Prague, August 16, 2018

Report on the PhD thesis of Jiří Vackář

"Automated determination of earthquake source parameters"

The presented thesis deals with determining of earthquake source parameters by automated method based on Bayesian probability. It tackles a weak point of any inverse problem, which is handling the uncertainties and projecting them from the data space to the parameter space. The existing most appropriate way of doing this is transforming the data covariance matrix (the prior) to the posterior probability density function of the model. Applying this approach is still rather rare in seismology, which makes the topic of the thesis actual and novel. Jiří Vackář does not limit his work to the inversion problem only; he also includes assessing the possible errors in data – velocity model and seismograms. This is reflected in the topics of the three involved papers. These address (i) Improving the velocity model with the use of observed fast long-period waves (Vackář et al. 2014); (ii) Automated detection of long-period disturbances in seismograms (Vackář et al. 2015) and finally (iii) Bayesian inversion for centroid moment tensors (Vackář et al. 2017). Instead of developing his own inversion method, the author has partially rewritten and significantly extended the existing and widespread used code ISOLA, which makes this extension available to a large community of researchers.

The thesis is of standard monograph-type, involves 120 pages and consists of nine chapters and one appendix.

The **Introduction** brings basic information about earthquakes and thorough review of centroid moment tensors and their uncertainty resulting from uncertainty and errors in data and wrong models. It also defines the aim of the thesis and its structure. I enjoyed reading this concise, comprehensive and easy to understand introductory part.

Chapter 1 introduces briefly **chapter 2,** which is a copy of Paper 1 on analysis of fast long-period surface waves, which were successfully simulated by the author in terms of leaking modes only after enhancing the available velocity model. Despite this chapter is not directly linked to the topic of the thesis it illustrates the necessity of high-quality velocity models for successful seismic source studies and documents the expertise of the author in structural seismology.

Chapter 3 is a part of Paper 3 and covers the Bayesian formulation of inverse problem of CMT**.** This part provides tools for inverting the data to CMT and derives a formula for the posterior probability density function of the parameters.

Chapter 4 reviews sources of uncertainties in data. I like classification of errors and uncertainties and also the thorough discussion of all their sources from seismograph parameters, signal disturbances and noise, uncertainties of the model and geometry of seismic network.

Chapter 5 is a copy of Paper 2 dealing with possibly the most serious disturbance in seismograms: the long-period pulses (Mouse) caused by acceleration step. I consider this paper an inherent part of the thesis because the provided code ISOLA-ObsPy code uses the MouseTrap code to detect these disturbances.

Chapter 6 is a part of Paper 3 and describes the construction of the data covariance matrix using a noise window preceding the seismic waves accounting also for the correlated noise between components. Additionally, a weighting scheme allowing to display waveform fit where the observed waveform is weighted by covariance matrix is developed.

Chapter 7 is devoted to description of the ISOLA-ObsPy code, its functionality and parametrization.

Chapter 8 describes synthetic test and application of the method to data of the Swiss Seismological Service.

The final **Chapter 9** gives an outlook of further developments and the thesis is summarized in **Conclusion**.

The thesis is easy to read and well organized, English language is also fine, I found only several minor grammatical mistakes. Below I summarize my impression and questions.

Positive points

- I like the application of noise covariance matrix as an implicit filter for noisy frequencies
- Object oriented programming is a great choice!
- Fig. 8.2: nice example how the high-noise stations bias the solution
- Fig. 8.7: nice illustration how weighting by covariance matrix acts as a frequency filter

Minor errors

- Fig 8.14: (c) low DC percentage should read high DC percentage

Questions

- In section 7.6.1 the maximum frequency is chosen based on the upper limit of number of wavelengths – this is a good choice; however, the noise covariance matrix could make the same job – did you test this option?
- In section 7.6.2 it is stated that
  "Then we determine common sampling rate Scomm as the greatest common divisor (GCD) from sampling rates of single components s1, s2… The decimation factor d, which is the quotient of these two numbers, is then rounded by the floor function to enable integer decimation factor. The used working sampling Sused is finally determined as a quotient of Scomm and d."
  This is a bit unclear, in particular
  (i) Does it mean that different components are sampled by different sampling rates?
  (ii) Why GCD is required - this means that integer resampling is possible only?
  To make it clear, please illustrate this approach with some example
- I like the way of displaying the waveform fit by weighting the waveforms by covariance matrix. However, in Figs 8.2, 8.6., 8.7. the unweighted seismogram amplitudes appear too small (only 1e-20 m/s). Possibly the reason is the dimensionless synthetic event?
- Covariance matrices in Figs 8.4 and 8.9 look different: the larger rectangles in Fig 8.4 (white noise) show non-zero inter-component covariance, which is missing in Fig 8.9 (colored noise). What is the reason?

- Description of scmtv implementation at SED (section 8.2): "The methodology assumes that the isotropic component is zero, the epicentral coordinates are fixed (though the depth can vary), and the source time function is fixed, so it is an MT rather than CMT approach." Possibly epicenter coordinates should read as Centroid coordinates? If this is right, does it makes sense to compare the distance to epicenter, if centroid in scmtv is fixed?
- Fig 8.12: In relation to my previous point it is not clear what is shown on X axis of (d)? The axis label reads "Hypocenter to centroid position" and caption tells "Difference in centroid position". At the same time "Difference in centroid depth" is announced in (c) caption.

In conclusion I find that the thesis presents original scientific results of direct impact to determining the source parameters of earthquakes and assessing the related uncertainties. Jiří Vackář focused on three interrelated topics, in which he had brought new approaches (development of Bayesian method for CMT inversion) or applied existing methodology (explaining the fast, long period waves) or introduced automated application of existing methodology (detection of LP disturbances). He has proven the ability for independent scientific research. I did not find any significant problems in the thesis. Accordingly, I fully recommend his thesis for defense.

Prof. RNDr. Tomáš Fischer, Ph.D.
Charles University
Faculty of Science