

[Note]

한반도 연안 조위자료의 결측 양상 Missing Pattern of the Tidal Elevation Data in Korean Coasts.

조홍연* · 고동휘** · 정신택**
Hongyeon Cho*, Dong Hui Ko** and Shin Taek Jeong

요 지 : 우리나라 연안 조위자료의 결측양상을 분석·제시하였다. 결측구간은 결측정보지시행렬을 이용하여 전체적인 결측양상을 파악할 수 있도록 도시하였으며, 시간적·공간적인 결측비율도 분석하여 제시하였다. 전반적으로 조위의 결측비율은 낮은 수준이나, 결측이 특정 조위관측소에 집중되는 경향을 보이고 있다. 또한 연속적인 결측자료 발생간격에 대한 자기상관함수를 분석한 결과, 조위자료의 결측은 무작위적으로 발생하고 있는 것으로는 파악되었다.

핵심용어 : 조위자료, 결측자료, 결측구간, 결측양상, 결측비율

Abstracts : The missing data patterns of tidal elevation data in Korean coasts are analysed and provided. The missing interval of the data is displayed for all stations using the missing data indicator matrix in order to identify the overall missing pattern. The spatial and temporal missing rates are also estimated. The total missing rate of tidal elevation data is low. However, most of the missing is mainly derived from just 1 or 2 specific stations. The autocorrelation function of the consecutive missing interval data also shows that the missing interval occurs randomly.

Keywords : Tidal elevation data, Missing data, Missing interval, Missing data pattern, Missing rate

1. 서 론

조위는 평균해수면의 변동양상 분석, 해안구조물의 설계조위 결정, 연안의 침수범위 분석 등에 이용되는 매우 필수적인 자료이다. 특히 해안 및 항만구조물의 신뢰성 설계, 즉 설계인자에 대한 분포함수 추정에 관한 연구가 중요한 사안으로 부각되고 있는 시점에서 조위자료의 빈도분포 함수에 대한 연구와 더불어 비조석 성분의 빈도분포 함수에 대한 연구도 설계조위 결정에 매우 중요한 요소이다(星谷·石井, 1986). 일반적으로, 관측된 조위자료의 빈도분포함수 형태는 쌍봉형 또는 단봉형의 형태를 보이게 되며(Pugh, 1987), 조 등(2004, 2006)은 쌍봉형 정규분포에 의하여 한반도 근해 조위자료와 비조석 성분의 분포를 표현할 수 있음을 제시하였다. 그러나 이러한 연구 결과는 결측 조위자료를 고려하지 않았으며, 결측 자료를 포함하여 분석하는 경우, 분포형의 인자가 변화될 수 있다.

우리나라의 조위는 국립해양조사원에서 관리하는 연안 조위관측소에서 관측되고 있으며, 관측된 자료는 품질검정을 수행한 후 1시간 간격자료의 형태로 제공되고 있다(국립해양조사원, 2011). 그러나 대부분의 장기간 현장 해양관측자료와 같이

조위 관측자료도 결측이 발생하고 있으며, 그 결측구간의 자료를 다시 관측하여 보충한다는 것은 이미 경과된 시간이기 때문에 불가능하다. 따라서 조위자료를 이용하는 경우, 결측자료에 대한 검토는 기본적으로 필수적인 단계이다. 조위자료의 결측정보는 이미 결정되어 있기 때문에, 우리나라 전체의 조위자료에 대한 결측정보를 정리·제시하면 조위자료를 이용하는 연구자가 매번 결측정보를 재검토하는 시간과 결측정보 검토과정에서의 오류를 저감할 수 있다. 본 논문은 결측자료 분석에서 기본적으로 이용되는 결측정보인자행렬(missing-data indicator matrix)의 도시를 통하여 우리나라 조위자료의 결측상황을 전체적으로 파악할 수 있도록 하였으며, 별도의 표에는 결측구간을 월별로 정리하여 제시하였다. 또한 지점별·관측시기별 결측비율도 추정하여 제시하였다.

2. 분석에 사용한 조위자료

우리나라 연안의 조위자료는 국립해양조사원 홈페이지(<http://www.khoa.go.kr/>) → 통합해양정보 → 조석예보 → 1시간자료에서 각 조위관측소별로 월별 1시간간격 조위자료를 누구나

*한국해양연구원 해양환경보전연구부 (Corresponding author, Marine Environment & Conservation Research Department, KORDI, Ansan PO Box 29, Seoul, 425-600, Korea, hycho@kordi.re.kr)

**원광대학교 토목환경공학과, 원광대학교 부설 공업기술개발연구소 연구위원(Department of Civil and Environmental Engineering, Wonkwang University, 344-2, Shinyong-dong, Iksan, 570-749, Korea)

측구간은 회색으로 표현)을 부여하면 조위 자료의 전체적인 결측양상을 시각적으로 파악할 수 있는 그림을 얻을 수 있다 (Fig. 2). 한편, 시각적인 정보와 더불어 각 조위관측소별로 구체적인 결측구간 정보(월 규모의 결측구간)는 별도로 정리하여 Table 1에 제시하였다.

3.2 조위자료의 시간적·공간적인 결측비율(missing rate)

어떤 자료의 결측비율은 결측구간이 자료의 통계분석 등에 영향을 미치는 정도를 판별하는 기준이 되기 때문에 매우 중요한 정보이다(Hair Jr. et al., 2010). 본 논문에서는 각 조위관측소별로 전 관측기간에 대하여 공간적인 결측비율을 계산하였으며, 관측

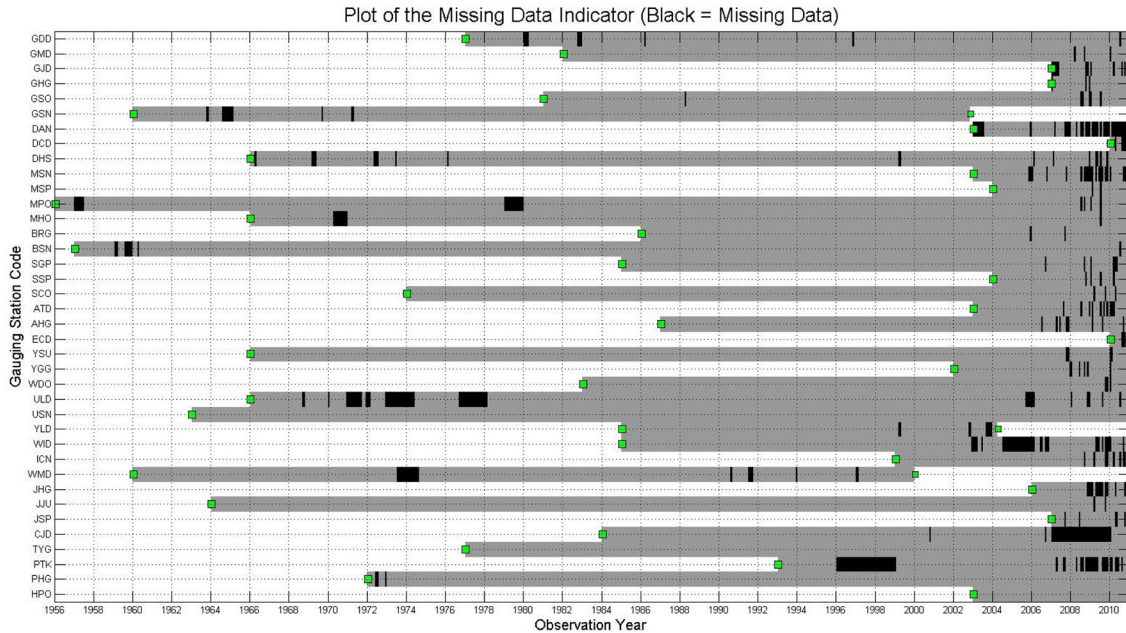


Fig. 2. Missing interval plot of tidal elevation data (Green Box = Start Year of the Tidal Gauging, Station Code = Table 1).

Table 1. Missing interval information of tidal elevation data (Month scale)

No.	Stations	Periods (Total [Year])	Missing period : Year(month)	Missing rate (%)
1	Gadeokdo (GDD)	1977.01-2010.12 (34years)	1980(01-03), 1982(10-12), 1986(03), 1996(11), 2010(07)	2.21
2	Geomundo (GMD)	1982.01-2010.12 (29years)	2008(03,09), 2010(01)	0.86
3	Geojedo (Gujora) (GJD)	2007.01-2010.12 (4years)	2007(01-05), 2008(10,11), 2009(01), 2010(03,08,10)	22.92
4	Goheung (Balpo) (GHG)	2007.01-2010.12 (4years)	2007(01), 2008(10,12)	6.25
5	Gunsan (Outer Port) (GSO)	1981.01-2010.12 (30years)	1988(04), 2008(07,08,12), 2009(01,07)	1.67
6	Gunsan (GSN)	1960.01-2002.10 (43years)	1963(10, 11), 1964(08-12), 1965(01,02), 1969(09), 1971(03,04)	2.33
7	Daesan (DAN)	2003.01-2010.12 (8years)	2003(01-07), 2005(12), 2007(03,09-12), 2008(04,07,08,10-12), 2009(02-05,07,09-11), 2010(02-12)	40.63
8	Daecheongdo (DCD)	2010.01-2010.12 (1years)	2010(04,08-12)	50.00
9	Daeheuksando (DHS)	1966.02-2010.12 (45years)	1966(04), 1969(03-05), 1972(05-07), 1973(06), 1976(02), 1999(03,04), 2006(02), 2007(02), 2008(12), 2009(04,05,07,11)	3.33
10	Masan (MSN)	2003.01-2010.12 (8years)	2005(11,12), 2006(01,10), 2007(10), 2008(07,09-12), 2009(01,02,04,06-08,10,11), 2010(01,09-12)	23.96
11	Moseulpo (MSP)	2004.01-2010.12 (7years)	2009(02,07)	2.38
12	Mokpo (MPO)	1956.01-2010.12 (55years)	1957(01-06), 1979(01-12), 2008(07,09), 2009(01,07)	3.33
13	Mukho (MHO)	1966.01-2010.12 (45years)	1970(04-12), 2009(07)	1.85
14	Boryeong (BRG)	1986.01-2010.12 (25years)	2005(12), 2007(09)	0.67
15	Busan (BSN)	1957.01-2010.12 (54years)	1959(02,03,08-12), 1960(04), 2010(07)	1.39
16	Seogwipo (SGP)	1985.01-2010.12 (26years)	2006(09), 2008(09), 2009(01), 2010(03-05)	1.92
17	Seongsanpo (SSP)	2004.01-2010.12 (7years)	2008(10), 2009(01,07), 2010(03)	4.76
18	Sokcho (SCO)	1974.01-2010.12 (37years)	2009(03,10), 2010(04)	0.68

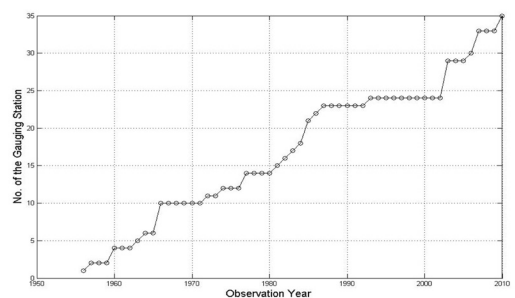
Table 1. (Continued)

No.	Stations	Periods (Total [Year])	Missing period : Year(month)	Missing rate (%)
19	Ansan (Tando) (ATD)	2003.01-2010.12 (8years)	2007(08), 2008(07,12), 2009(02,07,09,11), 2010(01-03)	10.42
20	Anheung (AHG)	1987.01-2010.12 (24years)	2006(07), 2007(04,06,10,11), 2009(02,08), 2010(09)	2.78
21	Eocheongdo (ECD)	2010.01-2010.12 (1years)	2010(08-10,12)	30.33
22	Yeosu (YSU)	1966.01-2010.12 (45years)	2007(10,11), 2010(01,02)	0.74
23	Yeonggwang (YGG)	2002.01-2010.12 (9years)	2007(12), 2008(01,06,09,11), 2010(01)	5.56
24	Wando (WDO)	1983.01-2010.12 (28years)	2009(10,11), 2010(01,12)	1.19
25	Ulleungdo (ULD)	1966.01-2010.12 (45years)	1968(09,10), 1970(01,12), 1971(01-09,12), 1972(01,02,12), 1973(01-12), 1974(01-05), 1976(09-12), 1977(01-12), 1978(01,02), 2005(09-12), 2006(01,02), 2008(01,11,12), 2009(08), 2010(07)	11.67
26	Ulsan (USN)	1963.01-2010.12 (48years)	-	0.00
27	Yuldo (YLD)	1998.01-2004.03 (20years)	1999(03-04), 2002(10-11), 2003(09-12)	3.46
28	Wido (WID)	1985.01-2010.12 (26years)	2002(12), 2003(01-03,06), 2004(07-12), 2005(01-12), 2006(01,02,06,07,09-11), 2009(04-06,08,10-12), 2010(01,09)	12.50
29	Incheon (ICN)	1999.01-2010.12 (12years)	2008(09), 2009(03,10,11), 2010(03,07,09,10)	5.56
30	Incheon (wolmido) (WMD)	1960.01-1999.12 (40years)	1973(07-12), 1974(01-08), 1990(08), 1991(07-09), 1993(12), 1997(01,02)	4.38
31	Janghang (JHG)	2006.01-2010.12 (5years)	2008(11,12), 2009(01,02,04-08,10,11), 2010(04,10-12)	25.00
32	Jeju (JJU)	1964.01-2010.12 (47years)	2009(03,10)	0.35
33	Jindo (Soopoom) (JSP)	2007.01-2010.12 (4years)	2007(09), 2008(06), 2010(04,05,10,12)	12.50
34	Chujado (CJD)	1984.01-2010.12 (27years)	2000(10), 2006(09), 2007(01-12), 2008(01-12), 2009(01-12), 2010(01,12)	12.35
35	Tongyeong (TYG)	1977.01-2010.12 (34years)	-	0.00
36	Pyeongtaek (PTK)	1993.01-2010.12 (18years)	1996(01-12), 1997(01-12), 1998(01-12), 1999(01), 2007(04,08,09), 2008(04,07,08,10-12), 2009(01-05,08-11), 2010(01,02,04-06,08, 12)	28.70
37	Pohang (PHG)	1972.01-2010.12 (39years)	1972(06,07,12)	0.64
38	Hupo (HPO)	2003.01-2010.12 (8years)	-	0.00

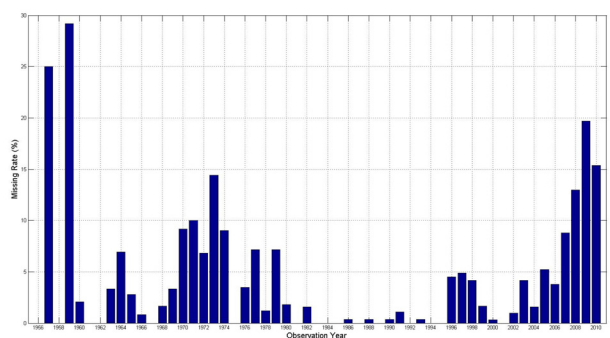
시기에 따른 시간적인 결측비율도 계산하여 그림과 표로 제시하였다(Fig. 3, Table 2). 각 조위관측소의 결측비율은 관측개시 이후의 전 기간에 대한 비율이며, 관측시기의 결측비율은 관측시기에 가용한 모든 조위관측소를 대상으로 추정된 결측비율이다. 결측비율은 전체적으로 높은 수준은 아니나 특정 조위관측소에 집중되는 양상을 보이고 있다. 조위자료의 결측은 다양한 이유보다는 관측기기의 오작동(malfunction)에 의한 영향이 대부분을 차지하고 있다. 따라서 결측구간을 저장하기 위해서는 기본적으로 정기적인 기기점검 및 유지관리가 가장 중요하다고 할 수 있다.

3.3 조위자료 결측시기의 양상분석

조위자료의 결측시기가 어떤 변화양상을 보이고 있는가를 분석하는 것은 조위자료의 보충(filling-in, imputation) 관점에서는 중요한 내용이다. 본 논문에서는 결측구간이 빈번한 지점을 대상으로 결측시기 자료의 간격에 대하여 자기상관분석을 수행하였다. 분석에 사용된 지점은 대산, 목포, 울릉도, 위도, 추자도, 평택지점으로 각각 39회, 32회, 65회, 39회, 40회, 62회의 결측구간이 발생하였다. 결측구간이 상대적으로 적은 지점은 통계적으로 소표본(small sample)으로 간주하여 오차범위가 크게 추정되기 때문에 분석에서 제외하였다(김 등, 2000). 연



(a) No. of gauging station



(b) Missing rate

Fig. 3. Temporal change of the gauging station number and missing rate.

Table 2. Temporal variation pattern on the station numbers and missing rate of the tidal elevation data

Year	Station no.	Missing rate (%)	Main Missing stations	Year	Station no.	Missing rate (%)	Main Missing stations
1956	1	0.00		1984	18	0.00	
1957	2	25.00	Mokpo	1985	21	0.00	
1958	2	0.00		1986	22	0.38	
1959	2	29.17	Busan	1987	23	0.00	
1960	4	2.08		1988	23	0.36	
1961	4	0.00		1989	23	0.00	
1962	4	0.00		1990	23	0.36	
1963	5	3.33	Gunsan	1991	23	1.09	
1964	6	6.94	Gunsan	1992	23	0.00	
1965	6	2.78		1993	24	0.35	
1966	10	0.83		1994	24	0.00	
1967	10	0.00		1995	24	0.00	
1968	10	1.67		1996	24	4.51	Pyeongtaek
1969	10	3.33		1997	24	4.86	Pyeongtaek
1970	10	9.17	Mukho	1998	24	4.17	Pyeongtaek
1971	10	10.00	Ulleungdo	1999	24	1.67	
1972	11	6.82		2000	24	0.34	
1973	11	14.40	Ulleungdo	2001	24	0.00	
1974	12	9.03	Ulleungdo	2002	24	0.96	
1975	12	0.00		2003	29	4.17	Daesan
1976	12	3.47		2004	29	1.56	
1977	14	7.14	Ulleungdo	2005	29	5.21	Wido
1978	14	1.19		2006	30	3.79	
1979	14	7.14	Mokpo	2007	33	8.80	Chujado
1980	14	1.79		2008	33	12.96	Chujado
1981	15	0.00		2009	33	19.68	Chujado
1982	16	1.56		2010	35	15.35	
1983	17	0.00					

*Wolmido Stop(1999.12), Gunsan Stop(2002.10), Yuldo Stop(2004.3)

속자료의 자기상관효과 분석은 조위자료의 결측이 발생하는 시기의 간격자료를 이용하여 수행하였다. 자기상관함수를 추정된 결과, 자기상관함수의 변화양상은 모든 지점이 매우 유사한 양상을 보였으며, 상관계수가 오차범위에서 변동하고 있는 것으로 파악되었다(Fig. 4). 따라서 조위자료의 결측은 무작위적으로 발생하는 것으로 판단된다.

4. 결론 및 제언

본 논문에서 결측자료 인자행렬을 이용하여 제시하는 조위자료의 결측양상 정보는 국립해양조사원에서 제공하는 조위자료를 이용하여 계산·도시되는 객관적인 정보이기 때문에 조위자료

를 이용하여 다양한 분석·연구를 수행하는 연구자에게 자료의 결측구간을 파악하거나 결측구간 파악과정에서의 오류를 절감하여 주는 정보로 활용가능하다. 한편, 결측비율은 전체적으로 높은 수준은 아니나 특정 조위관측소에 집중되는 양상을 보이고 있다. 결측자료 발생간격에 대한 자기상관함수를 분석한 결과, 조위자료의 결측은 무작위적으로 발생하고 있는 것으로는 파악되었다.

향후, 결측구간의 적절한 자료보충(imputation, data-filling)을 통하여 빈틈이 없는 완전한 자료세트(complete data set)가 구성되는 경우, 기존의 불완전한 자료세트(incomplete data set)로는 적용이 곤란하였던 다양한 통계분석 도구를 활용하여 보다 다양한 정보 분석이 가능할 것으로 판단된다.

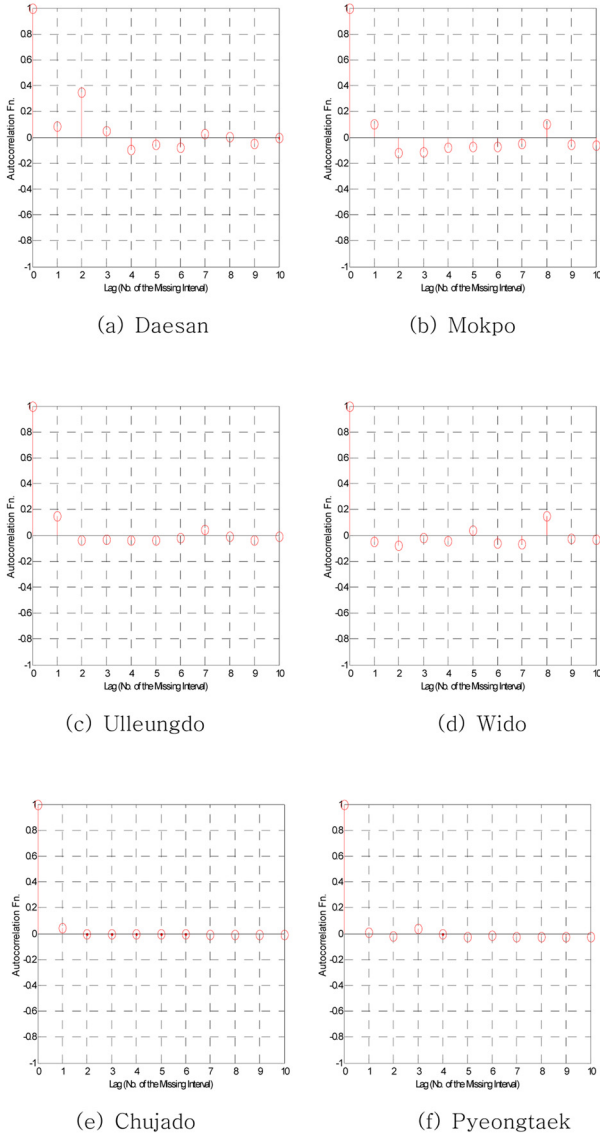


Fig. 4. Autocorrelation function of the consecutive missing interval data.

감사의 글

본 논문은 2010학년도 원광대학교의 교비지원으로 수행되었습니다. 연구비 지원에 감사드립니다. 또한 본 논문에서 사용한 조위자료를 제공해 주신 국립해양조사원에 감사드립니다.

참고문헌

국립해양조사원 (2011). <http://www.khoa.go.kr>.
 김우철, 김재주, 박병욱, 박성현, 박태성, 송문섭, 이상열, 이영조, 전종우, 조신섭 (2000). 통계학 개론, 제4개정판, 영지문화사.
 조홍연, 정신택, 오영민 (2004). 조위자료의 확률밀도함수 추정, 한국해양·해양공학회지, 16(3), 152-161.
 조홍연, 정신택, 윤중태, 김창일 (2006). 한반도 연안 비조석 성분자료의 통계적 특성, 한국해양·해양공학회지, 18(2), 112-123.
 星谷 勝, 石井 清 (1986). 構造物の信頼性設計法, 鹿島出版會.
 Hair Jr. J.F., Black, W.J., Babin, B.J., and Anderson, R.E. (2010). *Multivariate Data Analysis, A Global Perspective*, Seventh Edition, Pearson Education.
 Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, Second Edition, John Wiley & Sons.
 Pugh, D.T., (1987). *Tides, Surges and Mean Sea-Level, A Handbook of Engineers and Scientists*, John-Wiley & Sons.
 Allison, P.D. (2002). *Missing Data*, Sage Publications, Series/Number 07-136.

원고접수일: 2011년 10월 4일
 수정본채택: 2011년 11월 3일(1차)
 2011년 11월 22일(2차)
 게재확정일: 2011년 11월 22일