

Fragment 모형을 이용한 월 증발량 모의 발생에 관한 연구

김선영⁽¹⁾, 김병식⁽²⁾, 김형수⁽³⁾, 서병하⁽⁴⁾

1. 서론

물 순화과정의 구성요소 중 하나인 증발 및 증산과정은 저수지의 계획, 운영면에서 대단한 중요성을 가진다. 즉, 수자원을 개발하고자 할 때 순수하게 이용가능한 양을 추정하기 위해서는 증발량 산정이 필요하며, 이는 단기간보다는 비교적 장기간 동안의 자료가 필요하다(윤용남 1997). 그러나 현실적으로 이용 가능한 실측 증발량자료는 매우 짧은 편이다. 그러므로 미래의 발생가능성이 있는 증발량을 잘 이해하기 위해서는 추계학적 모의기법이 필요하다. 본 연구에서는 연/월 증발량 시계열자료를 모의발생하기 위해 청주기상대의 24년 기간의 관측자료와 대구기상대의 31년 동안의 월 자료를 이용하였다. 또한 자료의 통계적 특성을 파악하고자 동질성 검사와 정규화 검사를 실시하였다. 연 증발량 모의 발생을 위해 Monte Carlo, 자기회귀모형인 AR(1) 모형에 Bootstrap의 일종인 Balanced Bootstrap 방법을 적용시켰다. 월 증발량 모의에는 분해모형의 일종인 "Fragment 방법"을 이용하였다. 모의값의 불확실성을 고려해서 연/월 100회 Monte-Carlo 시험을 하였으며, 모형의 검정을 위해서 모의 발생된 증발량 자료와 관측증발량의 평균, 표준편차, 왜곡도, 상관계수를 비교하였다.

2. 모형의 구성

2.1 연 모형

본 연구에서는 난수발생에 있어 Bootstrap의 일종인 Balanced Bootstrap을 적용하였다. 가장 간단한 AR(1)모형을 예로 들면,

$$e_j = y_j - \hat{\alpha}y_{j-1}, \quad j=2, \dots, n \quad (1)$$

Bootstrap을 적용하기 앞서 잔차를 중심화 시킨다.

$$\epsilon_2, \dots, \epsilon_n = e_2 - \bar{e}, \dots, e_n - \bar{e}$$

중심화 된 잔차를 복원추출 한다. (여기서 Balanced Bootstrap은 R개의 Bootstrap집합을 하나의 집합으로 묶고 다시 재 배열 시킨다. 재 배열된 자료를 N개의 요소로 분해한다.)

- (1) 인하대학교 토목공학과 석사과정
- (2) 인하대학교 토목공학과 박사과정
- (3) 선문대학교 토목공학과 교수·공학박사
- (4) 인하대학교 토목공학과 교수·공학박사

$$y_j^* = \hat{a}y_{j-1} + \epsilon_j^*, \quad j=1, \dots, n \quad (2)$$

여기서, y_j^* 는 Bootstrap를 통해 복원 추출된 자료

\hat{a} 는 1차 자기상관계수, ϵ_j 는 관측치의 잔차

ϵ^* 는 Bootstrap를 통해 복원 추출된 잔차

2.2 분해모형 (Fragment 모형)

분해모형은 Lower-level(e.g. monthly)의 변량값을 higher-level(e.g. annual)의 변량값으로 나누는 기법이다. 이 분해모형은 두 가지의 장점을 가지고 있다. 첫째, 모의된 자료의 통계적 특성을 잘 재현할 수 있고, 매개변수의 수를 줄일 수 있다. 둘째, 모의발생을 위해 사용되는 방법에 적용하기가 쉽다.(Salas et al,1980) 이 중에서 간단한 분해방법을 "Fragment 방법"이라 부른다.

$$\text{Fragment}_{(i,j)} = \frac{E_{i,j}(\text{관측치의 월 자료})}{E_j(\text{관측치의 년 자료})} \quad (3)$$

$$M = \text{INTEGER}(N \cdot U + 1)$$

$$E_{i,j}(\text{모의월자료}) = \text{Fragment}_{M,j} \times E_i(\text{모의년자료}) \quad (4)$$

여기서 $I = 1, 2, 3, \dots, n$ 연, $j = 1, 2, 3, \dots, 12$ 월

3. 모형의 검증

모의된 연/월 증발량 자료가 관측된 자료의 기본 통계치를 잘 재현하는지를 검사하기 위해서 관측된 기본 통계치의 신뢰구간을 산정하였으며, 본 연구에서는 Bootstrap 방법을 적용하였다. Bootstrap 방법은 기존의 매개변수적 추정방법과는 달리 모집단의 정규분포라는 가정이 필요 없으며 원 자료의 분포의 편기성을 잘 재현할 수 있다.(전명식, 1997)

$$\hat{\theta}_\alpha = 2t - t_{((R+1)(1-\alpha))}^* \quad \hat{\theta}_{1-\alpha} = 2t - t_{((R+1)(\alpha))}^* \quad (5)$$

여기서 t^* = bootstrap에 복원 추출된 자료의 통계치

R = 반복횟수

α = 신뢰구간 95%의 유의수준

4. 대상지점의 적용

4.1 분석자료

본 연구에서 사용된 자료는 자료의 기간과 신뢰성을 고려해서 청주기상대의 1967. 1월-1990. 12월 기간인 24년치의 월 증발량 자료와, 대구기상대의 1960. 1월-1991.12월 기간인 31년동안의 월 증발량 자료를 이용하였다.

4.2 연/월 증발량 자료(mm/day)의 기본 통계치 분석

년/월 증발량 자료의 기본 통계치를 분석한 결과 5월-8월 사이에 가장 많은 증발량이 발생했으며, 왜곡도 계수 값에서 알 수 있듯이 대체로 정규분포를 따르고 있다.

표1. 원 자료의 기본 통계치

지점	통계치	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월	연
청주	평균	33.40	42.58	76.57	114.96	144.33	142.03	125.58	129.36	99.16	77.23	43.25	33.00	1061.46
	표준편차	4.68	6.81	8.47	13.59	17.29	22.66	20.71	16.87	12.43	8.11	4.50	4.61	69.88
	왜곡도	-0.56	0.18	1.13	0.28	1.24	0.02	0.49	-0.36	-0.56	0.38	0.33	-0.05	0.05
	lag-1	0.073	-0.176	-0.378	-0.268	-0.071	0.202	0.008	0.155	0.107	-0.115	0.245	-0.32	0.12
대구	평균	57.06	63.69	103.64	135.36	172.24	163.89	151.21	162.89	113	95.99	64.40	54.62	1337.97
	표준편차	10.91	11.15	15.78	17.48	21.90	29.36	31.67	33.88	16.62	11.67	8.75	9.03	111.48
	왜곡도	0.12	-0.12	0.59	-0.47	0.41	0.11	1.68	0.51	0.77	0.09	0.36	0.40	0.76
	lag-1	0.455	0.247	0.155	-0.017	-0.18	0.03	-0.155	0.017	0.002	-0.282	0.074	0.287	0.175

4.3 연 증발량 자료의 등질성 검정, 정규성 검정

수문자료를 해석하거나 모형화할 때 주어진 자료의 평균과 분산값이 정상적이라고 가정하는 경우가 많다. 그러나 주어진 자료에 변동이나 경향이 나타나는 경우에는 이러한 특성을 파악하고 해석하든지, 아니면 이러한 특성을 제거한 뒤 해석해야 한다. 본 연구에서는 Screening분석을 통해서 자료의 등질성과, 정규성이 인정되었다.

표4.2 원 자료의 등질성 및 정규성 Test

지 점	Mann-Whitney Test			simple T Test			Skewness Test		
	$ z $	$u_{1-\alpha/2}$	Test	T_c	$T_{1-\alpha}$	Test	g	$g_{1-\alpha}$	Test
청 주	0.577	1.96	○	0.494	1.96	○	0.0499	0.98	○
대 구	1.502	1.96	○	1.797	1.96	○	0.7641	0.8623	○

○ : cannot be rejected at a level of significance of 5%

4.4 연/월 증발량 자료의 모의발생

본 연구에서 사용된 시계열 모형에 Bootstrap을 적용시킨 결과 기존의 시계열 모형이 가지고 있던 단점인 난수분포의 가정을 해결할 수 있었으며, BMCS모형과 BAR(1) 모형의 통계치를 분석한 결과 BMCS모형이 더 타당하였다.

표3. Bootstrap + Monte Carlo모형(BMCS) 과 Bootstrap + AR(1)모형(BAR)에 의한 통계치

지 점	청 주(100회 평균치)		대 구(100회 평균치)	
	BMCS	BAR(1)	BMCS	BAR(1)
평 균	1060.994	1062.052	1338.553	1337.895
표준편차	68.659	65.896	109.724	104.911
왜 곡 도	0.052	0.13	0.712	0.957
lag-1	0.002	0.005	0.003	0.006

Fragment 방법은 최소한의 매개변수와 가장 간단한 방법으로 월 모의값을 발생시킨다. 표4와 그림 1~8에서 알수 있듯이 관측치의 월 통계치의 특성들을 잘 재현하였다. 특히 기존의 시계열 모형은 왜곡도 값에 있어서 모의기간이 길어질수록 0으로 수렴하지만(김병식, 2002) Fragment 모형은 월 왜곡도의 특성과 월 주기성을 잘 재현하였다.

표4. Fragment+BMCS모형(BMCS)에 의한 모의된 기본 통계치(100회 평균치)

지점	통계치	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
청주	평균	33.34	42.57	76.60	115.07	144.94	141.48	125.21	129.22	99.06	77.29	43.28	33.10
	표준편차	4.324	5.705	9.154	13.715	18.596	20.428	19.218	16.594	12.13	9.327	5.314	5.009
	왜곡도	-0.31	0.342	0.692	0.471	0.555	0.086	0.345	0.347	-0.07	0.458	0.114	-0.01
대구	평균	57.41	63.85	103.57	136.28	172.77	163.55	150.22	161.64	113.22	96.45	64.48	54.57
	표준편차	11.31	11.24	15.59	21.42	22.73	26.08	26.06	30.06	17.21	14.96	9.93	8.69
	왜곡도	0.04	0.46	0.73	-0.06	0.38	0.03	0.92	0.16	0.40	0.33	0.68	0.48

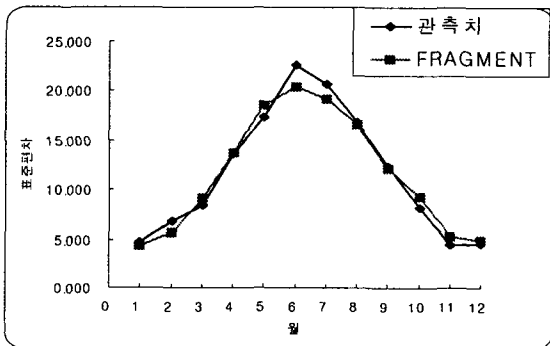
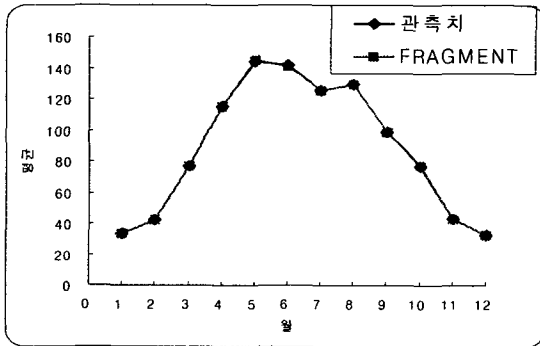


그림 1 Fragment 모형과 관측치의 평균비교(청주) 그림 2 Fragment 모형과 관측치의 표준편차비교(청주)

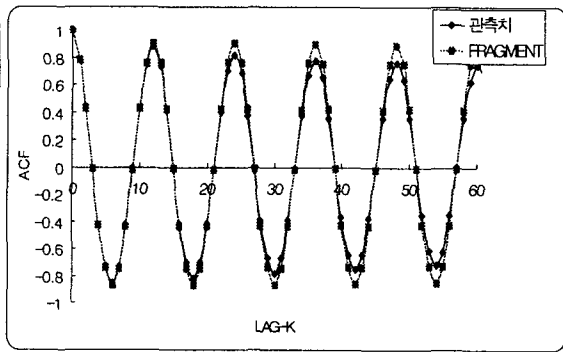
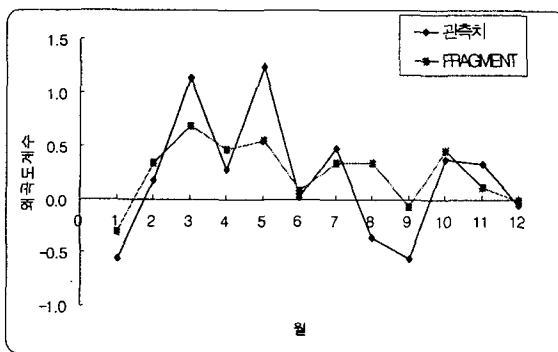


그림 3 Fragment 모형과 관측치의 왜곡도계수 비교(청주) 그림 4 Fragment 모형과 관측치의 자기상관계수 비교(청주)

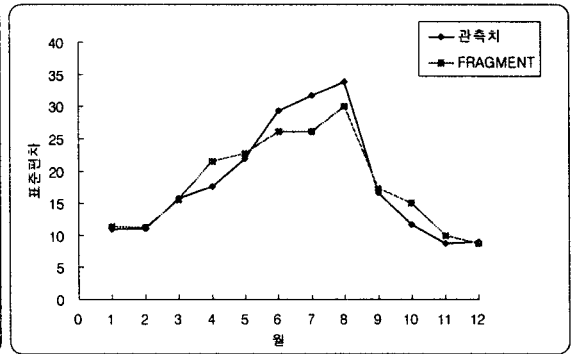
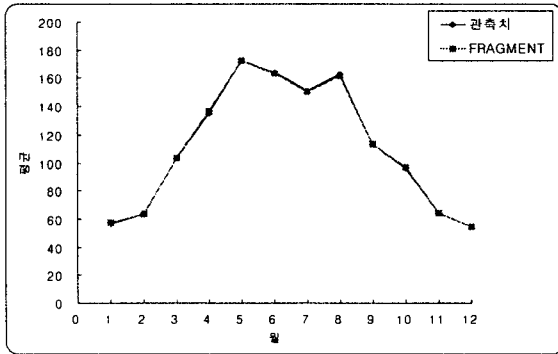


그림 5 Fragment 모형과 관측치의 평균비교(대구) 그림 6 Fragment 모형과 관측치의 표준편차비교(대구)

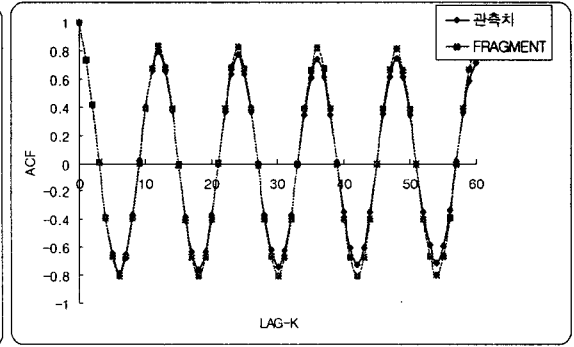
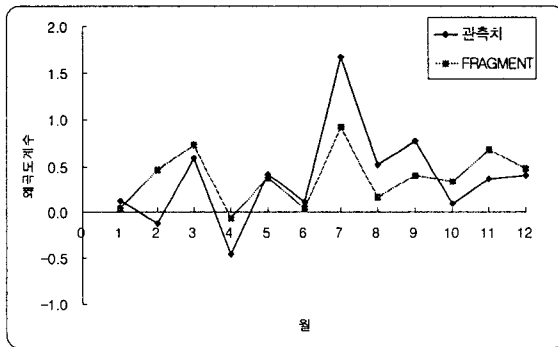


그림 7 Fragment 모형과 관측치의 왜곡도계수 비교(대구) 그림 8 Fragment 모형과 관측치의 자기상관계수 비교(대구)

4.5 연/월 모형의 검정

모형의 검증을 위해 관측자료의 신뢰수준95%에 대한 통계치의 신뢰구간이 산정하였다. 연 모의 값에 의해서 구해진 모든 통계치들이 신뢰구간에 만족하였으며, 월 모의값에 있어서는 2월, 7월의 왜곡도 값이 신뢰구간에서 다소 벗어났으나, 대체로 만족스러운 결과 값을 보여주었다. 따라서 본 연구에서 사용된 모형의 타당성을 입증하였다.

표5. Bootstrap 방법에 의해 산정된 연 자료의 신뢰구간

	청 주		대 구	
	하한계	상한계	하한계	상한계
신뢰구간				
평균	1036.279	1092.925	1298.655	1377.803
표준편차	58.353	87.717	89.304	146.677
왜곡도	-0.767	0.649	0.15	1.475
lag-1	-0.018	0.559	-0.003	0.684

표6. Bootstrap 방법에 의해 산정된 월 자료의 신뢰구간

지점	청 주						대 구					
	평 균		표 준 편 차		왜 곡 도		평 균		표 준 편 차		왜 곡 도	
신뢰구간	하한계	상한계	하한계	상한계	하한계	상한계	하한계	상한계	하한계	상한계	하한계	상한계
1월	31.57	35.4	3.96	5.77	-1.2	0.069	52.87	60.11	8.8	13.63	-0.49	0.77
2월	40.06	45.12	5.44	8.66	-0.641	1.0	59.86	67.68	8.86	14.01	-0.68	0.38
3월	73.38	79.28	5.30	12.28	0.13	3.2	98.09	109.83	12.65	19.96	-0.02	1.18
4월	109.54	120.20	11.55	17.43	-0.57	0.86	129.93	140.77	13.57	21.52	-1.17	0.11
5월	137.65	150.64	11.82	25.84	0.21	3.28	164.38	179.91	17.39	27.68	-0.27	1.03
6월	134.06	152.38	19.57	27.64	-0.67	0.5	153.85	175.15	24.11	37.27	-0.5	0.80
7월	117.53	134.50	15.87	28.04	-0.2	1.29	140.81	159.47	20.57	45.2	0.93	3.67
8월	122.24	139.79	13.56	21.34	-1.03	0.45	151.45	173.45	26.31	46.07	-0.59	1.73
9월	94.35	104.21	9.45	16.10	-1.43	0.3	108.03	118.55	12.45	21.50	0.16	1.59
10월	72.74	80.4	6.55	10.60	-0.24	1.09	92.46	100.96	9.67	13.89	-0.4	0.57
11월	41.64	44.82	3.54	5.95	-0.38	1.42	61.33	66.98	7.32	10.96	-0.13	0.97
12월	31.39	34.7	3.94	5.79	-0.63	0.51	51.85	57.62	7.79	11.25	-0.29	0.85

5. 결론

본 연구에서는 연/월 증발량 자료를 확충하기 위해서 연 모형에 있어서는 기존의 시계열 모형에 Bootstrap 방법을 적용하였고, 월 모형에 있어서는 분해모형의 일종인 “Fragment 방법”을 이용하였다. 각 모형에 대해 100회 반복모의하여 각 통계치의 평균값에 대한 신뢰구간 검정을 실시하여 다음과 같은 결론을 얻었다.

- (1)본 연구에서 사용된 자료는 Screening 분석을 통해 자료의 동질성과, 정규성이 인정되었다.
- (2)난수의 분포형을 가정하여 발생시키는 MCS, AR(1) 모형에 Bootstrap 방법을 적용함으로써 난수분포에 대한 가정을 배제하였고, 원 자료의 기본통계치 특성들을 잘 재현할수 있었다.
- (3)Fragment 방법을 평가하기 위해 매월 통계치들에 대한 신뢰구간 검정을 실시한 결과 2,7월의 왜곡도 계수 만 제외하고는 모두 만족함으로써 최소한의 매개변수와 간단한 계산과정인 Fragment 모형은 월 시계열의 모의발생에 적합성이 인정되었다.
- (4)Bootstrap을 이용하여 각 통계치에 대한 신뢰구간을 산정한 결과 기존의 신뢰구간 산정방법보다 폭이 더 좁은 신뢰구간이 산정되었다. 이는 원 자료의 분포와 편기성을 고려할 때 더 합리적이다.

참고문헌

1. 윤용남(1997, 공업수문학.청문각, pp.93-113
2. Michael, R. C.(1999), *Bootstrap Methods; A Practitioner’s Guide*.
3. 김병식(2002),Bootstrap 방법을 이용한 하천유출량 자료의 추계학적 확충, 수자원학회지(수록예정)
4. Dr.Abdulmohsen A. Al-Shaikh(2000), Evaporation data stochastic generation for king fahad dam lake in bishah, saudi arabia, ICH2000