Referee Report on doctoral thesis

“**Automated determination of earthquake source parameters**”

by RNDr Jiři Vackář

I think the advantage and benefit from this doctoral thesis is the development of a new, fully automated software tool for CMT inversion. This software package, and its associated sub-packages, is directly applicable to the daily routine data processing of various networks and for subsequent processing of large datasets after acquisition.

Overall, this is a well-written extensive thesis that develops publicly available algorithms and tools to test and model the CMT inversion procedure in a Bayesian framework. With some minor clarifications this is a well-organized doctoral thesis that fully deserves the PhD title.

In the following, I will make comments on various minor issues and potential clarifications on a per-chapter basis:

Introduction:

This is a short and concise introduction of the thesis subject and the approach followed. The candidate states clearly that his aim is to develop a highly automated processing schema. The structure of the thesis is described. This is necessary for the reader in order to understand the sequence of the chapters.

Chapter 1:

This is a short chapter that aims to describe the strong dependence of seismic source studies on the velocity models. On my personal taste, I would combined it with Chapter 2 but that implies that the following chapter would not be in a form of a published paper.

Chapter 2:

This chapter is actually a very interesting and innovative study concerning the long-period observed waves recorded at local and regional distances from the Efpalio 2010 earthquake. A combination with Chapter 1 with a description of full waveform inversion approach to obtain local velocity models would benefit the reader but it is also fully acceptable as it is.

The chosen subset of stations with observed strong PL waves contains 9 out of 17 stations marked as red in Figure 2.1. Why not all? In Loutraki area, in 2 adjacent stations only one has recorded the PL wave. Perhaps LTRA has a short period sensor.

The filtering in the inversion is flat between 10-20s whereas the observed PL waves have a period of 5-8s. I would expect that the flat part of the bandpass filter should include the observed periods of the fast long period waves.

Chapter 3:

This chapter is simple and concise. Although, it is a mathematical chapter, it is written in a clear manner that provides the reader a way to understand the Bayesian formulation of the inverse problem of centroid moment tensor.

Chapter 4:

The sources of uncertainty of the inverted parameters are generally described. Among those, uncertainties on instrument parameters and the Green function construction can not be dealt in this automatic Bayesian approach. Instrument disturbances are detected beforehand and excluded from the analysis whereas seismic noise, as the most common source of uncertainty in the inversion, is accounted for and incorporated in the data covariance matrix as weighting schema.

Chapter 5:

This chapter deals with the issue of detecting long period disturbances in seismic records. It is a common problem for analysts that use seismic records for source studies. The chapter is actually a published SRL paper that presents an automatic code to detect such disturbances in databases of seismic records. The *MouseTrap* code and its associate *SwissMouse* code have potential applications to other large datasets. In a nutshell, it is a very practical and useful chapter.

Chapter 6:

The data covariance matrix of the developed method is described in this chapter. It is straightforward. Perhaps some clarifications on how the **L** matrices are practically constructed from the $\mathbf{C_D}$ matrix are needed.

Chapter 7:
This chapter describes in detail the application and development of the method in a software package and the selection of the corresponding parameters. It is a technical text totally needed for the users of the newly developed code. The final presentation of the results and the selection and design of the plots is very useful for visual inspection. Points of clarifications:

The minimal distance ensures that the point source approximation is valid but for a large event (M>=5) might include clipped records. (As an example, for M_L=6 most records in a distance of 100km will be clipped. The corresponding minimal distance is only 20km.)

The *ArcLink* plugins in the code will soon be outdated since this technology is going to be totally replaced from the *fdsnws* webservices by the European EIDA datacenters. However, they should be easily implemented by new updated *ObsPy* libraries. The same holds for the retrieval of station metadata (*fdsnws_station*).

A database of pre-calculated Green functions in various bands and lengths might leverage the computational cost. As the author clearly states, this is an open-source code under GNU/GPL license so any interested developer might improve or alter it at later stages.

Chapter 8:

In this chapter various necessary synthetic tests show that the developed method is suitable to access and suppress the uncertainty of the CMT solutions taking into account the stationary seismic noise.

The newly developed code is applied in the very dense and well instrumented Swiss seismic network. The choice of this dataset is suitable for testing in reality the code. Switzerland shows low to moderate seismicity. Since 1999 there is a complete digital database with most stations well installed and equipped (most BB stations are equipped with STS-2). There are also appropriate velocity models for the area, local and regional MT databases and a great deal of background information. Surprisingly, the newly developed automatic code obtained 13 totally new events that were added in the Swiss MT database.

Chapter 9:
The future directions of this newly developed code are many. The author describes some of those but as he clearly states the new developments do not need to be done by himself only.

In the following section I will briefly describe some key aspects of the thesis in terms of:

<u>New scientific result</u>

All published papers in SCI referred journals constitute new scientific results. The main results of this thesis have been published as three papers in well-respected seismological and geophysical journals.

The first paper studies the strong fast long-period waves recorded during the Efpalio 2010 earthquake. It is the first time that such disturbances have been studied at local and regional distances. A full waveform inversion gave a suite of successful velocity models, not known before, that explain this observation. A numerical modeling of the multilayer media and three different methods of seismogram simulation were combined.

The second paper explore the automated detection of long-period disturbances in seismic records and develops the *MouseTrap* code. To my knowledge, this is the first code that detects such disturbances and has a great potential for future studies.

In the third paper, that is actually the core of the thesis, a newly developed method and code is developed to provide a fully automated CMT inversion that detects some disturbances in seismic records and constructs a data covariance matrix from the pre-event noise. This data covariance matrix works as an automated frequency filter and station weighting accounting for the recorded seismic noise. The Bayesian approach provides the posterior probability density function and the code itself is optimized to speed-up the inversion process.

<u>Importance and possible applications</u>

It is of great importance to network operators and seismological observatories to have a good and reliable method for automated CMT inversion. With this, large datasets with small and moderate events can be easily processed. The application to the Swiss dataset is a representative example. Rapid determination of MTs attached to near-realtime processing systems is the real challenge and future application of the developed method. Such successful implementation of the code eliminates the need of a 24/7 watch and analysis and produces fast results that might feed other processes in workflow observatory practices. Of course, this requires tests and estimates on calculation speed but a possible inclusion of GPUs Python modules on systems with multi-core graphic processing units might reduce considerably the inversion duration. A separate module that triggers calculation should be developed to “listen” to near-realtime processing

software (eg. *SeisComp3*). Application to larger than M_W 5 events should account for clipped records, non valid point source approximation and the inclusion of strong motion records. The method was successfully tested in a dense and well instrumented network. A future application to a sparser seismic network with great seismicity rates will be the new challenge.
The *MouseTrap* code could be separately included and applied as a QC tool on large databases. Potential application on specific earthquake catalogs in combination with the data availability of the European Integrated Data Archives (EIDA) will provide a European database of such disturbances. A correlation study of such disturbances with local site effects including strong motion records is also possible. Finally, the newly developed code should be applied to records of volcanic seismic networks to detect possible tilts.

To conclude, this is a well-organized doctoral thesis that fully deserves the PhD title.

Athens, August 20, 2018

Dr. Christos P. Evangelidis
National Observatory of Athens