

# Application of a weighted likelihood method to hypocenter determination

M. Imoto

National Research Institute for Earth Science and Disaster Prevention, 3-1 Ten-nodai, Tsukuba, Ibaraki 305-0006, Japan

(Received June 25, 2012; Revised July 25, 2013; Accepted July 28, 2013; Online published December 6, 2013)

The method of least squares is a standard approach to hypocenter determination in seismology. However, this method is not useful for data contaminated by systematic errors. To address this problem, we propose a weighted likelihood method (WLL) rather than a weighted least-squares method (WLSQ). Assuming a normally distributed random error and systematic errors, both methods give the same solution; however, variances of random errors estimated by WLSQ are much smaller than those estimated by WLL. Examining reasonable random errors, we simulate a case of systematic errors varying linearly with a given parameter, where the number of unknown parameters is reduced to one for simplification. We assume that a systematic error, two different arrays of stations, and three different weights are functions of distance. In the cases where biases affected by systematic errors are adequately reduced, the variances of random errors estimated by WLL become roughly equal to that assumed, but those estimated by WLSQ are much smaller than that assumed. This result implies that WLL is a better approach than WLSQ for data contaminated by systematic errors.

**Key words:** Weighted likelihood method, weighted least squares, hypocenter determination, systematic error.

## 1. Introduction

The method of least squares is a standard approach to the approximate solution of overdetermined systems. Use of this method can reduce the effect of random errors. The best solution in the least-squares method minimizes the sum of squared residuals—a residual being the difference between an observed value (arrival time of a seismic wave) and the fitted value provided by a model (origin time and location of an earthquake and a seismic velocity model). Hypocenter determination employs the nonlinear least squares method and is implemented by iterative refinement. If it is assumed that random error variances largely vary among observation stations, the weighted least-squares method (WLSQ) can be used to determine the hypocenter more reliably.

The residual is caused by both errors in the measurement of the arrival time and errors in the seismic velocity model for calculating theoretical arrival times. The former is a random error, but the latter is a systematic one. For a nationwide network in Japan, such as those operated by the Japan Meteorological Agency (JMA) and the National Research Institute for Earth Science and Disaster Prevention (NIED), a simple velocity model for hypocenter determination often involves systematic errors in the calculated travel time, since seismic velocities vary from region to region. JMA and NIED currently adopt WLSQ, in which the weight of each observation depends on its hypocentral distance. Their procedure, which may be a better approach than a simple least-squares method, obviously violates the condition that the least-squares method must be applied to data without systematic errors. Therefore, this procedure may give unre-

liable solutions, which has not yet been addressed.

In this study, we presume both normally distributed random errors and systematic errors and propose the use of the weighted likelihood method (WLL) (Hu and Zidek, 2002; Wang and Zidek, 2005) to address this problem. First, we demonstrate that WLSQ and WLL provide the same solution; however, the variance of random errors estimated by WLL exceeds that estimated by WLSQ. In order to clarify which method estimates more reasonable errors, both methods are applied to data contaminated with systematic errors in order to simulate hypocenter determination with an unsuitable velocity model. The solutions with both methods could be analytically obtained for the simulated data. This paper compares the variance of random errors estimated by the maximum likelihood method with the assumed ones. This comparison indicates that WLSQ underestimates the variance of the random errors and is unreliable as an approach to this problem.

## 2. Method

### 2.1 Weighted least squares

The simple least-squares method assumes that the standard deviation of the random error is constant for all samples. When the standard deviation varies, WLSQ can be applied. Less weight is given to less precise measurements, and more weight is given to more precise measurements. Here,  $\sigma$  and  $\sigma_i$  denote the standard deviation of the random error in a representative observation and the  $i$ th observation ( $i = 1, \dots, n$ ), respectively. The weight  $w_i$  is inversely proportional to the variance of the error, and is represented by  $\sigma^2/\sigma_i^2$ . The likelihood function in this case is given as:

$$L = \prod_{i=1}^n \left[ \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp \left\{ -\frac{1}{2\sigma_i^2} (o_i - c_i)^2 \right\} \right]$$

$$= \prod_{i=1}^n \left[ \frac{1}{\sqrt{2\pi\sigma^2/w_i}} \exp \left\{ -\frac{w_i}{2\sigma^2} (o_i - c_i)^2 \right\} \right], \quad (1)$$

where  $o_i$  denotes the observed value and  $c_i$  denotes the theoretical value at the  $i$ -th observation.

The maximum log-likelihood is obtained as:

$$\ln L_{\max} = -\frac{n}{2} \ln(2\pi\sigma^2) + \frac{\sum_{i=1}^n \ln w_i}{2} - \frac{\hat{R}^2}{2\sigma^2}, \quad (2)$$

where

$$R^2 = \sum_{i=1}^n \{w_i \cdot (o_i - c_i)^2\}, \quad (3)$$

where  $\hat{R}^2$  is the minimum of  $R^2$ .

The maximum likelihood estimate of  $\sigma^2$ ,  $\hat{\sigma}_{\text{WLSQ}}^2$ , is given by:

$$\frac{\partial \log_e L_{\max}}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}_{\text{WLSQ}}^2} + \frac{\hat{R}^2}{2\hat{\sigma}_{\text{WLSQ}}^4} = 0. \quad (4)$$

Then

$$\hat{\sigma}_{\text{WLSQ}}^2 = \frac{\hat{R}^2}{n}, \quad (5)$$

where the subscript WLSQ refers to the weighted least-squares method.

## 2.2 Weighted likelihood method

If reliabilities of observed samples differ from sample to sample, each sample can contribute to the maximum likelihood estimate in proportion to its reliability through WLL. The WLL function is given in the form of the weighted sum of log-likelihoods for each sample:

$$\ln L = \sum_{i=1}^n (w_i \cdot \ln L_i), \quad (6)$$

$$w_i \ln L_i = \ln \left[ \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^{w_i} \exp \left\{ -\frac{w_i}{2\sigma^2} (o_i - c_i)^2 \right\} \right]. \quad (7)$$

The maximum log-likelihood is obtained as:

$$\ln L_{\max} = -\frac{1}{2} \sum_{i=1}^n \{w_i \ln(2\pi\sigma^2)\} - \frac{\hat{R}^2}{2\sigma^2}. \quad (8)$$

The maximum likelihood estimate of  $\sigma^2$ ,  $\hat{\sigma}_{\text{WLL}}^2$ , is given by

$$\frac{\partial \log_e L_{\max}}{\partial \sigma^2} = -\frac{\sum_{i=1}^n w_i}{2\hat{\sigma}_{\text{WLL}}^2} + \frac{\hat{R}^2}{2\hat{\sigma}_{\text{WLL}}^4} = 0. \quad (9)$$

Then

$$\hat{\sigma}_{\text{WLL}}^2 = \frac{\hat{R}^2}{\sum_{i=1}^n w_i}, \quad (10)$$

where the subscript WLL refers to the weighted likelihood method.

When the same weights are applied in WLL and in WLSQ, the same solutions of  $c_i$  are obtained. However, the maximum likelihood estimates of  $\sigma^2$  are different.

## 3. Simulation

In applying WLL to hypocenter determination, the only difference from WLSQ is related to the formula for  $\sigma^2$ , where Eq. (10) can be compared with Eq. (5). When NIED and JMA determine the hypocenter, a weight with epicentral distance may be introduced to reduce bias due to systematic errors in the calculated travel times. The structure of seismic wave velocities in Japan cannot be modeled with a simple layered structure, as employed by NIED and JMA. Therefore, using arrival times of more distant stations in the determination results in systematic errors, and bias of the determined hypocenter. In this section, we examine which of WLSQ or WLL is better for hypocenter determination affected by systematic errors from an inappropriate velocity structure.

We schematically simulate this problem assuming simple conditions. We assume that an earthquake occurs on the surface of a uniform medium and that stations are densely deployed in a line from the epicenter in the first case, and on a two-dimensional surface in the second case. A systematic error due to an inappropriate velocity structure is as large as a random error at a short distance, and becomes much larger at more distant points. It is assumed that the epicenter of the earthquake is rather well constrained but the origin time of the earthquake is not, since the dense stations are well distributed. Thus, the problem of hypocenter determination is reduced to the problem of determining the origin time of the event.

The observed arrival time at the  $i$ th ( $i = 1, 2, \dots, n$ ) station is

$$t_i = t_0 + T(r_i) + B(r_i) + \varepsilon_i, \quad (11)$$

where  $r_i$  is the epicentral distance,  $T(r_i)$  is the travel time estimated by the model, and  $B(r_i)$  is a systematic error due to an inappropriate velocity model. In the single-layer model, this correction must be a function of epicentral distance, seismic wave velocity in the model, and the actual seismic wave velocity:

$$B(r_i) = \left( \frac{1}{V_c} - \frac{1}{V_m} \right) \times r_i, \quad (12)$$

where  $V_c$  is the actual value of the seismic wave velocity, and  $V_m$  is the seismic wave velocity used in the model. The term on the right-hand side is a linear function of epicentral distance:

$$B(r_i) = \alpha \cdot r_i. \quad (13)$$

Equation (11) can be represented as:

$$t_i = t_0 + T(r_i) + \alpha \cdot r_i + \varepsilon_i. \quad (14)$$

As the epicenter is well constrained under our assumption, the epicentral distance is considered to be a known parameter. Therefore, subtracting  $T(r_i)$  from the arrival time, we consider  $t_0$  to be an unknown parameter, and Eq. (14) can be replaced by:

$$o_i = t_i - T(r_i) = t_0 + \alpha \cdot r_i + \varepsilon_i. \quad (15)$$

Here, the problem of hypocenter determination is reduced to the problem of estimating  $t_0$  from observation  $o_i$  contaminated by a systematic error and a random error.

Without loss of generality, we fix the true value of the origin time to be 0. The sum of the residual squares  $R^2$  is calculated as:

$$R^2 = \sum_{i=1}^n w_i (B(r_i) + \varepsilon_i - x)^2, \tag{16}$$

where  $x$  refers to the estimated origin time. The maximum likelihood estimate of the origin time,  $\hat{x}$ , minimizes  $R^2$  and is calculated as follows:

$$\hat{x} = \frac{\sum_{i=1}^n w_i (B(r_i) + \varepsilon_i)}{\sum_{i=1}^n w_i}. \tag{17}$$

The minimum of  $R^2$  is calculated as:

$$\hat{R}^2 = \sum_{i=1}^n w_i (B(r_i) + \varepsilon_i)^2 - \sum_{i=1}^n w_i \hat{x}^2. \tag{18}$$

In the present study, we can estimate the expected values of Eqs. (17) and (18) with a probability density function for the epicentral distance of stations,  $S(r)$ , and the assumption that the random error of  $\varepsilon_i$  is normally distributed with a mean 0 and variance  $\delta^2$ . A weight function  $W(r)$  is applied and different limits of epicentral distance,  $u$ , are considered.

In order to derive an analytical solution of Eq. (17), we assume a two-step random value generation.

Step 1. Random generation of  $n$  stations within a distance of  $u$ .

Step 2. For each set of stations, generate a large number ( $l$ ) of series of random errors, ( $\varepsilon_{ij}$ : where  $j$  refers to the number of trials,  $j = 1, 2, \dots, l$ ).

For the  $j$ th trial of  $\varepsilon_{ij}$ , we obtain the maximum likelihood estimate of the origin time  $\hat{x}_j$  as follows:

$$\hat{x}_j = \frac{\sum_{i=1}^n w_i (B(r_i) + \varepsilon_{ij})}{\sum_{i=1}^n w_i}. \tag{19}$$

The average value of  $\hat{x}_j$  over  $j$ ,  $\overline{\hat{x}_j}$ , is estimated by:

$$\begin{aligned} \overline{\hat{x}_j} &= \frac{1}{l} \frac{l \cdot \sum_{i=1}^n w_i B(r_i) + \sum_{j=1}^l \sum_{i=1}^n w_i \varepsilon_{ij}}{\sum_{i=1}^n w_i} \\ &= \frac{1}{l} \frac{l \cdot \sum_{i=1}^n w_i B(r_i) + \sum_{i=1}^n \left\{ w_i \left( \sum_{j=1}^l \varepsilon_{ij} \right) \right\}}{\sum_{i=1}^n w_i}, \end{aligned} \tag{20}$$

where  $\varepsilon_{ij}$  refers to the random error at the  $i$ th station in the  $j$ th trial. For a large number of  $l$ ,  $\overline{\hat{x}_j}$  approaches  $E(\hat{x})$  as:

$$E(\hat{x}) = \frac{\sum_{i=1}^n w_i B(r_i)}{\sum_{i=1}^n w_i}, \tag{21}$$

where  $E$  denotes the expected value of a variable, since, for large enough values of  $l$ ,

$$\frac{1}{l} \sum_{j=1}^l \varepsilon_{ij} \approx 0, \tag{22}$$

which is an average of independent random errors  $\varepsilon_{ij}$  normally distributed with a mean 0. Thus, for each selection of stations, the expected value of Eq. (17) is given by Eq. (21).

For a set of large  $n$ , Eq. (21) can be represented as:

$$E(\hat{x}) = \frac{A}{C}, \tag{23}$$

where:

$$A = m \cdot \int_0^u B(r)S(r)W(r)dr, \tag{24}$$

$$C = m \cdot \int_0^u S(r)W(r)dr, \tag{25}$$

$$n = m \cdot \int_0^u S(r)dr, \tag{26}$$

and

$$\int_0^{10} S(r)dr = 1. \tag{27}$$

Here, we assume that  $m$  stations are distributed within an epicentral distance of 10 units.

Similarly, the expected value of Eq. (18) is derived as follows. For the  $j$ th trial of a set of stations:

$$\hat{R}_j^2 = \sum_{i=1}^n w_i (B(r_i) + \varepsilon_{ij} - \hat{x}_j)^2. \tag{28}$$

For a set of large  $n$ ,  $\hat{x}_j$  is replaced with  $E(\hat{x})$  of Eq. (21). This replacement will be justified by a Monte Carlo simulation. We obtain the average value of Eq. (28) over  $j$  for a large number of  $l$ :

$$\begin{aligned} \overline{\hat{R}_j^2} &= \frac{1}{l} \sum_{j=1}^l \sum_{i=1}^n w_i \left( B(r_i) + \varepsilon_{ij} - \frac{\sum_{k=1}^n w_k B(r_k)}{\sum_{i=1}^n w_i} \right)^2 \\ &= \sum_{i=1}^n w_i (B(r_i)^2 + \delta^2) - \left( \sum_{i=1}^n w_i \right) \cdot \left( \frac{\sum_{k=1}^n w_k B(r_k)}{\sum_{k=1}^n w_k} \right)^2. \end{aligned} \tag{29}$$

Equation (29) can be represented by:

$$E(\hat{R}^2) = D + \delta^2 \cdot C - \frac{A^2}{C}, \tag{30}$$

where

$$D = m \cdot \int_0^u B(r)^2 S(r)W(r)dr. \tag{31}$$

We analytically estimated the expected values of  $\hat{x}$ ,  $\hat{\sigma}_{WLL}$  and  $\hat{\sigma}_{WLSQ}$  for a large number of  $n$  in this way. Hereafter,  $\hat{x}$ ,  $\hat{\sigma}_{WLL}$  and  $\hat{\sigma}_{WLSQ}$  denote the expected values of  $\hat{x}$ ,  $\hat{\sigma}_{WLL}$  and  $\hat{\sigma}_{WLSQ}$ , respectively.

We consider two different observation station arrays and three different weight functions. In one case of the arrays, these stations are distributed in a line (Array1). The density function of  $S(r)$  is given as:

$$S(r) = 0.1. \tag{32}$$

Table 1. Formulas for estimating the values of  $A$ ,  $C$ , and  $D$ . Two different arrays of stations are listed in different columns. Three different weight functions are listed in rows.

	$S(r) = 0.1, m = 100$ $B(r) = \alpha \cdot r$	$S(r) = 0.02r, m = 100$ $B(r) = \alpha \cdot r$
$W(r) = 1.$	$A = 5\alpha u^2$ $C = 10u$ $D = \frac{10\alpha^2}{3}u^3$	$A = \frac{2}{3}\alpha u^3$ $C = u^2$ $D = \frac{\alpha^2}{2}u^4$
$W(r) = \begin{cases} 1 & 0 < r \leq 1 \\ \frac{1}{r^2} & 1 < r \leq 10. \end{cases}$	$A = 5\alpha + 10\alpha \cdot \log u$ $C = 20 - \frac{10}{u}$ $D = 10\alpha^2 \left(u - \frac{2}{3}\right)$	$A = 2\alpha \left(u - \frac{2}{3}\right)$ $C = 1 + 2 \log u$ $D = \alpha^2 \left(u^2 - \frac{1}{2}\right)$
$W(r) = \begin{cases} 1 & 0 < r \leq 1 \\ \frac{1}{r^4} & 1 < r \leq 10. \end{cases}$	$A = 10\alpha - \frac{5\alpha}{u^2}$ $C = \frac{10}{3} \left(4 - \frac{1}{u^3}\right)$ $D = 10\alpha^2 \left(\frac{4}{3} - \frac{1}{u}\right)$	$A = 2\alpha \left(\frac{4}{3} - \frac{1}{u}\right)$ $C = 2 - \frac{1}{u^2}$ $D = \frac{\alpha^2}{2} + 2\alpha^2 \log u$

In the other case, we consider stations deployed uniformly in a plane (Array2). The density function of  $S(r)$  is given as

$$S(r) = 0.02r. \tag{33}$$

Three different weighting functions are considered below. For the first case of applying a constant weight, the function of  $W(r)$  is given as:

$$W(r) = 1. \tag{34}$$

For the second case, where the weight decays inversely as the second power of distance, the weight is given as:

$$W(r) = \begin{cases} 1 & 0 < r \leq 1 \\ \frac{1}{r^2} & 1 < r \leq 10. \end{cases} \tag{35}$$

The assumed systematic error,  $B(r)$ , increases linearly with epicentral distance. WLSQ likely misapplies a weighting function that decays as the inverse of the second order of the epicentral distance, since the square of the systematic error increases in proportion to the second power of distance.

The last case, where the weight function decays as the inverse of the fourth power of epicentral distance, can be considered as an example of weight decaying faster than in the previous cases:

$$W(r) = \begin{cases} 1 & 0 < r \leq 1 \\ \frac{1}{r^4} & 1 < r \leq 10. \end{cases} \tag{36}$$

Table 1 lists formulas estimating the values of  $A$ ,  $C$ , and  $D$  for each of the six cases. Substituting these values into Eqs. (23), (30), and (5) or (10), we obtain standard deviations of random errors,  $\hat{\sigma}_{\text{WLSQ}}$  for WLSQ and  $\hat{\sigma}_{\text{WLL}}$  for WLL.

Figures 1(a) through (f) illustrate  $\hat{x}$ ,  $\hat{\sigma}_{\text{WLSQ}}$  and  $\hat{\sigma}_{\text{WLL}}$  with the upper limit  $u$  varying from 1 to 10 units, where  $\alpha = 1$  and  $\delta = 1$ . The systematic error is assumed to be equal to, or less than, the random error in a range of 0 to 1 units, where the weight is fixed to 1.  $\hat{\sigma}_{\text{WLSQ}}$  (dotted line) and  $\hat{\sigma}_{\text{WLL}}$  (dashed line) are compared with the assumed random error ( $\delta = 1$ ). In each set,  $\hat{x}$  represents the bias affected by the systematic error, since the correct origin time is set to 0.

In Figs. 1(a), 1(b), 1(d) and 1(e), biases caused by systematic errors are not adequately reduced, where  $\hat{x}$  departs

from 0. In these cases,  $\hat{\sigma}_{\text{WLL}}$  exceeds the standard deviation of the assumed random error. In both station arrays, it is clear that the most quickly decaying weight function (Figs. 1(c) and 1(f)), gives the best solutions, where  $\hat{x}$  becomes at most 1 even in cases including stations up to 10 units away. This suggests that biases are adequately reduced. In these cases,  $\hat{\sigma}_{\text{WLL}}$  becomes reasonable at around 1, mostly equal to the standard deviation of the assumed random error. In contrast,  $\hat{\sigma}_{\text{WLSQ}}$  values are much smaller than that assumed. Therefore, WLSQ underestimates the standard deviation of random errors, suggesting an invalid application of WLSQ to the present issue.

Table 2 summarizes  $\hat{x}$ ,  $\hat{\sigma}_{\text{WLSQ}}$  and  $\hat{\sigma}_{\text{WLL}}$  for  $u = 10$ . Results obtained from the weight function decaying as the inverse of the seventh power are added as a more quickly decaying weight function, which is the case of NIED. In this case,  $\hat{x}$  and  $\hat{\sigma}_{\text{WLL}}$  become slightly better than those of other cases, where  $\hat{x}$  becomes smaller and  $\hat{\sigma}_{\text{WLL}}$  approaches 1. These examples suggest that if  $\hat{\sigma}_{\text{WLL}}$  exceeds the standard deviation of the assumed random error, biases caused by systematic errors would not be adequately reduced.

Bias  $\hat{x}$  could not be compared with its standard error, the derivation of which is based on the Hessian/Fisher Information matrix of the log-likelihood function (Sakamoto *et al.*, 1983), since the standard error of  $\hat{x}$  concerns only random errors, while the bias originates from systematic errors. From a quantitative viewpoint, it should be noted that, in our definition,  $\hat{x}$ ,  $\hat{\sigma}_{\text{WLL}}$  and  $\hat{\sigma}_{\text{WLSQ}}$  are functions of  $u$  and are independent of  $m$ , but the standard error of  $\hat{x}$ ,  $\hat{\sigma}$  ( $\hat{\sigma}_{\text{WLL}}$  or  $\hat{\sigma}_{\text{WLSQ}}$ ) divided by the square root of  $C$ , is inversely proportional to the square root of  $m$ , which can easily be verified from Eqs. (5), (10), (23), (24), (25), (26), (30) and (31). The ratio of  $\hat{x}$  to its standard error becomes large in proportion to the square root of  $m$ . Consequently, bias  $\hat{x}$  cannot be compared with its standard error.

#### 4. Discussion and Summary

A Monte Carlo simulation was conducted to confirm Eq. (23) and justify the replacement in Eq. (29). Figures 2(a), 2(b) and 2(c) illustrate  $\hat{x}$ ,  $\hat{\sigma}_{\text{WLSQ}}$  and  $\hat{\sigma}_{\text{WLL}}$  values obtained for the set of Fig. 1(f). For each distance limit, the number of stations can be calculated by  $S(r)$ , and their locations are randomly generated. Random errors are also generated once for each station, which implies  $l = 1$  in

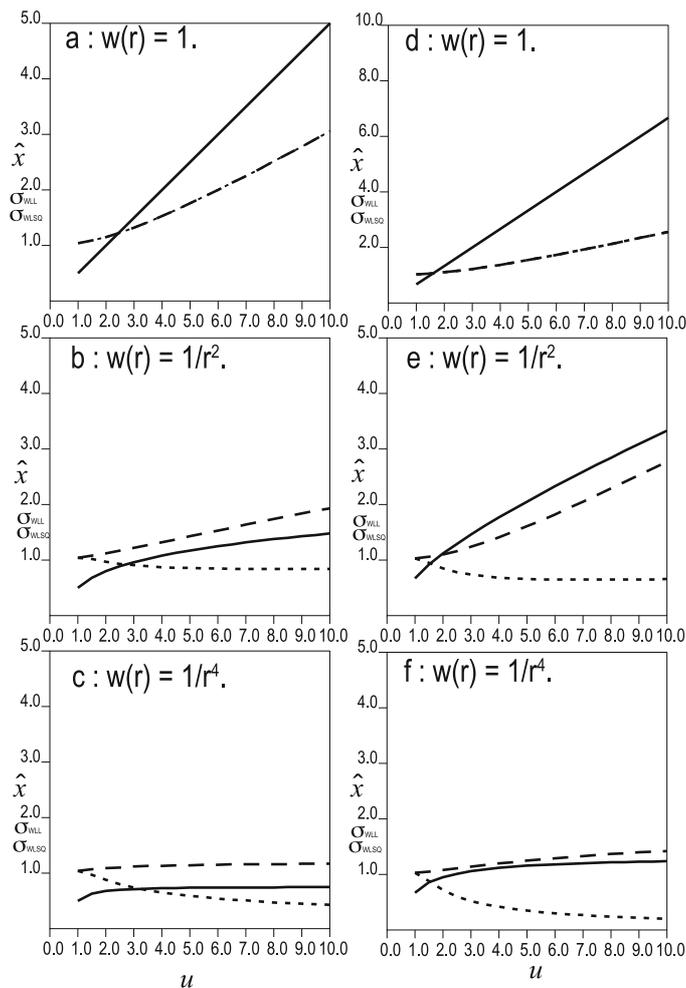


Fig. 1. Variation of  $\hat{x}$ ,  $\hat{\sigma}_{WLL}$  and  $\hat{\sigma}_{WLSQ}$ . The time scale (ordinate) is measured in a unit of  $\delta$ . The solid, dashed and dotted curves denote  $\hat{x}$ ,  $\hat{\sigma}_{WLL}$  and  $\hat{\sigma}_{WLSQ}$ , respectively. Figures 1(a), 1(b), 1(c). Different weight functions with Array 1. Figures 1(d), 1(e), 1(f). Different weight functions with Array 2. In Figs. 1(a) and 1(d), the dotted and dashed curves completely overlap each other.

Table 2. List of  $\hat{x}$ ,  $\hat{\sigma}_{WLSQ}$  and  $\hat{\sigma}_{WLL}$ . The listed values are estimated by stations at distances of up to 10 units. Results from the 7th power are also listed.

$S(r)$	Power of $w(r)$	$\hat{x}$	$\hat{\sigma}_{WLSQ}$	$\hat{\sigma}_{WLL}$
Array1	0	5.00	3.06	3.06
	-2	1.48	0.84	1.93
	-4	0.75	0.43	1.17
	-7	0.60	0.36	1.07
Array2	0	6.67	2.56	2.56
	-2	3.33	0.66	2.77
	-4	1.24	0.20	1.42
	-7	0.83	0.13	1.07

Eqs. (20) and (29). Ten thousand sets of stations are generated for each distance limit, which is shifted from 1 to 10 units for every unit step. In each set of the figure, the simple averages over ten thousand sets for  $m = 200$  (dashed line) and  $m = 500$  (dotted line) are compared with analytical solutions (solid line).

In general, as the number of stations increases (fixed distance limit), the results obtained by the simulation approach those of the analytical solution. Qualitative relations among  $\hat{x}$ ,  $\hat{\sigma}_{WLL}$  and  $\hat{\sigma}_{WLSQ}$  obtained from the Monte Carlo simulation do not largely differ from those of the analytical solutions, except in the case of a few tens of stations. Although

not shown here, the agreement between the simulation and the analytical solution in other cases (the other station array and/or different weights) becomes better than that of Fig. 2. We have only discussed qualitative relations between the parameters, so an analytical solution is an appropriate approach to this issue.

Ueno *et al.* (2002) reported that hypocenters of larger earthquakes tend to be deeper than those of smaller ones. They speculated that this tendency is related to the ratio of the number of long-distance stations to that of short-distance stations. This implies that a bias of hypocenters can be caused by systematic errors depending on the epi-

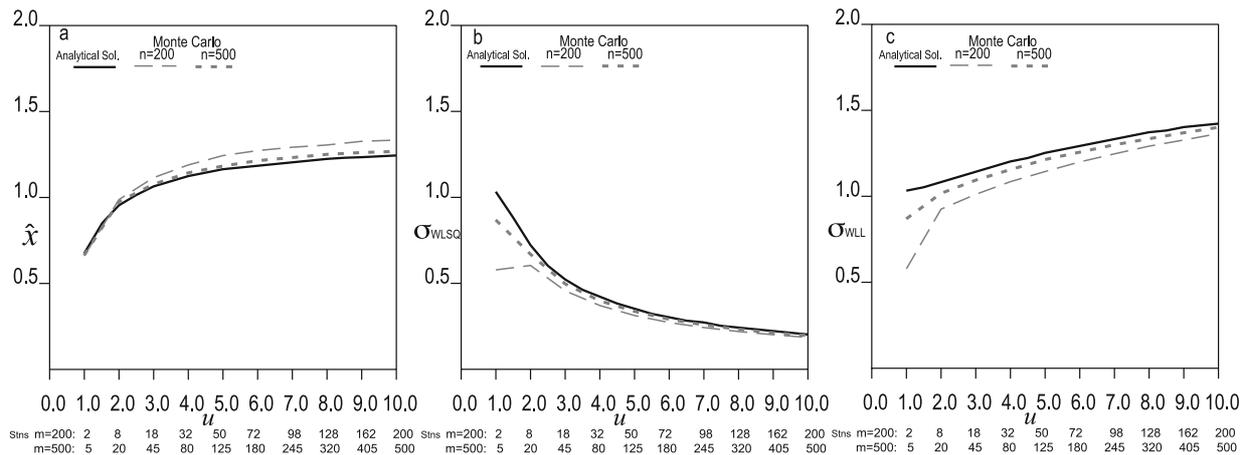


Fig. 2. Results obtained by the Monte Carlo simulation are compared with the analytical solution for the set of Fig. 1(f). Figures 2(a), 2(b) and 2(c) illustrate  $\hat{x}$ ,  $\hat{\sigma}_{\text{WLSQ}}$  and  $\hat{\sigma}_{\text{WLL}}$  values, respectively. In each set, the simple averages over ten thousand sets for  $m = 200$  (dashed curve) and  $m = 500$  (dotted curve) are compared with the analytical solution (solid curve). The number of stations within each distance limit is indicated.

central/hypocentral distance, which originate from the difference in velocity structure between the model used in hypocenter determination and the actual case. In order to reduce this bias, NIED (Shiomi, 2009 (personal communication)) introduced a weight depending on the hypocentral distance,  $Y$ :

$$W(Y) = Y_{\min}^7 / Y^7,$$

where  $Y_{\min}$  is the hypocentral distance of the station closest to the hypocenter (km) (if  $Y_{\min} \leq 50$ , then  $Y_{\min} = 50$ ; if  $W(Y) > 1$ , then  $W(Y) = 1$ ).

However, JMA introduced a similar weighting depending on the hypocentral distance except for the order of power:

$$\text{For } P\text{-waves } (W_p): W_p = Y_{\min}^2 / Y^2$$

$$\text{For } S\text{-waves } (W_s): W_s = W_p / 3.$$

Comparing weights in these methods with those of our simulation indicates that NIED, using a weight function decaying faster than the inverse fourth power, probably succeeds in reducing the bias by systematic errors (Table 2). But, JMA may fail to reduce bias with a weight function decaying inversely with the second power of distance. Both NIED and JMA must underestimate standard errors of hypocenters, which are calculated based on  $\hat{\sigma}_{\text{WLSQ}}$ . Standard errors of hypocenters are often critical in considering reliable hypocenter distributions of clustering earthquakes and their tectonic implications.

For deeper earthquakes, a qualitative consideration could be given as follows. At short epicentral distances compared with the focal depth, systematic errors reach a level depending on the focal depth. On the other hand, at longer distances, systematic errors increase with distance mostly similar to those of shallow earthquakes. These result in smaller variances of residuals than those of shallower earthquakes, since a range of systematic errors becomes smaller than those of shallower earthquakes. This implies that a significant bias may remain, even if  $\hat{\sigma}_{\text{WLL}}^2$  approaches the variance of random error. Such aspects should be discussed further as separate studies, since station arrays, the number of unknown parameters, and other factors must be rearranged to address these issues.

The present paper focuses on only the preliminary formu-

lation of applying WLL to hypocenter determinations. In practical applications of WLL, an optimal weight function should be determined taking into consideration the results in the case of various kinds of earthquakes in different areas and depths (Hu and Zidek, 2002). WLL can be applied to hypocenter determination with a minor revision of current methods which are based on misapplications of WLSQ.

We propose the use of WLL rather than WLSQ for hypocenter determination. Both methods give the same solution; however, the variance estimated by WLSQ is much smaller than that estimated by WLL. Our simulation indicates that a weight function that decays faster with distance gives a better solution, which could be realized by WLL. In contrast, such flexible weight functions are not justified by WLSQ, since weights should be inversely proportional to the variances of errors. In the cases where biases affected by systematic errors are adequately reduced, the variances of random errors estimated by WLL become roughly equal to the given one, but those estimated by WLSQ are much smaller than the given one. Therefore, WLSQ should not be used to address systematic errors in hypocenter determination. We conclude that WLL is a better approach than WLSQ for data contaminated by systematic errors.

**Acknowledgments.** The author thanks two anonymous reviewers for their critical reading and comments on this manuscript.

## References

- Hu, F. and J. V. Zidek, The weighted likelihood, *Can. J. Stat.*, **30**, 347–371, 2002.
- Japan Meteorological Agency, Users' Guide, [http://data.sokki.jmbasc.or.jp/cdrom/seismological/catalog/notes\\_e.htm](http://data.sokki.jmbasc.or.jp/cdrom/seismological/catalog/notes_e.htm) (as of April 1, 2013).
- Sakamoto, Y., M. Ishiguro, and G. Kitagawa, *Akaike Information Criterion Statistics*, 290 pp, D. Reidel, Dordrecht, 1983.
- Ueno, H., S. Hatakeyama, J. Funakaki, and N. Hamada, Improvement of hypocenter determination procedures in the Japan Meteorological Agency, *Quart. J. Seismol.*, **65**, 123–131, 2002 (in Japanese).
- Wang, X. and J. V. Zidek, Derivation of mixture distributions and weighted likelihood function as minimizers of KL-divergence subject to constraints, *Ann. Inst. Statist. Math.*, **57**, 687–701, 2005.