

## 지자기 전달함수의 로버스트 추정

양준모 · 오석훈 · 이덕기 · 윤용훈  
기상연구소 해양기상지진 연구실

### 요 약

일반적으로 지자기 전달함수는 관측치와 예측치의 차이를 최소화하는 관점에서 해가 추정된다. 오차의 구조가 가우스 분포를 따르면 최소자승 추정이 최적의 추정이지만, 그렇지 않은 경우 전달 함수 추정을 심각하게 왜곡시킬 수 있으므로 오차 구조에 대한 정보가 요구된다. 본 연구에서는 Q-Q plot을 이용한 오차 구조의 검증을 통하여 실제 오차 구조에 대한 정보를 획득하였고 가우스 분포 가정을 벗어나는 오차 구조에 대해 외치(outlier)에 의한 영향을 최소로 하며 해를 추정하는 로버스트 추정(regression M-estimate)을 적용하였다. 오차가 가우스 분포를 따르는 경우, 최소자승 추정과 로버스트 추정은 유사한 결과를 나타내나, 오차가 가우스 분포를 벗어나는 경우 로버스트 추정이 최소자승 추정보다 부드러운 결과를 나타냄을 확인하였다.

**주요어:** 지자기 전달함수, 로버스트 추정, Q-Q(Quantile-quantile) plot

**Jun-Mo Yang, Seok-Hoon Oh, Duk-Kee Lee and Yong-Hoon Youn, 2002, Robust estimation of geomagnetic transfer function. Journal of the Korean Geophysical Society, v. 5, n. 2, p. 131-142**

**ABSTRACT:** Geomagnetic transfer function is generally estimated by choosing transfer function to minimize the square sum of differences between observed and predicted values. If the error structure accords to the Gaussian distribution, standard least square(LS) can be the best estimation. However, for non-Gaussian error distribution, the LS estimation can be severely biased and distorted. In this paper, the Gaussian error assumption was tested by Q-Q(Quantile-Quantile) plot which provided information of real error structure. Therefore, robust estimation such as regression M-estimate that does not allow a few bad points to dominate the estimate was applied for error structure with non-Gaussian distribution. The results indicate that the performance of robust estimation is similar to the one of LS estimation for Gaussian error distribution, whereas the robust estimation yields more reliable and smooth transfer function estimates than the standard LS for non-Gaussian error distribution.

**Key words:** geomagnetic transfer function, robust estimation, Q-Q(Quantile-quantile) plot

*(Jun-Mo Yang, Seok-Hoon Oh, Duk-Kee Lee and Yong-Hoon Youn, Marine Meteorology & Earthquake Res. Lab., Meteorological Research Institute, 460-18, Shindaebang-Dong, Dongjak-Gu, Seoul, 156-720, Korea. email: yjm@metri.re.kr)*

### 1. 서 론

GDS(Geomagnetic Depth Sounding)는 자연적으로 존재하는 지구 자기장의 변화를 송신원으로 이용하여 지각구조를 규명하는 극심부 전자탐사 기술이다. GDS는 통상 서로 직교하는 지자기 3성분을 측정하여 지하의 전기적 구조에 대한 정성적인 정보를 제공하는 탐사 기법으로서 관측된 지자기 3성분은 시간 영역에서 주파수 영역으로

변환되며, 주파수 영역에서 세 성분간의 선형관계를 가정하여 지하의 전기적 구조를 반영하는 지자기 전달함수(geomagnetic transfer function)를 추정한다(Rikitake, 1985).

지자기 전달함수 추정은 오차가 가우스 분포를 따른다는 가정하에 오차의 제곱을 최소화하는 최소자승의 의미로서 해를 찾는 과정이다. 그러나 균질한 전자유도의 가정이 부적절한 자기 폭풍시기, 신호가 잡음에 독립적이라는 가정이 맞지 않

는 경우, 다양한 환경적 잡음이 존재하는 경우에는 오차의 가우스 분포 가정은 비현실적이다(Egbert · Booker, 1986). 자료가 가우스 분포를 따르지 않는 오차 구조에 대해서는 많은 연구가 이루어졌으며 (Huber, 1981; Egbert · Booker, 1986; Chave *et al.*, 1987; Egbert · Livelybrooks, 1996), 이러한 오차 구조에서 최소자승 추정은 몇 개의 잡음으로 오염된 자료에 대해 민감하므로 부정확한 해를 도출할 수 있다.

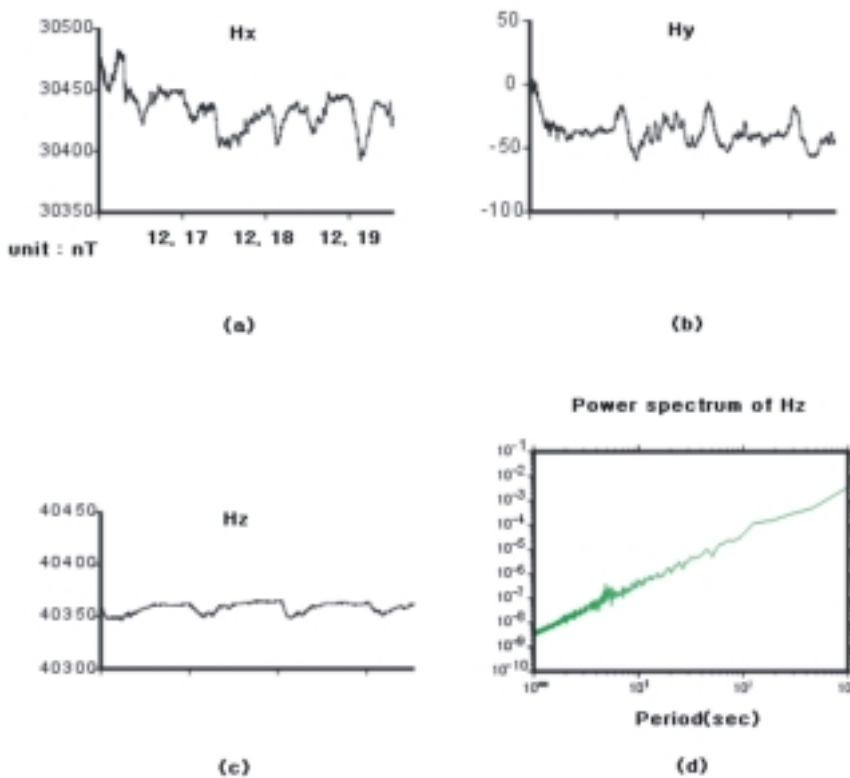
본 논문에서는 우선 지자기 자료의 처리 방법에 대해 간략히 소개하고, 최소자승의 의미에서 전달함수를 추정하는 방법에 대해 설명한다. 그 후 경기도 이천과 용인에 있는 지자기 고정관측소에서 획득한 지자기 자료를 이용하여 오차 구조의 가우스 분포 가정이 검증되는 Q-Q plot에 대한 해석 방법과 결과를 살펴보고 오차 구조에

따른 최소자승 추정과 로버스트 추정의 효율성에 대해 토의하고자 한다.

## 2. 자료 획득 및 처리

### 2.1. 자료 획득

본 연구에서는 전파 연구소가 이천과 용인에서 운영하는 지자기 고정관측소에서 플럭스 게이트 (Flux-gate)를 사용하여 측정된 지자기 3성분 자료를 이용하였다. 자료의 총 관측 기간은 2001년 12월 16일 - 12월 31일 까지 16일이며, 측정 간격은 5초, 측정 정밀도 0.01 nT이다. 또한 자료의 비교를 위하여 지자기 전달함수 추정을 위해 사용된 지자기 자료는 용인과 이천에서 동일한 시간대에 측정된 자료를 사용하였다. Fig. 1은 이천에서 획득한 약 3일 정도의 지자기 3성분 자료인



**Fig. 1.** The three-components of geomagnetic field that observed at Icheon and the power spectrum for the time variations of the magnetic field. (a), (b) and (c) denote northern, eastern and vertical magnetic field components, respectively. (d) represents the power spectrum of the time variations of the vertical magnetic field and its unit is  $\sqrt{nT^2/Hz}$ .

데 (a), (b), (c)는 각각 북쪽 방향, 동쪽 방향, 수직 방향의 지자기 성분이고 (d)는 수직 성분 (c)를 푸리에 변환하여 얻은 파워 스펙트럼의 간단한 예이다.

**2.2. 전 처리**

시간 영역에서 관측된 지자기 자료를 주파수 영역에서 처리하기 위해서는 푸리에 변환을 수행하여야 한다. 그러나 안정적이고 신뢰성 있는 추정을 위해서 주파수 영역으로 변환하기 전 단계로 전처리 과정(pre-processing)이 수행되어야 한다. 본 연구에서는 전 처리 과정으로서 스파이크 제거, 평균값 제거(demeaning), 테이퍼링(tapering), 저주파 차단 필터링(low-cut filtering), 고주파 차단 필터링(high-cut filtering)을 수행하였다.

첫 번째 과정은 스파이크와 평균값 제거이다. 스파이크는 매우 심한 불연속점이므로 자료에서 반드시 제거되어야 하며 평균값 제거는 계산상의 편의가 목적이다. 두 번째 과정은 테이퍼링으로서 푸리에 변환 시 자료 양단의 불일치가 불연속점으로 나타나게되므로 이 효과를 최소화하기 위해 적

절한 윈도우를 도입하여 푸리에 변환 전에 자료의 양단을 평균값(일반적으로 0)에 접근시키는 과정이다. 본 연구에서는 장형 윈도우(prolate window)의 일종인 Kaiser-Bessel 윈도우를 선택하여 테이퍼링과 평활화(smoothing)를 수행하였다. Kaiser Bessel 윈도우의 반응 파형과 주파수 반응은 alpha값이 커짐에 따라 반응 파형의 넓이는 좁아지고 주파수 반응은 주 로브(main lobe)와 부 로브(side lobe)의 스펙트럼 에너지 차이가 커지므로 적절한 alpha값의 선택으로 에너지 손실을 조절할 수 있다(Thomson, 1982). 본 논문에서는 일반적으로 이용되는 alpha=3.0인 Kaiser-Bessel 윈도우를 사용하였다(Fig. 2). 세 번째 과정은 필터링으로서, 본 연구에서의 가장 중요한 목적은, 관심 주파수 밖인 주기 3시간 이상의 저주파수 성분을 제거하는 것(low-cut filtering or detrending)과 환경적인 잡음으로 심각하게 오염되어 있다고 생각되는 주기 1분 이하의 고주파수 성분의 영향을 감소시키는 것(high-cut filtering)이다. 이러한 목적으로 주파수 영역에서 필터를 설계하여, 주 관심 주파수 대역인 1분~180분 이외의 신호는 필터링하여

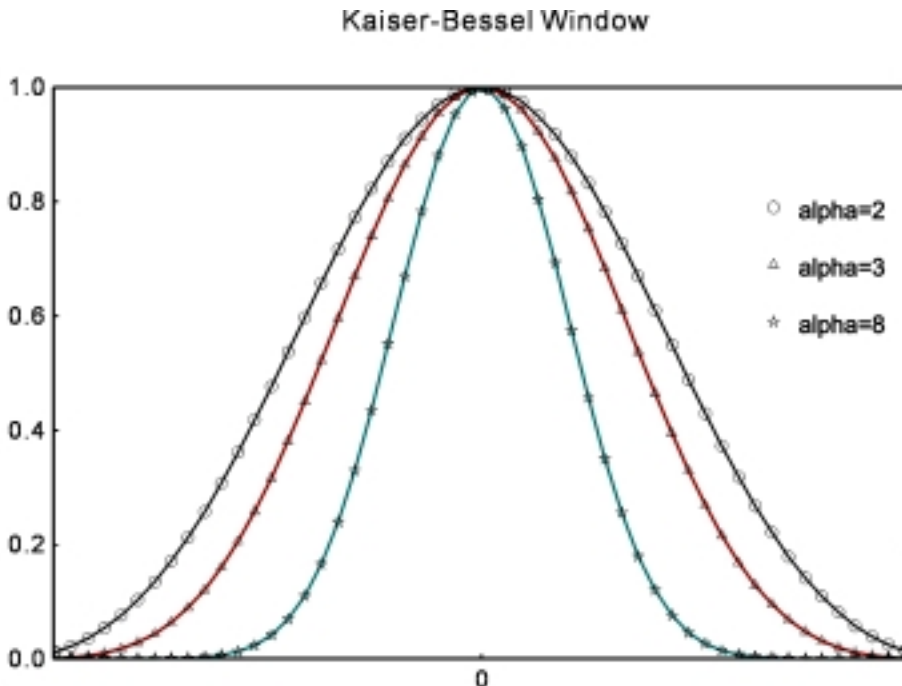


Fig. 2. Kaiser-Bessel window for alpha=2, 3, and 8.(Thomson, 1982).

제거하였다.

### 2.3. 스펙트럼 분석

전 처리 과정 후, 자료에 대한 신속한 푸리에 변환을 위해 FFT(Fast Fourier Transform)을 사용하였다. FFT 결과는 고주파수 부분에서 필요 이상의 주파수 성분을 제공하므로, 다루어야 할 주파수 수를 감소시키기 위해 이를 적절히 조절하여야 한다. 본 논문에서는 한 decade당 6-8개의 주파수를 얻었고, 중심 주파수의 25%를 대역폭으로 선택하였다(Vozzof, 1986; Nabighian, 1987).

과거의 연구자들은 계산시간 때문에, 긴 시계열 자료를 모두 푸리에 변환하지 않고, 짧은 길이의 시계열 자료에서는 고주파수 성분을 얻고, 보다 긴 길이의 시계열 자료에서는 저주파수 성분을 얻는 cascade decimation을 사용하였다(Wight · Bostick, 1980; Egbert · Booker, 1986). 예를 들어, 8초 간격으로 측정된 자료가 1024개 있다고 할 때, 첫 번째로 64개의 시계열을 푸리에 변환하여 고주파수 성분을 얻고(decimation level 1), 그 후 64개의 자료에 반 알리아스(anti-alias) 필터링을 수행한 후 32초 간격으로 샘플링을 하여 16개의 시계열을 만들어 낸다. 이런 과정을 4번 반복하며 32초로 샘플링된 64개의 자료가 생성되며, 이 자료를 푸리에 변환하여 전 단계보다 낮은 주파수 성분을 얻을 수 있다(decimation level 2). 위의 과정을 자료의 길이를 고려하여 반복 수행하면 원하는 수준까지 저주파수 성분을 얻을 수 있다. 따라서 계산하는 자료의 수는 동일하지만 고주파수에서는 짧은 시계열 자료를, 저주파수에서는 긴 시계열 자료를 사용하므로 계산 시간 단축에 효율적이다. 반면 적절하지 못한 decimation 간격을 사용하였을 경우 각 decimation level의 주파수 성분이 함께 고려되면 파워 스펙트럼(power spectrum)의 변화 양상이 불안정해지는 결과를 얻을 수도 있다. 또한 강한 신호들에 의해 주파수 성분이 결정되므로 안정된 주파수 성분을 위해서는 적절한 decimation 간격의 선택과 각 decimation 단계에 따른 자료 윈도우(data windowing) 효과에 대한 보정이 필요하다.

근래 들어 컴퓨터 계산 능력의 눈부신 향상은 전 시간 자료에 대한 분석(full time-series FFT)

을 강력하게 지원한다. Cascade decimation에 비해 계산하는 자료의 수는 많지만 컴퓨터의 계산 능력 향상으로 문제가 되지 않으며, 분석하는 자료의 길이가 길므로 주파수 분해능을 향상시키는 역할을 한다. 그러나 강한 파워를 가진 신호의 경우, 그 신호에 관련된 주파수의 범위가 주변 주파수로 퍼지는 결과가 발생하며 이는 전체적으로 파워 스펙트럼을 부드럽게 변화시키지만 전체적인 신호의 파워를 감소시킨다. 그러나 이 효과는 대역 평균(band averaging)을 수행함으로써 극복할 수 있다고 판단된다.

한편 본 연구에서는 지자기 자료로부터 얻을 수 관심 주파수 대역이 상당히 저주파 대역( $10^{-4}Hz \sim 10^{-1}Hz$ )이기 때문에, 지자기장의 파워 스펙트럼 곡선이 저주파수에서 고주파로 가면서 부드럽게 감소한다는 가정하에(Vozzof, 1972) 부드러운 파워 스펙트럼을 도출하는 데 초점을 맞추었다. 본 연구에서는 cascade decimation을 수행과 함께, 5초로 샘플링 된 22시간 길이의 시계열 자료를 분석하여 주파수 성분을 얻었다. 주파수 범위는 1분~150분까지이며, 이 주파수 범위는 분석 대상인 시계열의 길이가 22시간이므로 신뢰성 있는 주파수 범위라 판단된다(Shimoizumi *et al.*, 1997). 앞의 두 방법으로 계산한 주파수 성분 중 본 연구에서는 보다 부드러운 주파수 성분을 나타내는 22시간 자료 분석 결과를 선택하였다.

본 연구에서는 총 10일의 지자기 자료를 분석하였는데, 이 자료는 22시간 단위의 하위 자료로 나누어졌다. 각 하위 자료는 독립적으로 대역 평균을 수행하여 주파수 성분을 얻고, 그 다음 모든 하위 자료의 동일한 대역의 주파수 성분을 평균(block averaging)하였다(Chave *et al.*, 1987). 위의 평균과정은 주파수 성분의 신뢰구간을 좁게 하고 무작위적인 잡음의 영향을 최소화 할 수 있어 주파수 성분을 안정적으로 얻을 수 있으므로, 지자기 전달함수의 부드러운 추정에 큰 기여를 한다.

### 2.4. 지자기 전달함수의 추정

GDS의 관점에서는 임의의 주파수  $\omega$ 에서 수직 성분의 자기장은 수평 성분의 자기장과 오차의

선형적 관계로 표현될 수 있다고 가정한다. 예를 들어 N개의 i번째 Z와 H의 관측값은

$$Z_i(\omega) = H_i(\omega)^T T(\omega) + e_i(\omega) \quad (1)$$

or

$$Z_i(\omega) = T_1(\omega)H_{i1}(\omega) + T_2(\omega)H_{i2}(\omega) + e_i(\omega)$$

으로 표현된다. 여기서  $Z_i$ 는 지자기장의 수직 성분,  $H_i$ 는 지자기장의 수평성분( $H_{i1}$ 은 북쪽 방향의 자기장 성분,  $H_{i2}$ 는 동쪽 방향의 자기장 성분),  $T(T_1, T_2)$ 는 지자기 전달함수,  $e_i$ 는 에러, 위첨자  $T$ 는 전치행렬을 나타낸다.

식 (1)은 주파수 영역에서 다루어지므로 에러를 포함하여 모두 복소량이다. 본 연구에서는 계산의 편의상, 모든 양을 실수 값으로 변환하였고 이를 표현하면 다음과 같다.

$$Z = \begin{pmatrix} Re Z_1 \\ Im Z_1 \\ \dots \\ \vdots \\ \dots \\ Re Z_N \\ Im Z_N \end{pmatrix},$$

$$H = \begin{pmatrix} Re H_{11} & -Im H_{11} & Re H_{12} & -Im H_{12} \\ Im H_{11} & Re H_{11} & Im H_{12} & Re H_{12} \\ \dots & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \dots & \dots \\ Re H_{N1} & -Im H_{N1} & Re H_{N2} & -Im H_{N2} \\ Im H_{N1} & Re H_{N1} & Im H_{N2} & Re H_{N2} \end{pmatrix}$$

$$e = \begin{pmatrix} Re e_1 \\ Im e_1 \\ \dots \\ \vdots \\ \dots \\ Re e_N \\ Im e_N \end{pmatrix}, T = \begin{pmatrix} Re T_1 \\ Im T_1 \\ Re T_2 \\ Im T_2 \end{pmatrix} \quad (2)$$

이를 실수 행렬로 표시하면 다음과 같다.

$$Z = HT + e \quad (3)$$

식 (3)은 "다중 회귀분석(multiple regression)"이라는 통계학적인 문제로 귀결되며 T에 대한 추정은

$$T^* = (H^T H)^{-1} H^T Z \quad (4)$$

이고, 위의  $T^*$ 는 아래의 잔차들의 제곱을 최소화시키는 최소자승 추정이다.

$$\sum_{i=1}^{2N} r_i^2 = \sum_{i=1}^{2N} (Z_i - H_i^T T^*)^2 \quad (5)$$

식 (4)는 오차가 가우스 분포를 따를 때, 즉 오차가 서로 독립적이며, 동일한 분산을 갖는 다는 조건에서 최적의 추정이 되며, 이 조건에서 오차 벡터 e의 공분산 행렬은

$$E(ee^T) = \Sigma = \sigma^2 I \quad (6)$$

이고, 이 때 최적의 불편 추정(linear unbiased estimation)은 다음과 같다.

$$T^* = (H^T \Sigma^{-1} H)^{-1} H^T \Sigma^{-1} Z \quad (7)$$

그러나 조건을 완화하여 오차의 분산이 각 자료마다 다를 경우( $Var(e_i) = \sigma_i^2$ ), 다음의 가중 합을 최소화하는 가중 최소자승 추정을 얻을 수 있다.

$$\sum_{i=1}^{2N} w_i r_i^2 = \sum_{i=1}^{2N} w_i (Z_i - H_i^T T^*)^2 \quad (8)$$

여기서 가중치  $w_i$ 는  $1/\sigma_i^2$ 에 비례한다.

최소자승 추정은 오차가 가우스 분포를 따른다면 최적이지만, 실제 오차 분포에 많은 영향을 받는다. 위와 같이 가정을 완화한 가중 최소자승 추정 혹은 가우스 분포를 따르지 않는 오차 구조에 대한 많은 연구가 이루어졌는데 본 연구에서는 Egbert · Booker(1986)가 소개했던 regression M-estimate를 적용하였다. Regression M-estimate는

예측치와 관측치의 차이를 최소화한다는 의미에서 최소자승 추정과 유사하지만 잔차의 측정이 몇몇의 외치(outlier)가 추정에 큰 영향을 미치지 않도록 정의된다. 이는 L1-norm의 특별한 경우로 볼 수 있다.

통계적인 입장에서 문제를 정의하면, M-estimation은 다음 식을 최소화시키는 것이다.

$$\sum_i \rho \left( \frac{Z_i - H_i^T T^*}{\sigma} \right) \quad (9)$$

여기서  $\rho(r)$ 은 적당한 손실 함수(loss function)이다. 최소자승 추정에서는  $\rho(r) = r^2/2$ 를 사용하며 L1-norm의 경우  $\rho(r) = |r|$  이다.  $\rho(r)$ 에 대한 여러 가지 함수가 제안되었으나, 본 연구에서는 L1, L2-norm의 혼합형태로 구성된 손실 함수(loss function)를 이용하였고(Huber, 1981), Fig. 3는 Huber가 제안한 손실 함수, 영향 함수(influence function), 가중 함수(weighting function)의 형태를 개략적으로 나타낸 것이다.

$$\rho(r) = \begin{cases} r^2/2, & |r| < r_0 \\ r_0|r| - r_0^2/2, & |r| \geq r_0 \end{cases} \quad (10)$$

여기서  $r_0 = 1.5$  이다.

식 (9)은 일반적으로 잔차를 정규화 하는 scale parameter  $\sigma$  에 의존적이다. 식 (10)에 정의된 손실 함수는  $r_0$  보다 작은 잔차에 대해서는 L2-norm을,  $r_0$  보다 큰 잔차에 대해서는 L1-norm을 최소화 킨다. 식 (10)를 최소화는 것은 다음의 식의 해를 구하는 것과 동일하며,

$$\sum_i \psi \left( \frac{Z_i - H_i^T T^*}{\sigma} \right) H_i = 0 \quad (11)$$

여기서  $\psi(r) = \rho'(r)$  이며, 영향 함수(influence function)라 정의된다. 실제 식 (11)의 해는 외치에 하향 가중(down weighting)을 주는 반복적인 최소자승 추정으로 계산할 수 있다. 우선

$$w(r) = \psi(r)/r \quad (12)$$

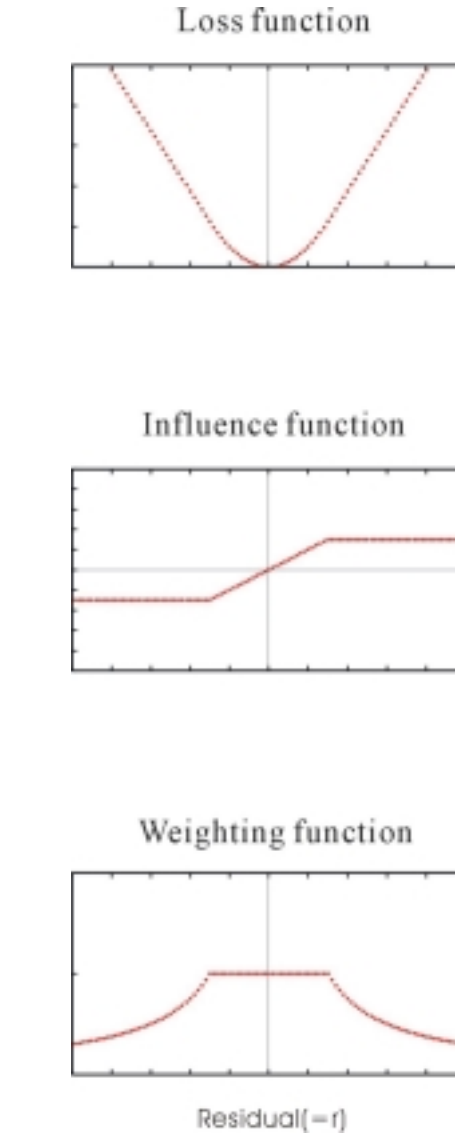


Fig. 3. Example of Huber's M-estimator (Hber, 1982). The discontinuous point is  $-r_0, r_0$ .

이라는 가중 함수(weighting function)을 도입하여 최소자승 추정을 이용해 첫 반복 단계의 error scale  $\sigma_0^*$ 와 전달함수 추정값  $T_0^*$ 를 이용하여 Z의 예측치와 잔차를 계산한다.

$$\begin{aligned} Z_{\hat{d}}^* &= H_i^T T_0^* \\ r_{\hat{d}} &= Z_i - Z_{\hat{d}}^* \end{aligned} \quad (13)$$

그 다음 보정 된 관측치(modified observation)를 계산한다.

$$Z_{d1} = Z_{d0}^* + w(r_{d0}/\sigma_0^*)r_{d0} \quad (14)$$

여기서 가중 함수  $w$ 는

$$w(r) = \begin{cases} 1, & |r| \leq r_0 \\ r_0/|r|, & |r| \geq r_0 \end{cases} \quad (15)$$

으로 주어지며  $r_0$ 보다 작은 잔차에 대해서는 보정된 관측치와 전 단계의 관측치가 동일하게,  $r_0$ 보다 큰 잔차에 대해서는 추정된 관측값에 가깝도록 정의한다. 그 다음 반복단계에서는 보정된  $Z_1$ 가 갱신되며 다음과 같이 에러 스케일(error scale)  $\sigma_1^*$ 을 가진 새로운 최소자승 추정  $T_1^*$ 이 계산된다

$$T_1^* = (H^T H)^{-1} (H^T Z_1) \quad (16)$$

위의 과정을 반복하여 regression M-estimate를 수행할 수 있다. 여기서 에러 스케일  $\sigma_k$ 는 첫 단계에서는 표준편차와 동일하나 두 번째 단계부터는 외치에 영향을 덜 받도록 스케일링(scaling)을 해야하는데

$$\sigma_k'^2 = \left[ \frac{1}{\beta(2N-4)} \sum_i (Z_{ik} - Z_k')^2 \right] \quad (17)$$

$r_0=1.5$ 일 때,  $\beta = 0.7784$ 라고 알려져 있다(Egbert · Booker, 1986). 본 연구에서 수행한 가중 최소자승의 반복 과정은 수렴이 보장되며(Huber, 1981), 수렴 조건에 도달한 후 마지막 단계로서 외치라 판단되는 자료를 제외하고 다시 최소자승 추정을 수행하여 최종적인 해를 얻게 된다(Egbert · Booker, 1986).

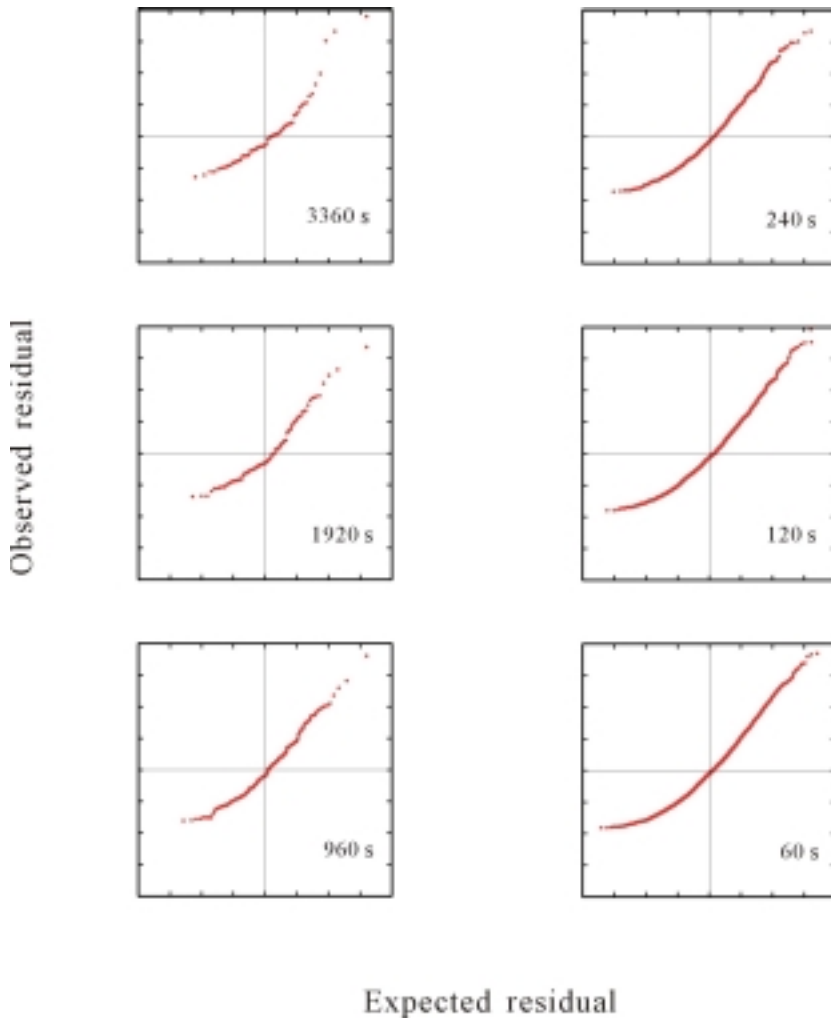
### 3. 결과 및 토의

앞 절에서 언급했던 바와 같이, 오차의 구조는

지자기 전달함수의 추정에서 중요한 요소이다. Gauss-Markov 이론에 따르면 오차가 가우스 분포를 따를 때 또는 근사적으로 가우스 분포를 따른다면 최소자승 추정은 모든 선형 추정에서 최적의 추정이다. 그러나 오차가 가우스 분포를 따르지 않을 때 최소자승 추정은 외치에 심각한 영향을 받으며, 극단적인 경우 전달함수는 몇 개의 외치에 의해 결정될 수도 있다(Egbert · Booker, 1986).

오차의 가우스 분포 가정을 검증하는 방법은 오차가 정규 가우스 분포를 따를 때 예상되는 잔차와 실제 잔차를 비교하여 도시하는 것으로서, 이를 Q-Q(Quantile-quantile) plot 이라고 한다. 최소자승 추정에 의해 계산된 잔차를 크기 순으로 정렬하고, 적당한 통계값으로 스케일링 한 후, 가우스 분포를 따르는 경우 예상되는 값과 실제 잔차가 위치하는 곳에서의 값을 쌍으로 하여 그래프로 도시하는 과정을 수행하면 Q-Q plot을 얻을 수 있다(Egbert · Booker, 1986; Chave *et al.*, 1987). 오차가 가우스 분포를 따른다면, Q-Q plot은 직선 형태로 나타나며, 꼬리가 긴 형태일 수록 오차가 가우스 분포에서 벗어난다는 것을 의미한다. 본 연구에서는 이천과 용인에서 획득한 지자기 자료를 이용하여 첫 반복단계에서 추정된 전달함수를 기준으로 각 주파수 별 잔차를 계산하였다.

Fig. 4는 이천 관측소에서 얻은 지자기 자료를 분석하여 얻은 Q-Q plot을 나타낸 것이다. 전 주파수에서 대체적으로 직선 형태를 보이나 주기 3360초의 저주파 대역과 240초 이하의 고주파 대역에서 약간의 꼬리의 형태가 보인다. 반면 1920초와 960초의 중간 대역에서는 거의 직선 형태이며 가우스 분포를 가정해도 무방하다고 판단된다. 저주파 대역의 꼬리 형태는 균질한 소스 가정이 깨지는 시기에 주로 나타나는 현상으로서 이 대역의 자기 전달 함수는 오차가 가우스 분포를 근사적으로 따르지만 불균질한 소스에 의해 가우스 분포를 벗어나는 몇몇의 외치가 자료에 존재함을 의미한다. 한편 주기 240초 이하의 고주파 대역의 꼬리 형태는 환경적인 잡음의 영향이라고 판단되지만 꼬리의 형태가 직선의 형태에서 크게 벗어나 있지는 않다.



**Fig. 4.** Q-Q(Quantile-quantile) plots at the Icheon observatory. They show tailed residual distributions for long periods due to source effect and short periods due to local environmental noises.

Fig. 5는 이천에서 최소자승 추정과 로버스트 추정으로 계산된 전달 함수의 크기와 위상(실수 부분)을 나타낸 것이다. Q-Q plot에서 예상되었던 것처럼, 거의 가우스 분포를 따르는 중간 주파수 대역에서의 전달함수 추정은 두 방법 모두 상당히 유사한 결과를 나타내고, 저주파 대역과 고주파 대역의 경우 최소자승 추정보다 외치에 하향 가중을 주는 로버스트 추정에 의한 전달 함수가 보다 부드럽게 조정되었음을 확인할 수 있다. 물론 부드러운 해가 정확하다는 근거는 없지만, 부드럽지 않은 해는 필연적으로 잡음으로 오염되어

있다는 것은 의미한다(Egbert · Booker, 1986).

Fig. 6은 용인 관측소에서 얻은 지자기 자료를 분석하여 얻은 Q-Q Plot을 나타낸 것이다. 용인의 경우, 주기 3360초의 저주파 대역에서 약간의 꼬리가 보이나 근사적으로 직선 형태이며 240초 이하의 고주파 대역에서는 상당히 긴 꼬리 형태가 나타난다. 반면 주기가 1920초와 960초의 중간 주파수 대역은 직선 형태라 가정해도 문제없다고 생각된다. 이천 같은 시기에 측정된 지자기 자료이기 때문에 이천 관측자료와 마찬가지로, 저주파 대역의 약간의 꼬리 형태는 균질한 소스가



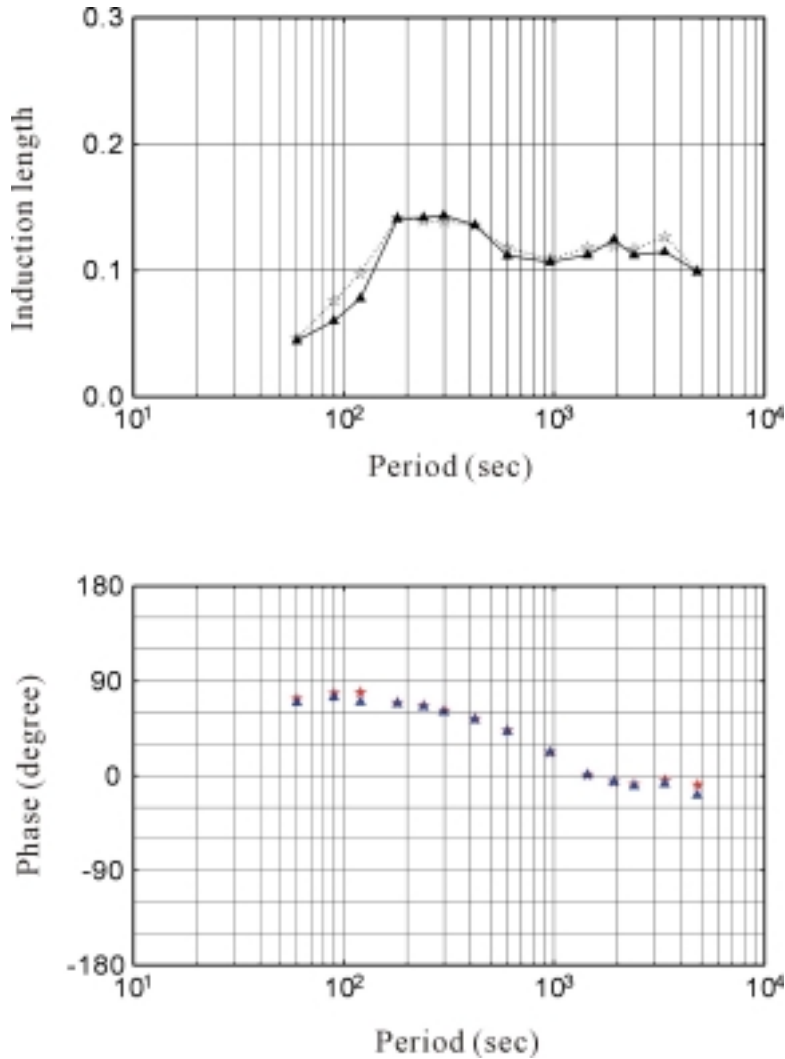


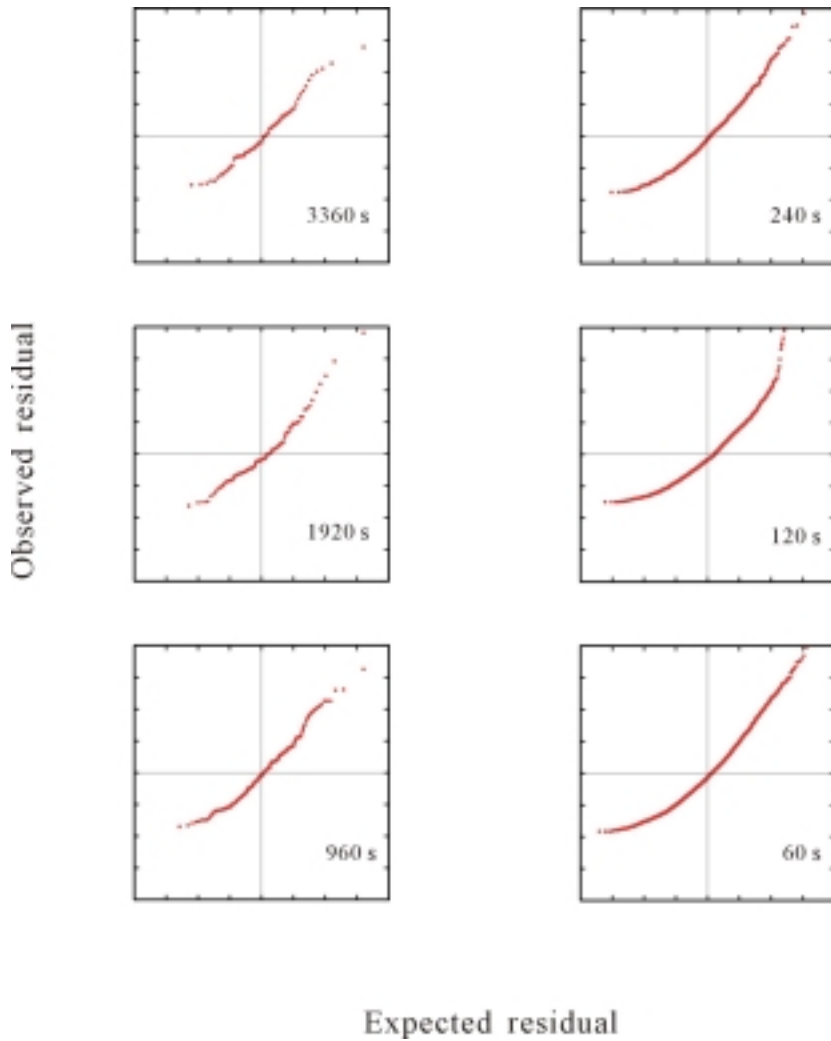
Fig. 5. Induction length and phase of geomagnetic transfer function at Icheon observatory(after Parkinson's convention). The star(☆) and the triangle(▲) denote conventional least square estimation, robust estimation, respectively.

정이 깨지는 시기 때문이라고 판단할 수 있으며, 고주파 대역의 긴 꼬리 형태는 환경적인 잡음의 영향 때문이라고 생각된다.

Fig. 7은 용인에서의 표준 최소자승 추정과 로버스트 추정으로 계산된 전달함수의 크기와 위상(실수 부분)을 나타낸 것이다. 저주파수 대역에서는 최소자승 추정과 로버스트 추정에 의한 결과가 유사하나 로버스트 추정으로 보다 부드러운 형태로 조정이 되었다. 고주파수 대역에서는 외치에 의해 과평가 되었던 전달함수의 크기가 감소

하는 형태로 조정된 것을 확인할 수 있다. 반면 중간 주파수 대역의 경우는 최소자승 추정과 로버스트 추정의 결과가 거의 차이가 나지 않음을 확인할 수 있다.

위의 결과들은 오차의 구조가 가우스 분포 가정을 만족하거나 또는 근사적으로 만족한다면 최소자승 추정과 로버스트 추정의 결과가 차이가 없다는 것을 보여주며, 가우스 분포에서 벗어나는 오차의 경우 최소자승 추정보다 로버스트 추정이 보다 부드러운 결과를 얻을 수 있음을 의미한다.



**Fig. 6.** Q-Q(Quantile-quantile) plots at the Yongin observatory. They show tailed residual distributions for short periods due to local environmental noises.

그러나 오차가 심각하게 외치에 의존하는 경우 (외치가 자료의 20-30% 이상을 차지할 경우), 최소자승 추정에 비해 로버스트 추정이 안정적이고 부드러운 결과를 낸다고 보장할 수 없다(Chave *et al.*, 1987). 또한 로버스트 추정은 첫 반복단계의 최소자승 추정에 민감하다. 첫 반복단계의 추정에서 가장 좋은 신호 대 잡음비(S/N)를 가진 신호, 즉 가장 강한 파워를 가진 신호가 가장 큰 잔차를 가진 자료로 인식될 수 있으며, 이 자료에 하향 가중이 주어지면 결과가 심하게 왜곡될 수 있다. Egbert · Livelybrooks(1996)에 의하면 낮은

신호 대 잡음비를 가진 자료의 경우 로버스트 추정은 최소자승 추정보다 더 왜곡된 결과를 보이는 것으로 나타났다.

로버스트 추정은 오차의 분산이 동일하다는 표준 최소자승 추정의 조건을 완화시켜 자료마다 오차의 분산이 다르다는 조건을 사용하고, 또한 여러 스케일의 경우도 현 단계가 아닌 전 단계의 결과를 이용하기 때문에 근사적인 추정 방법이며, 자료에 외치가 일정 수준이상 포함되거나(20-30% 이상) 낮은 신호 대 잡음비를 가진 자료에 대해서는 최소자승 추정에 비해 안정적이고 부드러운

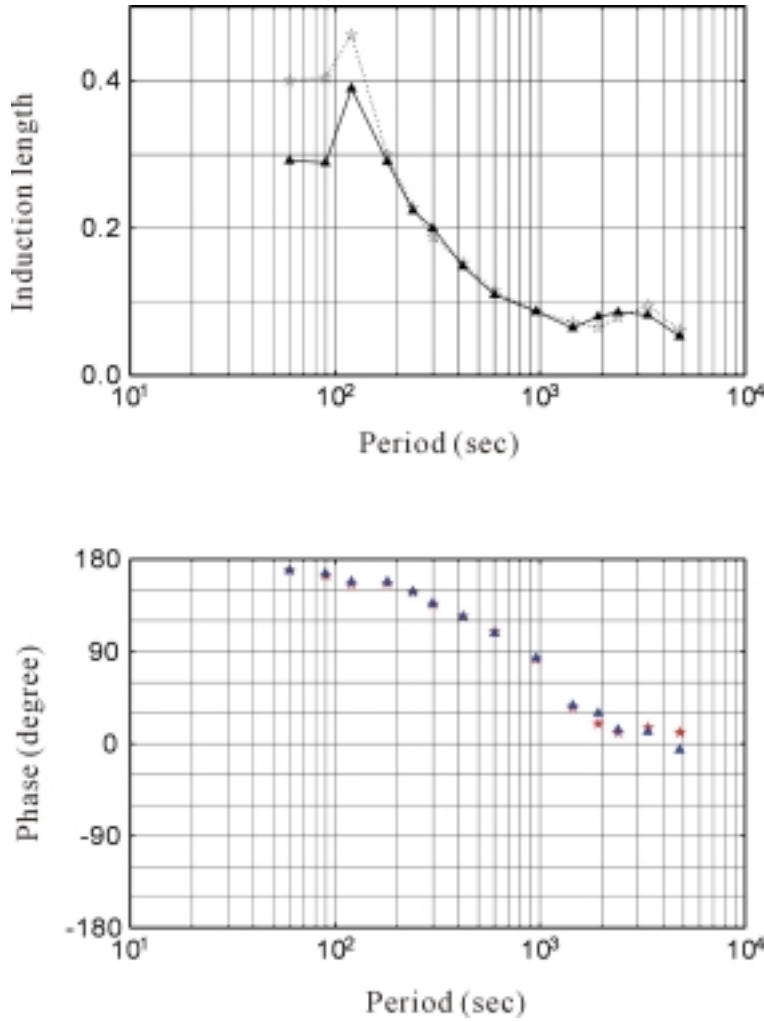


Fig. 7. Induction length and phase of geomagnetic transfer function at Yongin observatory(after Parkinson's convention). The star(☆) and the triangle(▲) denote conventional least square estimation, robust estimation, respectively.

해를 보장할 수 없다는 의미이다.

#### 4. 결론

지자기 전달함수 추정은 전통적으로 관측치와 예측치의 차이를 최소화한다는 최소자승의 의미이며, 오차가 독립적이며 동일한 분산을 갖는다는 가우스 분포를 만족한다고 가정한다. 그러나 지자기장의 특성상 비정상적인(non-stationary) 신호나 균질한 소스 가정이 깨지는 자기 폭풍 시기, 환경적인 잡음이 강한 경우 오차의 가우스 분포 가정

이 비현실적이다.

따라서 본 연구에서는 지자기 전달함수 추정 시 최소자승 추정이 가정하는 오차의 가우스 분포 가정의 타당성을 Q-Q plot을 이용하여 검증하였다. Q-Q plot을 통해 얻은 오차 구조에 대한 정보는 최소자승 추정의 신뢰성을 판단할 수 있는 근거가 되며, 이를 근거로 적절한 추정방법이 선택될 수 있다.

오차 구조가 근사적으로 가우스 분포를 따르는 경우 최소자승 추정과 로버스트 추정의 결과는 차이가 없으며 최소자승 추정만으로도 충분히 신

뢰성 있는 결과를 도출할 수 있다. 반면 오차 구조가 가우스 분포를 벗어나는 경우 외치에 하향 가중을 주는 로버스트 추정 방법이 최소자승 추정보다 부드러운 결과를 보여준다. 그러나 외치가 일정 수준이상(20-30% 이상) 포함되어 있거나 신호 대 잡음비가 낮은 경우에는 로버스트 추정이 최소자승 추정보다 안정적이고 부드러운 해를 도출한다고 보장할 수는 없다.

결론적으로 지자기 전달함수의 추정 시 전달함수에 대한 신뢰성과 추정 방법의 타당성 확인하기 위해서는 본 연구에서 제시한 바와 같이 오차 구조에 대한 조사가 수행되어야 하며, 오차의 구조에 따른 적절한 추정 방법의 선택은 추정 결과에 대한 정확성과 신뢰성을 향상시킬 것으로 사료된다.

### 사 사

이 연구는 기상청 기상연구소 기상지진기술개발 사업의 「지진예지 및 지진해일 예측기술 개발」의 지원으로 수행되었으며 이에 사의를 표합니다.

### 참고문헌

Chave, A. D., Thomson, D. J., and Ander, M. E., 1987, On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res.*, 92, 633-648

Egbert, G., and Booker, J. R., 1986, Robust estimation of geomagnetic transfer functions, *Geophys. J. Roy. Astr. Soc.*, 87, 173-194  
Egbert, G. D., and Livelybrooks, D. W., 1996 Single station magnetotelluric impedance estimation: Coherence weighting and the regression M-estimate, *Geophysics*, 40, 964-970  
Huber, P. J., 1981, *Robust Statistics*, Wiley, New York  
Nabighian, N., 1987, *Electromagnetic methods in applied geophysics-Applications, Part B*, Soc. Expl. Geophys., Tulsa, OK  
Rikitake, T., Honkura, Y., 1985, *Solid earth geomagnetism*, pp. 295-347, Terrapub, Tokyo  
Shimoizumi, M., Mogi, T., Nakada, M., Yukutake, T., Handa, S., Tanaka, Y., and Utada, H., 1997, Electrical conductivity anomalies beneath the western sea of Kyushu, Japan, *Geophys. Res. Lett.*, 24, 1551-1554  
Thomson, D. J., 1982, Spectrum estimation and harmonic analysis, *Proc. IEEE*, 70, 1055-1095,  
Vozzof, K., 1972, The magnetotelluric method in the exploration of sedimentary basin, *Geophys.*, 37, 98-141  
Vozzof, K., 1986, *Magnetotelluric Methods*, Soc. Expl. Geophys. Reprint Ser. No 5, Soc. Expl. Geophys., Tulsa, OK  
Wight, D. E., Bostick, F. X., 1980, Cascade decimation-A technique for real time estimation of power spectra, *Proc. IEEE International Conference on Acoustic Speech and Signal Processing*, 626-629

---

2002년 4월 30일 원고접수  
2002년 6월 21일 원고채택