the completeness and recurrence methods provide values of the output parameters as expected.

The remaining functionalities in the toolkit implement an initial set of methods for the inference of maximum magnitude from seismicity, and the smoothed seismicity methodology of Frankel (1995). The maximum magnitude tools contain the algorithms of Kijko (2004) and an adaptation of the cumulative strain energy method of Makropoulos and Burton (1983). The latter has been adapted for usage with cumulative moment, and the option is available to infer uncertainty on M_{MAX} making a bootstrap sampling of magnitude uncertainties.

The smoothed seismicity tools are in an early stage of development and therefore the offered functionalities remain limited. Whilst many different bivariate kernels could be invoked to undertake the smoothing process, it is not necessarily the choice of kernel that will drastically affect the results, but the manner in which the smoothing is applied, and in particular, the effective bandwidth (or similar parameter) that controls the degree of smoothing. In an object-oriented computational framework such as the one implemented inside the "hmtk", particular effort is needed in understanding potential suite of tools that might be needed for a particular functionality, in order to define clearly the common elements between the possible tools. In the case of smoothed seismicity this is challenging, as not only are the potential kernel functions different, the manner in which the smoothing is implemented can also vary considerably. The Frankel (1995) algorithm provided an obvious choice for an initial smoothing methodology in a hazard modeller's toolkit largely because of its clear precedent for use in a seismic hazard context. This is particularly important, as the code of Frankel (1995) compensates for time-variation in catalogue completeness, assuming a fixed b-value. Nevertheless, the method of Frankel (1995) has some details of the implementation that distinguish it from other type of kernel smoothing approaches, even those such as Zechar and Jordan (2010) which are adopting the same kernel type. Scrutiny of existing literature on smoothed seismicity methods leads to the recognition that divergence does not just exist in the choice of kernel, but also in the implementation. Therefore additional smoothed seismicity methods are in development that will largely be treated as stand-alone smoothed seismicity tools.

GEOLOGICAL TOOLS

In many regions of the world earthquake fault geology is now a common feature of PSHA models. In spite of this widespread adoption, the manner in which earthquake recurrence is characterized from estimates of the slip is still an area of some diversity (Bungum, 2007). Sidestepping, to some extent, debates regarding the use of characteristic, exponential or hybrid models, the geological tools allow for the recurrence from different models to be compared, assuming the same geological characteristics of the fault. At the time of writing, four different types of fault recurrence model are available: i) a set exponential models (Type 1, 2 and 3 based on the formulation of Anderson and Luco (1983) in which the recurrence relation is referred to the total area of the fault (which we describe by the term "Arbitrary"), ii) a set of exponential models (Type 1, 2 and 3) based on the formulation of Anderson and Luco (1983) in which the recurrence is referred to the rupture area of the maximum earthquake (which we describe by the term "Area-Mmax", iii) a characteristic model in which the rate is distributed according to a truncated Gaussian distribution, and iv) the exponential and hybrid models of Youngs and Coppersmith (1985). A comparison of the different recurrence models, assuming the same fault configuration, is shown in Figure 2, with corresponding parameters listed in Table 2.

The comparison of magnitude-frequency distribution models based on geological data may be a helpful tool, but it alone does not necessarily serve the purpose of constructing a fault source model for probabilistic seismic hazard analysis. Built into the geological tools, however, is support for the characterization of epistemic uncertainty on parameters such as slip-rate, shear modulus, magnitude-scaling relation (and its variability) and supported magnitude frequency distribution parameters (e.g. b-value). These can be used to develop a seismogenic fault model input to OpenQuake with activity rates reflecting the model-to-model variability.

Whilst the geological tools can be used in isolation from those of seismicity, it is also possible to link them together, with the help of ancillary functions. For example seismicity around an active fault can be selected, using Joyner-Boore distance (i.e. distance to the surface projection of the fault rupture) or distance to the rupture, and comparisons can be made between the activity rates derived from geology and those from the surrounding seismicity. Such comparisons may assist in selecting, or at least appropriately weighting, the preferred form of the magnitude frequency distribution. Conversely, the modeler may even prefer to identify the possibly range of slip parameters (e.g. slip rate, seismogenic coupling coefficient, coupling depths etc.) that may suitable account for the observed seismicity on and around the fault.

Table 2: Configuration of the geological properties of the fault and magnitude frequency distributions used for comparison in Figure 2.

Fault Configuration	
Length = 100 km Width = 20 km Area = 2000 km Slip-Type = Strike-Slip (Rake = 0˚)	Slip-Rate = 10 mm/yr Shear Modulus = 30 GPa Magnitude Scaling Relation = Wells & Coppersmith (1994)
Recurrence Model	**Parameters**
Anderson & Luco (1983) "Arbitrary"	b-value = 0.8 Minimum Magnitude (M_{MIN}) = 5.0
Anderson & Luco (1983) "Area-Mmax"	b-value = 0.8 Displacement/Length = 1.25×10^{-5} M_{MIN} = 5.0
Characteristic	Standard Deviation (σ) = 0.15 magnitude units Truncation Level = $\pm 3\ \sigma$
Youngs & Coppersmith (1985) Exponential	b-value = 0.8
Youngs & Coppersmith (1985) Hybrid	b-value = 0.8 $\Delta M = 0.5$
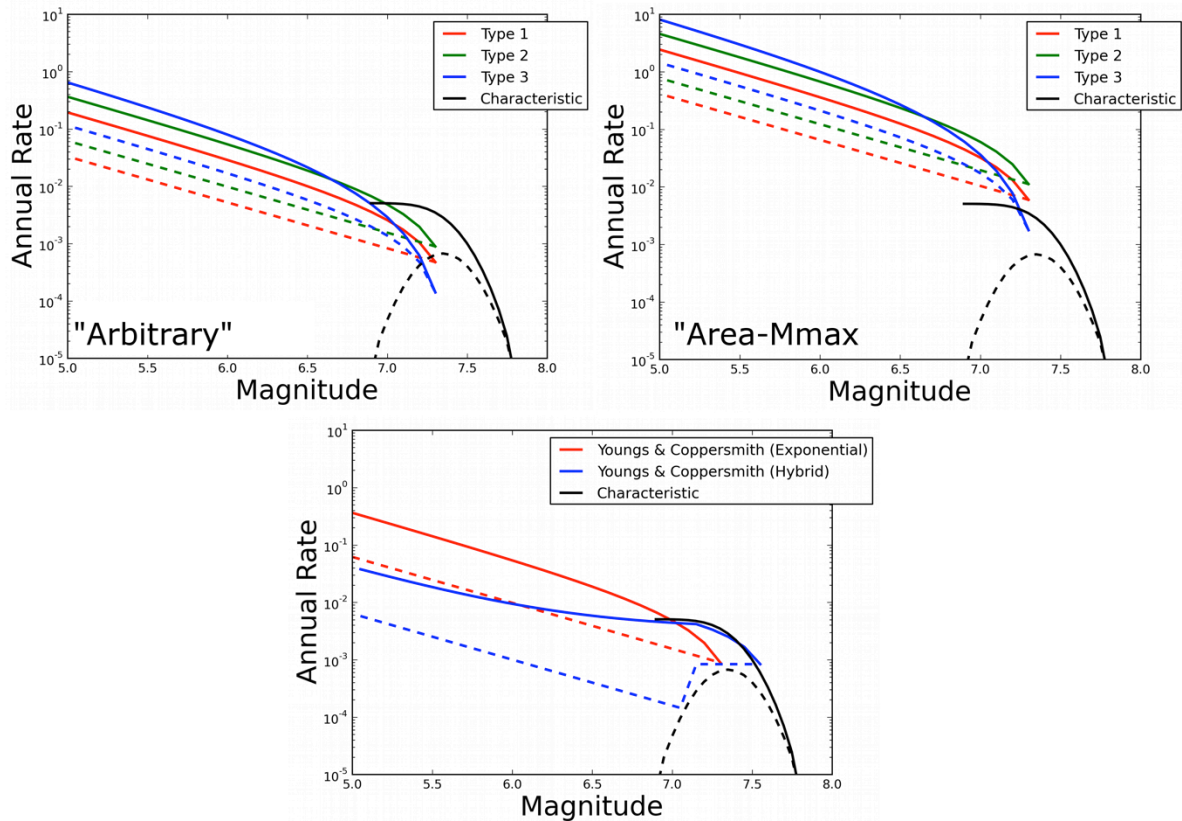
Figure 2. Comparison of magnitude frequency distributions with using the fault configuration shown in Table 2. Anderson and Luco (1983) "Arbitrary" type (*top left*), Anderson and Luco (1983) "Area-Mmax" type (*top right*), and Youngs and Coppersmith (1985) (*bottom)*

The incorporation of geological tools into the "hmtk" may help in the initial stages of model construction in regions where geological information is becoming sufficient to characterize fault sources in PSHA. Further work is ongoing to expand the fault data structure to incorporate more detailed paleoseismic information, including displacements observed at sites and detailed slip histories for the fault. This can help strengthen a critical link between the earthquake geology and hazard modeling communities by making paleoseismic data visible to the seismogenic source characterization. It is hoped that by putting these tools in place, seismic hazard modellers will have

new means of exploring the links between geology and seismicity (including historical earthquakes), optimally leading towards the creation of models in which critical conditions, such as the balance of seismic moment, are more carefully addressed.

STRAIN TOOLS

The third category of tools is more experimental as it is designed to help begin the process of incorporating geodetic strain information into the seismic hazard model. As of the time of writing, the geodetic tools support one established methodology for creating activity rate models from geodetic strain. This methodology is the "Seismic Hazard Inferred from Tectonics (SHIFT)" approach of Bird and Liu (2007), applied globally by Bird et al. (2010). Detail on the methodology itself can be found in the publications highlighted, but in essence the SHIFT methodology is one of several possible approaches to relate strain rate across a deforming continua to geodetic moment rate and subsequently, using a moment-balanced magnitude frequency distribution, to a seismic activity rate. For a two-dimensional strain rate tensor described by three independent terms $\dot{\varepsilon}_{xx}$, $\dot{\varepsilon}_{yy}$ and $\dot{\varepsilon}_{xy}$. At the ground surface the three principal strain rate directions ($\dot{\varepsilon}_1, \dot{\varepsilon}_2$ and $\dot{\varepsilon}_3$ ordered such that $\dot{\varepsilon}_1 \leq \dot{\varepsilon}_2 \leq \dot{\varepsilon}_3$) can be derived, assuming incompressibility ($\dot{\varepsilon}_1 + \dot{\varepsilon}_2 + \dot{\varepsilon}_3 = 0$) such that:

$$\dot{\varepsilon}_1 = \frac{\dot{\varepsilon}_{xx} + \dot{\varepsilon}_{yy}}{2} - \sqrt{\dot{\varepsilon}_{xy}^2 + \left(\dot{\varepsilon}_{xx} + \dot{\varepsilon}_{yy}\right)^2 / 4} \text{ and } \dot{\varepsilon}_2 = \frac{\dot{\varepsilon}_{xx} + \dot{\varepsilon}_{yy}}{2} + \sqrt{\dot{\varepsilon}_{xy}^2 + \left(\dot{\varepsilon}_{xx} + \dot{\varepsilon}_{yy}\right)^2 / 4} \quad (1)$$

The seismic moment production of an area (*A*) of lithosphere is estimated according to:

$$M_0 = A \langle cz \rangle \mu \begin{cases} 2\varepsilon_3 & \text{if } \varepsilon_2 < 0 \\ -2\varepsilon_1 & \text{if } \varepsilon_2 \geq 0 \end{cases} \quad (2)$$

where $\langle cz \rangle$ is the coupled seismogenic thickness of the lithosphere, and µ is the elastic shear modulus. This derivation of moment rate from strain rate for a deforming continuum largely emerges from the model of Kostrov (1974), with other modifications proposed by Ward (2007) and Savage and Simpson (1997). Support for alternative models could be implemented in future.

To define the magnitude frequency distribution from the moment rate, Bird and Liu (2007) propose the use of the tapered Gutenberg-Richter distribution:

$$\dot{N}\left(m > m_T\right) = \dot{N}^{CMT} \left(\frac{\dot{M}_0}{\dot{M}_0^{CMT}}\right) \left(\frac{M_0\left(m_T\right)}{M_0\left(m_T^{CMT}\right)}\right)^{-\beta} \exp\left(\frac{M_0\left(m_T^{CMT}\right) - M_0\left(m_T\right)}{M_0\left(m_c\right)}\right) \quad (3)$$

In this formulation the parameters of the tapered Gutenberg-Richter distribution (namely β, the exponent, and m_C the "corner magnitude") are constrained according to the regionalisation initially presented by Bird and Kagan (2004). This same region also provides the coupled seismogenic thickness and the elastic shear modulus. The SHIFT model itself adjusts the activity rates and moment rates with respect to those found in the Global Centroid Moment Tensor (CMT) catalogue (Ekström et al., 2012).

To allow for the development of an accurate testing strategy, the geodetic tools implement faithfully the version of SHIFT presented in Bird et al. (2010). It is recognized, however, that this methodology is strongly dependent on the regionalisation parameters proposed by Bird and Kagan (2004), which may be something that a modeler may wish to adapt to more appropriate values, given the context. It is possible to supply different values, according to a tectonic regionalisation defined by the user. However, at present tools to allow the user to update or re-adjust the regionalisation

parameters from existing data, using the methodologies presented Bird and Kagan (2004), is not yet available.

As the geodetic strain rate tools themselves are currently experimental, it is important to emphasise that we anticipate their development to evolve, partly in response to feedback and requirements from users of the software, and partly in response to future developments that may occur in this area. We may also hope that such tools could, as with the geological tools, provide a means of linking the different areas of seismotectonics, geology and geodesy into a more tightly connected environment for the construction of seismic hazard models.
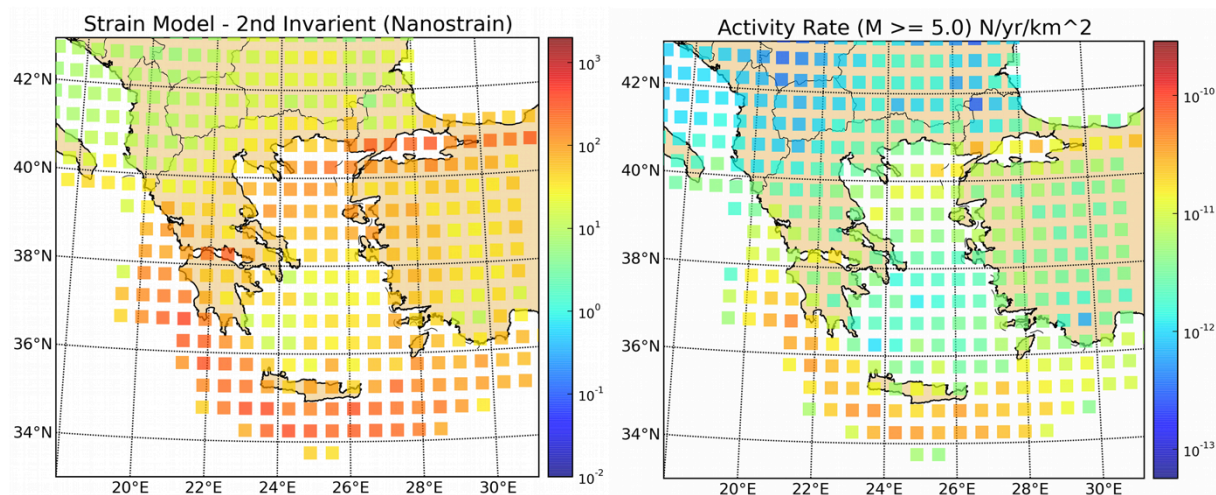

Figure 3. Example usage of the "hmtk" geodetic tools to derive an activity rate map from a geodetic strain rate model. Geodetic strain rate from the Global Strain Rate Model (Kreemer, 2003) (*left*) and corresponding activity rate (M ≥ 5.0 yr^{-1} km^{-2}) from the "SHIFT" methodology in the "hmtk" (*right*)

VISUALISATION AND INTERACTIVITY

Whilst the primary objective in the creation of the "hmtk" is the implementation of the scientific functionalities, the importance of good visual tools must not be underestimated. The development of a full graphical user interface, whilst desirable as a long-term objective, presents significant challenges to ensure that the user has full operability of the functions and can customize the necessary workflows. To enhance the visual capabilities of the "hmtk", however, we include a set of built-in plotting functions that should allow the modeller to generate pertinent graphics and maps to understand the properties of the data and the results of the analysis. Selected examples of such plots have been shown in the figures within this paper. New plotting functions are added regularly; however, all plots are generated using the widely-support python plotting library (Matplotlib) and its mapping extension (Basemap), so the modeler may also have the means to create their own plots with these same libraries if they are familiar with their basic functionality.

A novel development in the Python software language that greatly enhances the usage of the "hmtk" is the interactive Python (IPython) notebook (http://ipython.org/notebook.html). The IPython notebook provides an environment in which the interactive python console can be embedded in a web-based environment, where Python commands can be not only executed but also annotated with text, markup language and graphical figures. This provides a perfect setting in which a user of the "hmtk" can explore the tools provided for construction of the hazard model and save particular workflows (processes that link together multiple tools) to allow them to be retrieved, recalculated and modified at a later time. An example of the "hmtk" usage in an IPython notebook environment is shown in Figure 4.
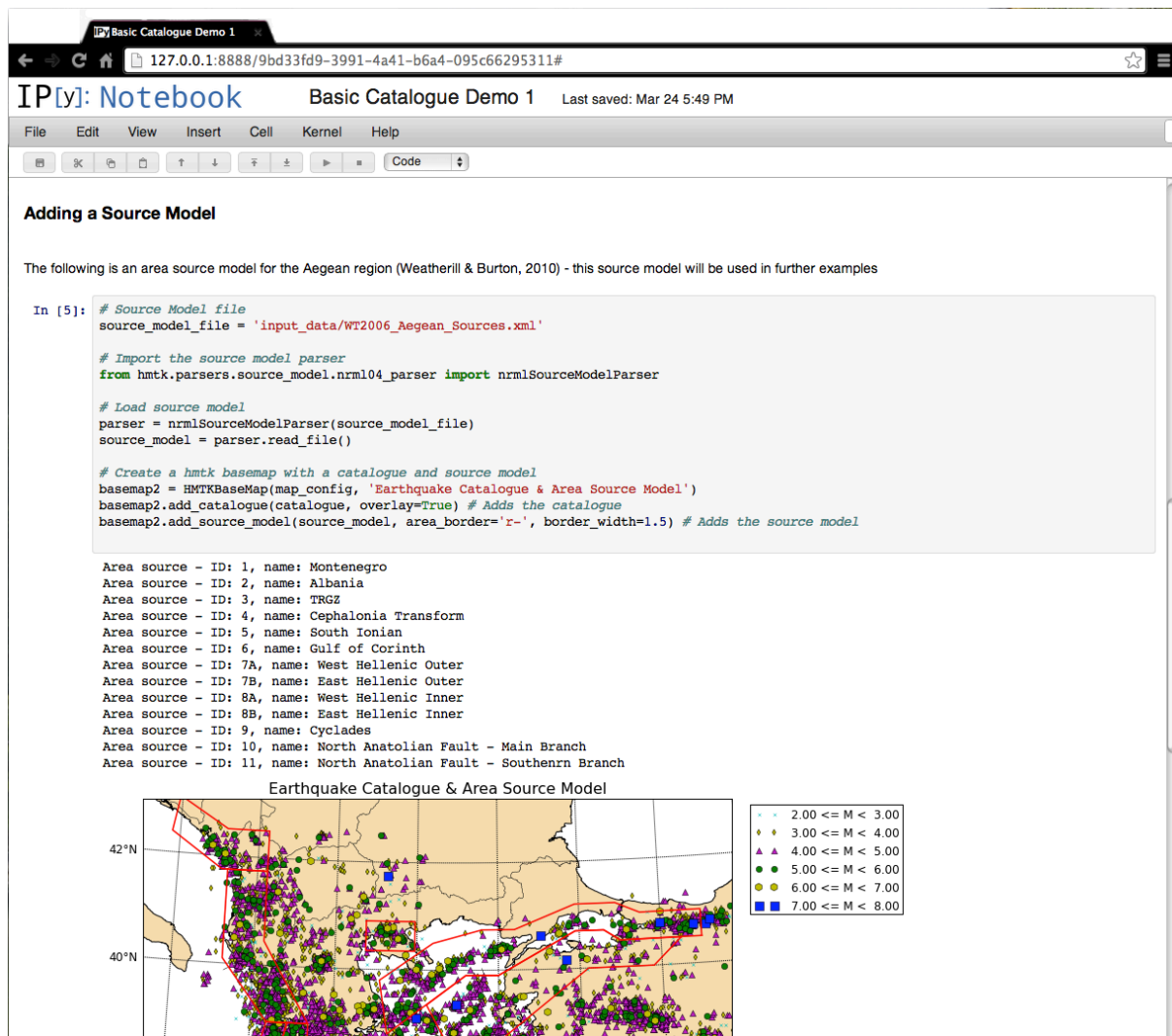
Basic Catalogue Demo 1

127.0.0.1:8888/9bd33fd9-3991-4a41-b6a4-095c66295311#

IP[y]: Notebook   Basic Catalogue Demo 1   Last saved: Mar 24 5:49 PM

File   Edit   View   Insert   Cell   Kernel   Help

Code

Adding a Source Model

The following is an area source model for the Aegean region (Weatherill & Burton, 2010) - this source model will be used in further examples

```
In [5]:  # Source Model file
         source_model_file = 'input_data/WT2006_Aegean_Sources.xml'

         # Import the source model parser
         from hmtk.parsers.source_model.nrml04_parser import nrmlSourceModelParser

         # Load source model
         parser = nrmlSourceModelParser(source_model_file)
         source_model = parser.read_file()

         # Create a hmtk basemap with a catalogue and source model
         basemap2 = HMTKBaseMap(map_config, 'Earthquake Catalogue & Area Source Model')
         basemap2.add_catalogue(catalogue, overlay=True) # Adds the catalogue
         basemap2.add_source_model(source_model, area_border='r-', border_width=1.5) # Adds the source model
```

```
Area source - ID: 1, name: Montenegro
Area source - ID: 2, name: Albania
Area source - ID: 3, name: TRGZ
Area source - ID: 4, name: Cephalonia Transform
Area source - ID: 5, name: South Ionian
Area source - ID: 6, name: Gulf of Corinth
Area source - ID: 7A, name: West Hellenic Outer
Area source - ID: 7B, name: East Hellenic Outer
Area source - ID: 8A, name: West Hellenic Inner
Area source - ID: 8B, name: East Hellenic Inner
Area source - ID: 9, name: Cyclades
Area source - ID: 10, name: North Anatolian Fault - Main Branch
Area source - ID: 11, name: North Anatolian Fault - Southenrn Branch
```

Earthquake Catalogue & Area Source Model

2.00 <= M < 3.00
3.00 <= M < 4.00
4.00 <= M < 5.00
5.00 <= M < 6.00
6.00 <= M < 7.00
7.00 <= M < 8.00

Figure 4. Illustration of the "hmtk" used within an IPython Notebook environment.

CONCLUSIONS

The development of the "hmtk" is intended to move toward the key objectives of adoption of open tools, transparency and reproducibility in the hazard modelling process. It is also hoped that they provide a natural, open repository in which earthquake scientists engaged in different aspects of the hazard model development process can contribute new tools and methodologies allow for greater ease of use, and ultimately greater understanding and wider usage of different methods used in the model construction process. In designing the architecture of the "hmtk", however, and in thinking critically about the nature of testing and quality assurance in the model construction process, new challenges are frequently encountered. In attempting to adopt a test-driven approach to development, greater attention will need to be paid to the means of testing, not just to ensure the correctness of implementation of the functionalities, but also to understand the factors that can have the greatest influence both on the characteristics of the seismogenic source model and on the resulting hazard analyses. Such insights may form a critical step in the adoption of "good-practice" in PSHA model construction worldwide. Furthermore, they may also create opportunities to explore new ideas for integration of different types of seismic, seismotectonic and geological information into the PSHA modelling process.

REFERENCES

Anderson JG and Luco JE (1983) "Consequences of Slip Rate Constraints on Earthquake Occurrence Relations," *Bulletin of the Seismological Society of America*. 73: 471 – 496

Bird P and Kagan YY (2004) "Plate-Tectonic Analysis of shallow Seismicity: Apparent Boundary Width, Beta, Corner Magnitude, Coupled Lithosphere Thickness and Coupling in Seven Tectonic Settings," *Bulletin of the Seismological Society of America*, 94(6): 2380 - 2399

Bird P and Liu Z (2007) "Seismic Hazard Inferred from Tectonics: California," *Seismological Research Letters*. 78(1): 37 – 48

Bird P, Kreemer C, Holt WE (2010) "A Long-Term Forecast of Shallow Seismicity Based on the Global Strain Rate Map," *Seismological Research Letters*. 81(2): 184 - 194

Bungum H (2007) "Numerical Modelling of Fault Activities," *Computers and Geosciences*. 33: 808 - 820

Cornell, CA, (1968) "Engineering Seismic Risk Analysis," *Bulletin of the Seismological Society of America*. 58(5): 1583 – 1606

Ekström G, Nettles M, Dziewonski AM (2012) "The Global CMT Project 2005 – 2010: Centroid Moment Tensors for 13,017 Earthquakes," *Physics of the Earth and Planetary Interiors*. 200 – 201: 1 – 9.

Field EH, Jordan TH, Cornell CA (2003) "OpenSHA: A developing community-modelling environment for seismic hazard analysis," *Seismological Research Letters,* 74: 406 – 419

Frankel, A (1995) "Mapping Seismic Hazard in the Central and Eastern United States," *Seismological Research Letters*, 66(4): 8 - 21

Gardner JK and Knopoff L (1974) "Is the Sequence of Earthquakes in Southern California, with Aftershocks Removed, Poissonian?," *Bulletin of the Seismological Society of America*. 64: 1363 - 1367

Kijko A (2004) "Estimation of Earthquake Maximum Magnitude, m_{max}," *Pure and Applied Geophysics,* 161: 1655 - 1681

Kijko A and Smit A (2012) "Extension of the Aki-Utsu *b*-Value Estimator for Incomplete Catalogues," *Bulletin of the Seismological Society of America,* 102(3): 1283 – 1287

Kostrov VV (1974) "Seismic Moment and Energy of Earthquakes and Seismic Flow of Rock," *Izy. Acad. Sci. USSR Physics of the Solid Earth*. 1: 13 – 21

Kreemer C, Holt WH, Haines AJ (2003) "An integrated global model of present-day plate motions and plate boundary deformation," Geophysical Journal International, 154(1): 8 - 34

Makropoulos KC and Burton PW (1983) "Seismic Risk of Circum-Pacific Earthquakes I: Strain Energy Release," *Pure and Applied Geophysics*, 121: 247 - 266

McGuire RK, (1976) "Fortran programme for Seismic Risk Analysis," *US Geological Survey Open-File Report,* 76-67

Monelli D, Pagani M, Weatherill G, Danciu L, Garcia J, (2014) "Modeling Distributed Seismicity for Probabilistic Seismic-Hazard Analysis: Implementation and Insights with the OpenQuake Engine," *Bulletin of the Seismological. Society of America*, 104(4), *In press*

Musson RMW (1999) "Probabilistic Seismic Hazard Maps for the North Balkan Region," *Annali di Geofisica*. 42: 1109 - 1124

Pagani, M, Monelli, D, Weatherill, G, Danciu, L, Crowley, H, Silva, V, Henshaw, P, Butler, L, Nastasi, M, Panzeri, L, Simionato, M, Vigano, D. (2014) "OpenQuake-engine: An Open Hazard (and Risk) Software for the Global Earthquake Model," *Seismological Research Letters*, *In press.*

Ordaz M, Martinelli F, D'Amico V, Meletti C. (2013) "CRISIS2008: A Flexible Tool to Perform Probabilistic Seismic Hazard Assessment." *Seismological Research Letters*. 84(3): 495 – 504

Savage JC and Simpson RW (1997) "Surface Strain Accumulation and the Seismic Moment Tensor," *Bulletin of the Seismological Society of America*. 87(5): 1345 - 1353

Stepp JC (1971) <u>An investigation of earthquake risk in the Puget Sound area by use of the Type I Distribution of largest extremes</u>, Ph.D. Thesis, Pennsylvania State University, Pennsylvania, United States

Thomas P, Wong I, Abrahamson, NA (2010) <u>Verification of Probabilistic Seismic Hazard Analysis Computer Programs</u>, Pacific Earthquake Engineering Reseach Center (PEER) Technical Report, 2010/106, University of California, Berkeley, California

Ward, SN (2007) "Methods for Evaluating Earthquake Potential and Likelihood in and around California," *Seismological Research Letters*. 78(1): 121 - 133

Weichert DH (1980) "Estimation of the Earthquake Recurrence Parameters for Unequal Observation Periods of Different Magnitudes," *Bulletin of the Seismological Society of America*, 70(4): 1337 - 1346

Wells DL and Coppersmith KJ (1994) "New Empirical Relationships Among Magnitude, Rupture Length, Rupture Width, Rupture Area and Surface Displacement," *Bulletin of the Seismological Society of America*, 84: 974 - 1002

Youngs RR and Coppersmith KJ (1985) "Implications of Fault Slip Rates and Earthquake Recurrence Models to Probabilistic Seismic Hazard Estimates," *Bulletin of the Seismological Society of America*. 75: 939 – 964

Zechar JD and Jordan, TH (2010) "Simple smoothed seismicity earthquake forecasts for Italy," *Annals of Geophysics*, 53(3): 99 - 105