

다중빔 음향측심기 자료의 이상치 보정에 관한 연구

이정숙 · 김수영 · 이용국¹ · 신동완 · 김한준^{1*} · 주형태¹

이화여자대학교 통계학과, ¹한국해양연구소 환경기후연구본부

A Study on Outlier Adjustment for Multibeam Echosounder Data

JUNG-SOOK LEE, SOO-YOUNG KIM, YONG-KOOK LEE¹, DONG-WAN SHIN,
HAN-JOON KIM^{1*} AND HYEONG-TAE JOU¹

Department of Statistics, Ewha University, Seoul 120-750, Korea

¹Korea Ocean R & D Institute, Ansan P.O. Box 29, Seoul 425-600, Korea

해저면의 지형과 형상을 조사하기 위해 획득되는 다중빔 음향측심자료는 조사선의 움직임과 빔의 위치에 따른 음압의 보정부족 등에 기인하는 이상치를 포함하고 있다. 이 연구에서는 자기회귀이동평균(ARMA)기법을 이용하여 이상치에 의해 왜곡된 자료를 보정해주는 통계학적 절차를 소개하고 이를 동해에서 얻어진 자료에 적용하였다. 이 방법에서는 자료의 자기 상관을 1차 자기회귀모형 AR(1)으로 표현하고 이를 추정된 후 추정된 모형에 현저하게 부합하지 않는 자료(이상치)는 추정된 모형을 이용하여 수정한다. 이러한 방법을 이상치가 발견되지 않을 때까지 반복적으로 되풀이한다. 자료의 처리결과 해저지형을 나타내는 신호에 비해 월등히 큰 이상치들이 대부분 제거되었음을 볼 수 있다.

Multibeam echosounder data, collected to investigate seabed features and topography, are usually subject to outliers resulting from the ship's irregular movements and insufficient correction for pressure calibration to the positions of beams. We introduce a statistical method which adjusts the outliers using the ARMA (Autoregressive Moving Average) technique. Our method was applied to a set of real data acquired in the East Sea. In our approach, autocorrelation of the data is modeled by an AR (1) model. If an observation is substantially different from that obtained from the estimated AR (1) model, it is declared as an outlier and adjusted using the estimated AR (1) model. This procedure is repeated until no outlier is found. The result of processing shows that outliers that are far greater than signals in amplitude were successfully removed.

서 론

해저면 탐사를 위하여 한국해양연구소에서는 종합해양조사선 “온누리” 호의 건조와 맞추어 미국의 SeaBeam Instruments Inc.의 다중빔 음향측심기 SeaBeam 2000을 도입하였다. SeaBeam 2000은 12 KHz의 주파수 특성을 가진 28개의 음원 발생기를 동시에 동작시켜, 84개의 hydrophone으로 해저면에서 반사되어 돌아온 신호를 수신하여 해저면의 지형을 측정한다. 이때, 조사선 항로에 수직 방향으로 일정한 폭(90~120°)에 대한 수심을 동시에 측정하여 지형자료를 보정함으로써 정확한 수심자료를 2차원 혹은 3차원으로 나타낼 수 있다. SeaBeam 2000 자료가 가지는 문제점의 하나로서, 순간적인 배의 요동 등에 의해 일부 자료가 유실되거나 미관측되며, 측정치와는 상관없이 큰 값을 갖는 오차가 이상치(outlier)로서 많이 나타난다. 이러한 이상치들은 일반적으로 측정치보다 더 큰 진폭을 가지므로 자료의 해석을 위해서 제거되어야 한다. 이상치를 제거하기 위한 쉬운 방법은 주파수 필

터링이지만 그 적용결과는 이상치 주변의 신호들을 왜곡시키며 전체적으로 자료의 분해능을 떨어뜨린다. SeaBeam 2000에서 측정하는 자료의 또 다른 문제점은 측정되는 반사신호가 음원의 발생각도(beam의 위치)에 따라 값이 달라지는 것으로 음원의 발생각도에 따른 음압이 저하된다. 중앙의 beam은 해저면에 수직으로 향해 있지만 양 끝으로 갈수록 beam의 방향은 해저면에 대하여 더 큰 경사를 가지게 된다. 그러므로 beam의 위치가 중앙에서 멀어짐에 따라 신호가 해저면에서 반사되어 돌아올 때 이동하는 경로가 길어져서 수신 신호의 음압이 체계적으로 감소하게 되며 이에 따라 수신 신호의 신뢰성이 저하되게 된다. 그 결과로서, SeaBeam 2000으로 얻는 해저면의 수심도에는 beam의 양단에 해당되는 선을 따라 불연속선이 나타나게 된다. 이러한 문제를 해결하기 위한 여러 방안 중의 하나는 beam의 위치에 따른 음압 저하를 정확하게 추정하고 이 저하된 양을 정확하게 보정해 주는 것이다.

이 연구에서는 자기회귀이동평균(ARMA: Autoregressive Moving Average)법을 이용하여 이상치를 제거하는 통계학적 기법을 유도하고 이것을 이용하여 동해에서 얻은 SeaBeam 2000자료를 보정

*Corresponding author: hanjkim@kordi.re.kr

한 예를 예시하고자 한다. 최종적으로 결측치와 이상치가 보정된 자료로부터 음압저하를 추정하고자한다.

관측치와 이상치

Fig. 1a는 온누리호가 일정속도로 이동하면서 121개의 beam에서 등간격으로 250개 시점에서 측정한 반사음압을 3차원 그림으로 나타낸 것이다.

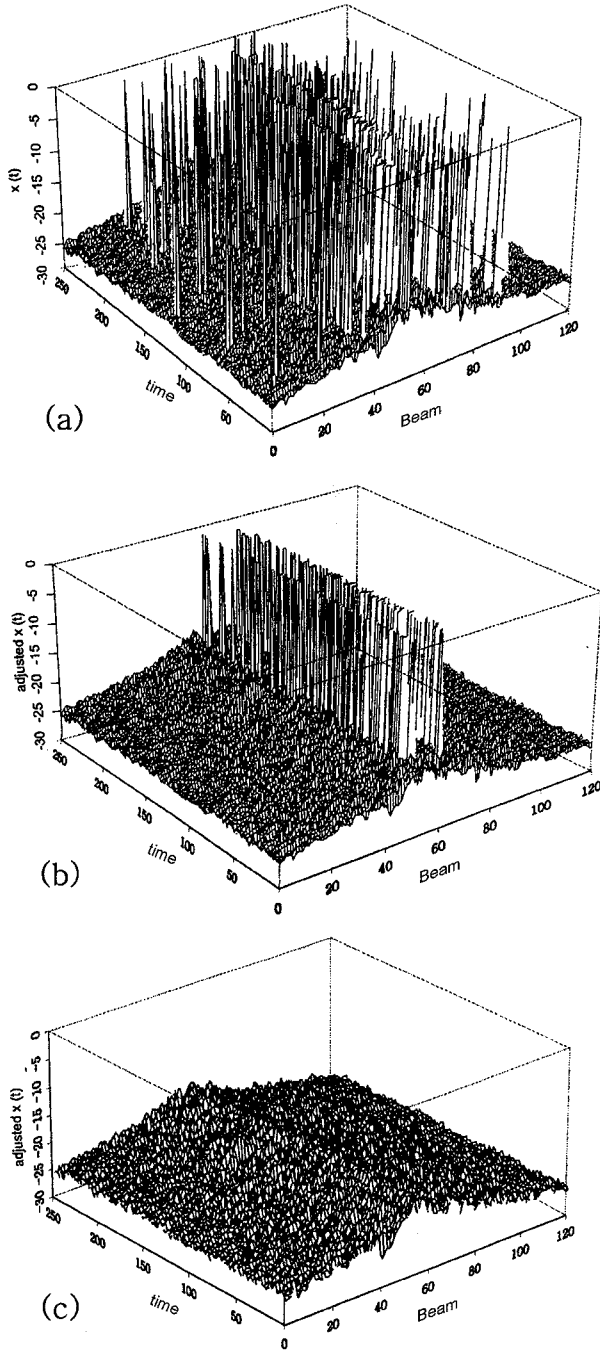


Fig. 1. Surface plot of amplitude: (a) observed data, (b) adjusted for outliers along time direction, and (c) adjusted first for outliers along time direction and then for beam direction.

로 나타낸 것이다. 다중빔 음향측심기로부터 얻어지는 자료는 다른 자료와 마찬가지로 풍랑 또는 불안정한 기상조건에 따른 순간적인 배의 요동과 기기의 오작동 등에 의해 일부자료가 유실되거나 미관측될 뿐만 아니라(결측치, missing value) 자료가 비정상적인 값(이상치, outlier)을 갖게된다. Fig. 1a는 이러한 면을 극명하게 보여준다. 결측치는 0으로 기록되어 그림상에서는 수직으로 치솟는 선으로 나타나는데 그림을 보면 상당히 많은 결측치가 있음을 알 수 있다. 이 그림에는 결측치 뿐만아니라 이상치들도 많이 숨어있다. 다음 절에서 제시하는 통계학적 방법으로 Fig. 1a를 보정한 결과는 Fig. 1b와 1c에 나타나 있다. 결측치도 그 값이 0으로 기록되어 실제 지형을 나타내는 값과 현저히 다르므로 이상치로 간주할 수 있다. 따라서 결측치도 이상치로 간주하여 서술한다.

Fig. 2a는 64번째 beam에서 얻어진 자료를 시간별로 나타낸 것이다. 여기에서, 0의 값까지 수직으로 내려가는 결측치 뿐만 아니라 거의 -6 dB에 육박하는 10개 이상의 이상치들을 볼 수 있다. 자료를 측정할 해역은 일정한 수심을 갖는 평탄한 지역이고 또 beam의 위치도 64번째로 고정되었으므로 이 값들은 다른 평탄한 해저면을 나타내는 값들과는 현저히 다른 “이상한” 잡음인 것이다.

이러한 결측치와 이상치는 자료의 신뢰도를 현저히 저하시켜 거리에 따른 beam의 음압저하 추정의 통계적 정밀도를 현저히 감소시키며 해저 지형도의 불연속정도를 심화시키는 한 요인으로 작용하게된다. Fig. 2b는 이 논문에서 제시하는 방법으로 Fig. 2a의 결측치와 이상치들을 제거한 결과이다.

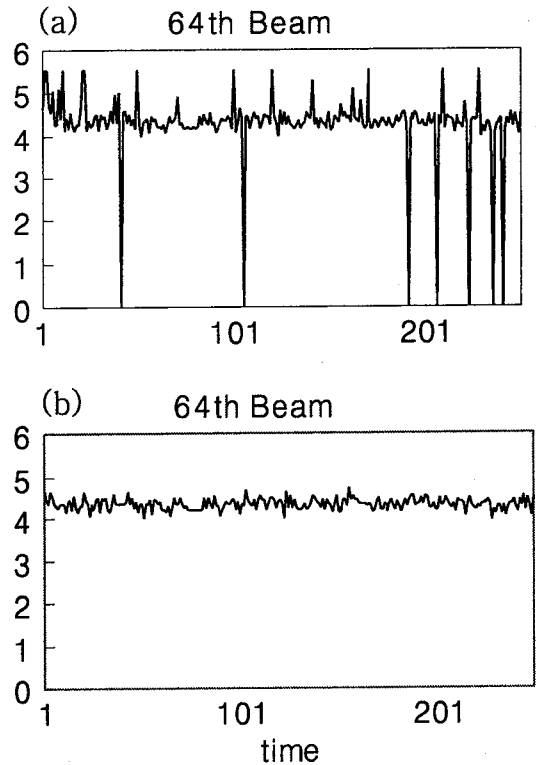


Fig. 2. Time series plots of (a) original data and (b) outlier adjusted data for beam 64.

이상치의 보정

다중빔 음향측심기로부터 얻어지는 자료에는 해저지형에 대한 신호외에도 배의 흔들림(rolling and pitching) 등에 의한 잡음이 포함되며 이들은 서로 관련성을 갖게 된다. 배의 진행 방향에서 생각했을 때 각 beam에서 얻어지는 자료는 시계열 자료(time series data)로 생각할 수 있고 여기에는 자기 상관(autocorrelation, serial correlation)이라는 관련성이 있다. 또 어떤 정해진 시점에서 각 beam에서 얻어지는 자료들을 일렬로 늘어놓았을 때 이 또한 하나의 시계열로 간주할 수 있고 자기상관을 내포하고 있다.

이상치를 발견하고 보정하는 통계적 방법의 첫 단계는 주어진 시계열 자료를 자기회귀 이동평균(ARMA)으로 모형화하는 것이다. 이는 자기상관을 하나의 통계모형으로 나타낸 것이다. 그 후 자료를 ARMA 모형으로 적합시킨 후 각 자료들에 대해 추정된 ARMA 모형과 부합하는가를 판단하여 그렇지 못하면 이를 이상치로 간주하고 추정된 ARMA 모형에 근거하여 수정하게 된다. 수정된 자료를 가지고 위 절차를 이상치가 나타나지 않을 때까지 되풀이함으로써 이상치를 제거할 수 있다.

ARMA 모형중 가장 단순하나 널리 사용되는 것이 1차 자기회귀모형 AR(1)이다. 이 모형의 장점은 이상치 보정에 있어서 비교적 간단한 절차를 제공해 주는 것이다. 실제의 자기상관은 대부분 알려지지 않았고 자기상관의 상당부분이 AR(1)에 의해 포착되기 때문에 AR(1) 모형을 적용해도 고차의 ARMA 모형을 적용했을 때에 비해 자기상관 표현의 부정확성은 그리 크지가 않다. 신호처리 이론에서 AR 프로세스의 파워스펙트럼은 최소위상(minimum-phase)이므로(Robinson, 1980) AR(1)이 주어진 신호의 주된 특성을 표현한다고 볼 수 있다. 이러한 이유로 이 연구에서는 자기상관을 AR(1)으로 표현하였다.

자기상관을 이용하여 이상치의 보정을 하는 여러 방법들을 통계학 문헌에서 찾아볼 수 있다. Fox(1972)는 이상치를 유형화하여 찾아내고 제거하는 과정에 대한 최초의 논의를 시작했다. Guttman과 Tiao(1978)는 이상치가 AR과 MA모수들을 추정하는 데에 심각한 편의(bias)를 일으킬 수 있다고 발표했다. 만일 이상치의 생성 구조를 이용할 수 있다면 이런 편차들은 Box와 Tiao(1975)의 개입분석기술을 사용함으로써 제거될 수 있다. 반면 이상치의 시점과 유형이 알려져 있지 않을 때 이런 추정 문제를 해결하기 위해서 Abraham과 Box(1979)는 Bayesian 접근을 했고, Chang 등(1988)은 반복적인 절차를 제시하였다. Tsay(1986)는 이상치가 ARMA 모형 적합에 일으키는 문제를 해결하기 위하여 모형을 추정하고 이상치를 탐색하여 제거하는 일련의 반복 절차를 제시하였다. 여기에서는 Tsay(1986)의 방법에 근거하여 음압자료의 이상치를 보정하는 방법을 제시하고자 한다.

AR(1) 모형에서의 이상치 보정 절차

이상치에 의해 오염된 일차 자기회귀(AR(1)) 과정 $x_t, t=1, \dots, n$ 이 주어졌을 때 이상치를 발견하고 보정해주는 절차를 기술한다. 아래의 Algorithm 1에서 AO(Additive Outlier)는 한 시점 T_A 에서 만 관측치에 가해진 이상치로 이는 시점 T_A 에서만 관측치에 영향을 미친다. IO(Innovation Outlier)는 시점 T_I 에 오차항에 가해진

이상치로 자기상관때문에 그 영향이 시점 T_I 이후에도 잔류하는 이상치를 말한다.

Algorithm 1.

Step 1 (AR(1) fitting)

자료 $x_t, t=1, \dots, n$ 에 대해 AR(1) 모형을 적합시켜 일차 자기회귀계수와 오차분산 $\hat{\rho}, \hat{\sigma}^2$ 를 다음과 같이 구한다.

$$\hat{\rho} = \frac{\sum_{t=2}^n (x_t - \bar{x})(x_{t-1} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

$$\hat{\sigma}^2 = \sum_{t=2}^n \{x_t - \bar{x} - \hat{\rho}(x_{t-1} - \bar{x})\}^2 / (n-1), \bar{x} = \sum_{t=1}^n x_t / n$$

Step 2 (Estimation of residuals, AO and IO effects)

$$e_t = (x_t - \bar{x}) - \hat{\rho}(x_{t-1} - \bar{x}), t = 2, \dots, n,$$

$$\hat{\omega}_{At} = (e_t - \hat{\rho}e_{t-1}) / (1 + \hat{\rho}^2), \lambda_{At} = (1 + \hat{\rho}^2)^{1/2} \hat{\omega}_{At} / \hat{\sigma}, t = 3, \dots, n,$$

$$\hat{\omega}_{It} = e_t, \lambda_{It} = \hat{\omega}_{It} / \hat{\rho}, t = 2, \dots, n,$$

Step 3 (Maximal AO and IO effects and their times)

$$T_A = \operatorname{argmax}_{3 \leq t \leq n} \lambda_{At}, \quad T_I = \operatorname{argmax}_{2 \leq t \leq n} \lambda_{It}$$

$$\lambda_A = \lambda_{AT_A}, \quad \lambda_I = \lambda_{IT_I}$$

Step 4 (Outlier adjustment)

만일 λ_A 가 3.5보다 크면 이상치를 AO유형으로 판별하고 λ_I 가 3.5보다 크면 IO유형으로 판별한다. λ_A 와 λ_I 중 큰 값에 대해 AO인 경우엔 다음과 같이 자료를 보정하고,

$$\tilde{X}_t = X_t \quad \text{for } t \neq T_A$$

$$\tilde{X}_{T_A} = X_{T_A} - \hat{\omega}_{AT_A}$$

IO인 경우에는 다음과 같이 보정한다:

$$\tilde{X}_t = X_t \quad \text{for } t = 1, \dots, T_I - 1$$

$$\tilde{X}_t = X_t - \hat{\rho}^{t-T_I} \hat{\omega}_{IT_I} \quad \text{for } t = T_I, \dots, n$$

Step 5: 보정된 자료를 가지고 이상치가 모두 없어질 때까지 Step 1부터 다시 반복한다.

음압자료의 이상치 보정

Fig. 1a에서 보는 바와 같이 음압자료는 121개의 beam에서 일정시간 간격으로 250개 시점에서 얻어졌다. 어떤 정해진 beam에서 얻어지는 자료는 크기가 $n=250$ 인 시계열이고 이러한 시계열이 121개 있다. 이를 종단면 시계열이라 정의한다. 또 어느 시점에서 얻어지는 자료는 크기가 $m=121$ 인 시계열로 생각할 수 있고 이러한 시계열이 모두 250개가 있다. 이를 횡단면(cross-sectional) 시계열이라 정의한다. 이상치의 보정절차는 아래와 같다.

1. 자료의 부호를 반대로 바꾸고 제곱근을 취한다. 그 이유는 beam의 중앙에서 멀어짐에 따라 자료값의 변동폭이 커지기 때문

에 등분산조건을 만족시키기 위해 분산안정화(variance stabilizing transform)의 일환으로 제곱근 변환을 한 것이다. 이후의 그림은 처리된 자료를 다시 역변환(제곱변환)하여 작성한 것이다.

2. 각 beam의 종단면 방향 250개 관측치에서 처음과 끝에 분실값이 많아 이상치 보정에 문제가 있으므로 양끝에서 10개 중에 0이 발견되면 근방의 0이 아닌 10개의 평균으로 대체한다.

3. 먼저 각 종단면 시계열을 AR(1)로 모형화시키고 이에 근거하여 Algorithm 1에 의해 이상치를 보정한다. Fig. 2a와 2b에는 각각 이상치 보정전과 후의 beam 64의 종단면 시계열의 그림이 나타나 있다. 이를 m=121개의 모든 종단면 시계열에 대해 반복한다. Fig. 1a의 종단면 방향의 이상치가 보정된 전체자료의 그림이 Fig. 1b에 나타나 있다.

4. 1에 의해 이상치가 보정된 시계열에 대해 각 beam의 250개 보정된 자료의 평균을 뺀다. 이를 평균수정 자료라 한다. 이렇게 함으로써 Beam의 위치에 따른 음압의 차이가 보정되어 평균수정된 횡단면 시계열자료는 모두 평균 0을 갖는 시계열로 간주할 수 있다. 평균 수정 횡단면 시계열을 AR(1)로 모형화시키고 Algorithm 1에 의해 이상치를 보정한다.

단계 4에서 평균 수정 자료를 횡단면으로 그려보았을 때 Fig. 3a에서와 같이 이상치가 있을 수 있다. 그 이유는 Fig. 1b에서 보는 바와 같이 beam 61, 69의 종단면자료는 거의 전부가 이상치이어서 종단면 자료에 근거한 이상치 보정은 성공적이지 못했기 때문이다. 아마도 이는 beam 61, 69에 장착된 수신기가 고장난 때문으로 생각된다. 아직 보정이 되지 않은 이들 이상치를 처리하기 위해 평균 수정 횡단면 시계열을 AR(1)로 모형화시키고 이상치를 보정하였다. Fig. 3b는 이상치가 보정된 평균수정 횡단면 시계열의 그림이다. 이러한 횡단면 방향으로의 이상치 보정을 n=250개의 평균수정 횡단면 시계열에 대해 되풀이한다. 횡단면 방향의 이상치가 보정된 전체 자료의 그림은 Fig. 1c와 같다.

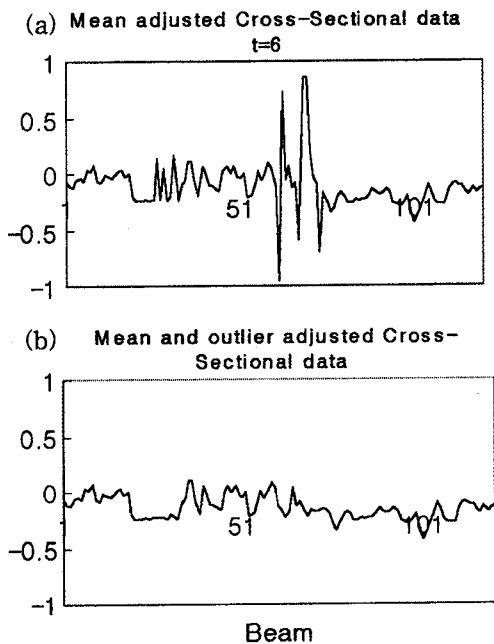


Fig. 3. Mean adjusted cross-sectional data at time 6 (a) before and (b) after outlier adjustment.

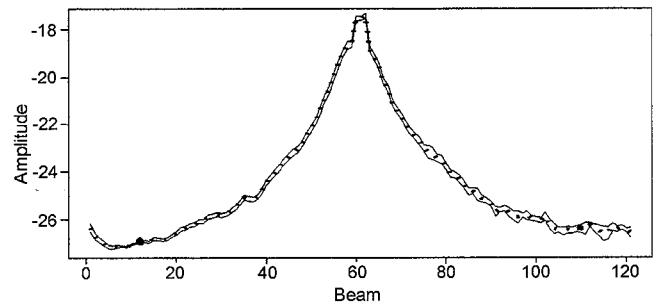


Fig. 4. Amplitude variation along beam positions. The surrounding band represents 95% confidence bounds.

Beam별 평균음압의 추정

각 beam별 평균 음압 추정은 위의 절차에 의해 보정된 자료의 각 beam별 종단면 자료의 단순 표본 평균에 의하여 구하였다. 종단면 자료가 자기상관을 갖기 때문에 일반화 최소자승 추정치가 통계적으로 최적임이 알려져 있으나 평균을 추정할 때에는 단순 표본평균도 시계열의 크기가 작지 않으면 거의 최적임이 알려져 있다(Fuller, 1996). 그렇기 때문에 계산이 간편한 단순평균을 사용했다. Fig. 4에 각 beam별 평균음압 추정치와 95% 신뢰 구간을 표시하였다. 이 그림에서는 beam의 중앙에서 멀어짐에 따른 음압 감소가 선명히 나타나 있다. 이 결과를 다중빔 음향측심자료로부터 해저면 지형도를 작성시 측면왜곡의 보정에 유용하게 사용할 수 있을 것이다.

결론

해저지형과 해저면의 형상을 얻기 위해 획득되는 다중빔 음향측심자료에는 해저지형자료와는 관계없는 큰 진폭의 이상치들이 존재한다. 이 연구에서는 자기회귀이동평균(ARMA)과정에 의해 이러한 이상치들을 제거하는 통계학적 기법을 유도하고 이것을 이용하여 실제로 동해에서 획득한 SeaBeam 2000자료를 보정한 결과를 제시함으로써 그 타당성을 검증하였다. 이 연구에서는 SeaBeam 2000의 종단면 및 횡단면자료를 각각 자기회귀모형 AR(1)로 모형화시켜 이상치를 보정하였다. 보정된 자료는 원래의 자료와 비교하였을 때 신호보다 그 진폭이 월등히 큰 이상치가 대부분 제거되어졌음을 보여준다.

사사

이 연구는 한국해양연구소의 기관고유사업 PE00783-1의 지원으로 수행되었다.

참고문헌

Abraham, B. and Box, G.E.P., 1979, Bayesian analysis of some outlier problems in time series. *Biometrika*, **66**: 229-236.
 Box, G.E.P. and G.C. Tiao, 1975. Intervention analysis with applications to economic and environmental problems. *J. Am. Stat.*

- Assoc.*, **70**: 70–79.
- Fox, A.J., 1972. Outliers in time series. *J. Roy. Stat. Soc., Series B*, **34**: 350–363.
- Chang, I., C.G. Tiao and C. Chen., 1988. Estimation of time series parameters in the presence of outliers. *Technometrics*, **30**: 193–204.
- Fuller, W.A., 1996. Introduction to Statistical Time Series. 2nd ed., Wiley, New York, 698 pp.
- Guttman, I. and G.C. Tiao, 1978. Effect of correlation on the estimation of a mean in the presence of spurious observations. *Can. J. Stat.*, **6**: 229–247.
- Robinson, E.A., 1980. Geophysical Signal Analysis, Prentice-Hall Inc., Englewood Cliffs, 466 pp.
- Tsay, R.S., 1986. Time series model specification in the presence of outliers. *J. Am. Stat. Assoc.*, **81**: 132–141.
-
- 2000년 9월 26일 원고접수
2000년 11월 23일 수정본 채택
담당편집위원: 김대철