Journal of Rock Mechanics and Geotechnical Engineering

Journal online: www.rockgeotech.org

Source location error analysis and optimization methods

Maochen Ge*

Missouri University of Science and Technology, Rolla, MO 65409, USA

Received 9 November 2011; received in revised form 1 March 2012; accepted 12 March 2012

Abstract: The efficiency of an optimization method for acoustic emission/microseismic (AE/MS) source location is determined by the compatibility of its error definition with the errors contained in the input data. This compatibility can be examined in terms of the distribution of station residuals. For an ideal distribution, the input error is held at the station where it takes place as the station residual and the error is not permitted to spread to other stations. A comparison study of two optimization methods, namely the least squares method and the absolute value method, shows that the distribution with this character constrains the input errors and minimizes their impact, which explains the much more robust performance by the absolute value method in dealing with large and isolated input errors. When the errors in the input data are systematic and/or extreme in that the basic data structure is altered by these errors, none of the optimization methods are able to function. The only means to resolve this problem is the early detection and correction of these errors through a data screening process. An efficient data screening process is of primary importance for AE/MS source location. In addition to its critical role in dealing with those systematic and extreme errors, data screening creates a favorable environment for applying optimization methods.

Key words: source location; residual; error; least squares method; absolute value method; acoustic emission (AE); microseismic (MS)

1 Introduction

Errors in acoustic emission/microseismic (AE/MS) source location data are inevitable because of a variety of practical reasons. A common practice in dealing with these input errors is to incorporate an optimization scheme, such as the least squares method or the absolute value method, for source location. The perception is that the impact of these input errors will be automatically diminished or minimized as a result of the optimization process.

Although optimization methods are important and used widely for AE/MS source location, the perception that these methods can automatically handle the input errors is misleading from a theoretical point of view and can be very harmful from a practical point of view. The intention of this paper is to provide readers a perspective view of the optimization process and the general principles for handling source location errors.

The discussion is divided into three parts. The first part discusses how the source location errors are defined by the optimization methods and their relations with the input errors. It is to show that the difference among the optimization methods is how the error is defined. Therefore, an optimization method in essence is an error definition and its effectiveness depends on whether the error defined by the method is compatible with the errors contained in the input data.

The focus of the second part is the distribution of station residuals. This section explores how the errors in the input data might be mapped as station residuals and what is the ideal distribution of station residuals that would minimize the impact of the input errors. The efficiency of the least squares method and the absolute value method for the AE/MS source location purpose is then analyzed and compared in this regard.

In the third part, the application condition of the optimization methods is further examined in terms of the typical errors encountered in AE/MS source location. This section demonstrates that none of the optimization methods will be able to function if the errors in the input data are systematic or extreme in that it damages the basic data structure. The only means to resolve this problem is to detect these errors by data screening and to make the necessary corrections prior to the source location process.

Doi: 10.3724/SP.J.1235.2012.00001

*Corresponding author. Tel: +1-573-3416029; E-mail: gem@mst.edu

2 Basic concepts on source location residuals

Ideally, a source location error refers to the difference between the actual location and the calculated location. In reality, however, the actual locations of AE/MS events are generally not known. As a result, source location errors are mostly measured indirectly by some parameters related to the physical data utilized for the source location.

When the signal arrival times recorded from a sensor array are utilized for source location, the place which would best match the recorded arrival times is considered as the event location. The mismatched portions between the recorded arrival times and the calculated arrival times are termed as residues. Source location, from a mathematical point of view, is therefore a process of searching for the point with the minimum residual.

2.1 Event residual

It is known from the previous discussion that the residual used for defining an event location is a measurement of total error. In statistics, a total error is not an arbitrary term; it is defined precisely by a regression method. Two best known regression methods are the least squares method and the absolute value method. The least squares method, also known as the L2 norm approach, defines the total error as the sum of the squares of individual errors. In the case of AE/MS source location, it may be expressed by the following equation:

$$Res = \left[\left(\sum \gamma_i^2 \right) / (n - m) \right]^{1/2} \quad (1)$$

where *Res* is the event residual, γ_i is the residual associated with the *i*th station, *n* is the number of stations, and *m* is the degree of freedom. If multiple arrivals from the same station are used, for instance, when both P- and S-wave arrival times are used, γ_i can be interpreted as the residual associated with the *i*th arrival time and the number of arrival times, *n*.

The least squares method has long been used in science and engineering to obtain the so-called *best fit* for over-determined problems. Based on statistical considerations, the *best fit* is unbiased only for linear approximations with the assumption that the errors associated with each variable follow a normal probability distribution (Hines and Montgomery, 1980; Burden et al., 1981).

A difficulty with the least squares method for AE/MS source location is that the input errors do not often follow a normal distribution as assumed by the method. For instance, the number of stations used for source location is often limited. If a large error occurs at one station, the assumption of the normal distribution is violated.

The absolute value method, or L1 norm approach, defines the event residual as the sum of the absolute values of the individual errors:

$$Res = \left(\sum | \gamma_i | \right) / (n - m) \quad (2)$$

An important advantage of the absolute value method is that it is relatively insensitive to large errors. This method was introduced to the analysis of source locations in the 1980s (Anderson, 1982; Prugger and Gendzwill, 1989).

2.2 Station residual

By definition, a station residual is the difference between the observed arrival times and calculated arrival times which can be expressed by the following equation:

$$\gamma_i = t_i - tc_i \quad (3)$$

where t_i and tc_i represent the observed and calculated arrival times, respectively. The calculated arrival time consists of origin time, *t*, and travel time tt_i:

$$tc_i = t + tt_i \quad (4)$$

It can be shown that the origin time for the least squares method can be expressed by the following equation (Ge, 1995):

$$t = \sum (t_i / n) - \sum (tt_i / n) \quad (5)$$

and, therefore, a precise expression of the station residual for the least squares method is

$$\gamma_i = t_i - tt_i - \sum (t_i / n) + \sum (tt_i / n) \quad (6)$$

The origin time for the absolute value method is the median of all $(t_i - tt_i)$s and is denoted by t_m. The station residual for the absolute value method, therefore, is

$$\gamma_i = t_i - tt_i - t_m \quad (7)$$

It is understood from the above discussion that, although both the least squares method and the absolute value method share the same format of the station residual as defined in Eq. (3) and the same format of the calculated arrival time as defined in Eq. (4), the contents of these equations are different for the two methods because the origin time is uniquely defined by each regression method as shown in Eqs. (6) and (7) (Ge, 1995).

2.3 Error space and source location

It is known from Eqs. (6) and (7) that the data required for the residual calculation include observed arrival times, t_i , calculated travel times, tt_i , and number of travel times, *n*. As such, for a given set of