

엔트로피를 이용한 흐름분배 알고리즘 해석

이 학 수*, 김 남 원**, 강 창 용***, 김 상 현****

1. 서 론

흐름분배 알고리즘은 복잡한 지형굴곡을 따른 흐름분산 특성을 잘 기술하도록 발전되어 왔는데, 이는 단방향 흐름분배 알고리즘(O'Callaghan과 Mark, 1984), DEMON 알고리즘(Costa-Cabral과 Burges, 1994), 다방향 흐름분배 알고리즘(Quinn 등, 1991), 구배역급수 알고리즘(Holmgren, 1994), SDFAA 알고리즘(김상현 등, 2002) 등의 다양한 알고리즘을 유도하였다. 다양한 흐름분배 알고리즘으로 계산된 습윤지수의 분포함수는 모형내 포화지역, 지표면 흐름, 지하수 흐름의 계산을 위한 기본정보일 뿐 아니라, 유역내 모든 지점의 수문과정을 기술해주는 지표이므로 토양수분의 구조성(organization)과 임의성(randomness)(Western, 1999)이 혼재하는 유역수문과정의 기술에 적합한 습윤지수 분포함수의 산정은 필수적인 과정이라 할 수 있다. 엔트로피는 습윤지수와 같은 이산 무작위 변량에 의한 정보량의 측정(Shannon과 Weaver, 1949)에 이용될 수 있으며, 최대정보량을 기술해주는 흐름분배 알고리즘의 선정에 활용될 수 있다. 엔트로피를 이용한 연구로는 TOPMODEL 구조내에서 습윤지수 분포함수 추정 정보이론을 도입한 Mendicino와 Sole(1997)의 연구와, 격자 해상도 변화에 대한 토양도와 토지이용도의 정보량 분석에 관한 Kuo 등(1999)의 연구가 있지만, 기존 흐름분배 알고리즘들의 비교분석에 대한 엔트로피 해석이 요구된다. 따라서, 본 연구에서는 다양한 흐름분배 알고리즘들로 계산한 습윤지수를 엔트로피 개념을 이용하여 해석하고, 실측된 지표토양수분자료와 최대 정보량을 가지는 습윤지수 분포함수와의 분석을 통한 기존 흐름분배 알고리즘들에 대한 해석을 목적으로 한다.

2. 방 법 론

2.1 흐름분배 알고리즘

김상현등(2002)의 Spatially Distributed Flow Apportioning Algorithm (SDFAA)은 Holmgren(1994)의 구배역급수(h)의 개념과 Quinn 등(1995)의 수로형성면적(Channel Initiation Threshold, CIT)의 개념을 이용하여 지형을 따른 흐름이 사면에서의 흐름분산 특성이 수로에서의 흐름수렴과정으로 변화하는 흐름특성을 기술할 수 있도록 구성한 알고리즘이다. 식 (1)은 SDFAA 알고리즘의 공간흐름 분배인자($H_{(i,j)}$)를 가변수로형성면적($A_{ECIT_{(i,j)}}$)에 따라 각 격자에 할당하는 식이다.

$$\frac{dH_{(i,j)}}{dA_{ECIT_{(i,j)}}} = cA_{ECIT_{(i,j)}}^n \quad (1)$$

여기서, $H_{(i,j)}$ 는 I, J 격자지점에서 가변수로형성면적에 따라 공간적으로 변화하는 구배역급수이며, $A_{ECIT_{(i,j)}}$ 는 I, J 격자지점의 가변수로형성면적으로서 단위격자 크기에서 수로형성면적(CIT)크기까지 변화한다. c 와 n 은 해당유역에 대해 보정되는 매개변수이다. 식 (1)에서 구한 $H_{(i,j)}$ 를 이용한 i 격자방향으로의 흐름분배율은 식 (2)로서 기술된다.

$$f_i = \frac{L_i(\tan \beta)^{H_{(i,j)}}}{\sum (L_i(\tan \beta)^{H_{(i,j)}})} \quad (2)$$

* 부산대학교 환경기술·산업개발연구센터 연구원

** 한국건설기술연구원 수자원환경연구부 선임연구원

*** 부산대학교 환경공학과 석사과정

**** 부산대학교 공과대학 환경공학과 조교수

2.2 엔트로피를 이용한 습윤지수 분포함수의 불확실성 해석

흐름분배 알고리즘에 의해 결정되는 유역의 습윤지수가 특정모집단을 가지며, 각 흐름분배 알고리즘에 의해 기술되는 습윤지수의 분포가 모집단에 대한 하나의 추정치들이라고 하면, 습윤지수가 가질수 있는 임의의 분포형에 대한 정보가 없는 경우 엔트로피($H(x)$)를 이용한 불확실성의 정량적 평가는 다음과 같다(Kottegoda와 Rosso, 1996).

$$H(x) = - \sum_{i=a}^b p_X(x_i) \ln p_X(x_i) \quad (3)$$

$$\sum_{i=a}^b p_X(x_i) = 1 \quad (4)$$

식 (3)은 이산무작위변량 X 의 확률(p_X)분포가 가지는 엔트로피($H(x)$)량으로서, 이는 X 가 가지고 있는 불확실성, 또는 정보량을 의미한다(Shannon과 Weaver, 1949). a , b 는 X 의 하한과 상한을 의미한다. 식 (3)과 식 (4)를 Lagrangian 함수($L(p, \lambda)$)의 형태로 표현하면 식 (5)로 표현된다. 식 (5)에서 λ 는 Lagrangian multiplier이며 $L(p, \lambda)$ 를 최대화 하는 확률분포함수가 최대정보량을 가지게 되며 이는 $dL(p, \lambda) / dp = 0$ 를 만족하는 것으로서 식 (6)로 기술된다.

$$L(p, \lambda) = - \sum_{i=a}^b p_X(x_i) \ln p_X(x_i) + \lambda \left(\sum_{i=a}^b p_X(x_i) - 1 \right) \quad (5)$$

$$p_X(x) = \exp(-1 + \lambda) \quad (6)$$

식 (6)을 식 (4)에 적용하면 식 (7)로서 표현되는 확률분포함수를 구할 수 있으며, 식 (8)은 식 (7)로서 구한 정보량을 나타낸다. 식 (8)은 습윤지수의 분포범위($b-a$)가 넓어질수록 정보량이 커짐을 의미하며, 이를 만족하는 습윤지수분포를 산출해 주는 흐름분배 알고리즘을 최적 알고리즘으로 표현하고 있다.

$$p_X(x) = \frac{1}{b-a} \quad (7)$$

$$H(x) = \ln(b-a) \quad (8)$$

식 (7)은 균등분포(Uniform Distribution)가 최대정보량을 가지는 확률분포함수임을 나타내며, 습윤지수의 분포가 균등분포이면 식 (3)과 식 (8)의 계산결과가 동일해야 함을 의미한다. 그러나, 실제 습윤지수의 분포형을 알지 못하므로, 식 (3)과 식 (8)의 계산 결과는 다르게 산출될 것이며, 최대한 균등분포에 접근하려는 분포형을 도출할 것으로 기대된다. 이는 구조성과 임의성이 동시에 존재하는 유역내에서 임의성을 최대한 반영하는 형태로 습윤지수 분포를 기술함으로 예상되는 결과라 판단한다(Western 등, 1999).

3. 연구대상지역

본 연구의 대상이 되는 유역은 강원도 삼척시 원덕읍 월천리로 0.286025 km^2 의 유역면적에 1%에서 87%의 경사도를 가지고 있는 산지유역이다. 연평균 기온은 연최고 24.6°C , 연최저 -4.7°C 이며, 연평균 1100 mm 의 강우량을 기록하는 월천리 유역은 유역의 출구가 해안과 접해있고, 사면유출흐름의 발달에 대한 관찰과 모의가 용이한 밀폐유역이다. 본 연구의 두 번째 대상유역으로써 95년부터 한국건설기술연구원에서 시험유역으로 운영하고 있는 임진강 하구에서 약 46 km 상류의 경기도 파주시 적성면 마지리와 설마리의 설마천 유역을 선정했다. 설마천 유역은 유역면적 8.5 km^2 , 유로연장 5.8 km , 연평균 강수량은 $1,600 \text{ mm}$, 고도는 최저 59 m 에서 최고 675 m , 유로경사 2%인 전형적인 급경사 산지 사행하천 지형이며 지질은 경기편마암 복합체로 구성되고 단층과 엽리가 잘 발달되어 있다(한국건설기술연구원, 1998). 그림 1은 대상유역의 수치고도모형을 나타낸다.

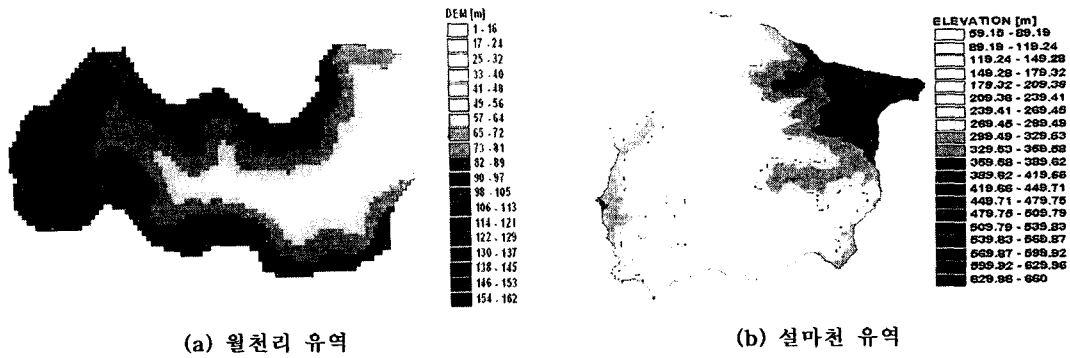
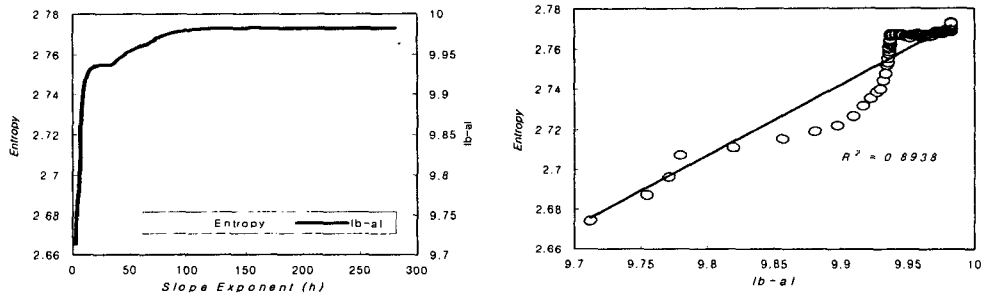


그림 1. 대상유역의 수치고도모형 : (a) 월천리 유역 (b) 설마천 유역

4. 결과 및 토의

그림 2는 구배역급수(h)에 따른 엔트로피의 경향성을 기술하고 있다. 엔트로피와 $|b-a|$ 는 절대도 0.8938의 양의 상관관계를 가지며, 이는 식 (3)와 식 (8)에 의해 구배역급수 알고리즘에 의한 습윤지수의 분포함수가 균등 분포에 근접함을 알 수 있다.

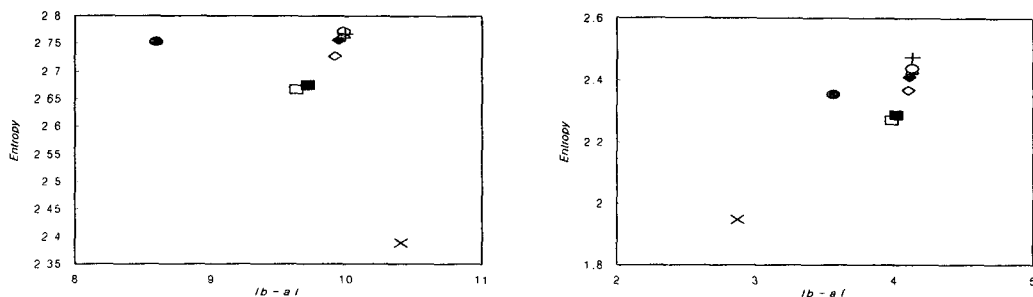


(a) 구배역급수(h) 변화에 따른 $|b-a|$ 와 Entropy의 분포

(b) 구배역급수(h) 변화에 따른 $|b-a|$ 와 Entropy의 상관도

그림 2. 구배역급수(h) 변화에 따른 엔트로피와 $|b-a|$ 의 분석

그림 3은 각 흐름 알고리즘에서 계산한 엔트로피와 $|b-a|$ 를 나타낸다. 그림 2에서 고찰한대로, 그림 3에서 구배역급수(h)의 증가에 따라 습윤지수의 엔트로피와 분포범위가 증가함을 알 수 있다. 큰 h 값은 *sfd*적 흐름을 유도하는데, 실제 *sfd*흐름 알고리즘에서는 h 알고리즘에 비해 엔트로피가 적게 산출되었으며, 또한 DEMON 알고리즘에서는 작은 엔트로피에 큰 분포범위를 가지고 있다. 이것은 유역내 혼재하는 임의성(randomness)과 구조성(organization)(Western 등, 1999)중에서 수로에 집중적으로 분포하는 구조적 특성에 따라 유역 흐름분배의 불확실성의 증가를 방해함으로 판단되며, 이에 따라 수로가 형성가능한, 즉, 구조적인 유역성분을 제외하고 분석해 보겠다.



(a) 전체유역면적

(b) 수로형성가능지역 제외한 유역면적

그림 3. 각 흐름분배알고리즘들로 구한 습윤지수의 엔트로피와 분포범위

□ *mfd*; ■ $h=2$; ◇ $h=10$; ◆ $h=20$; △ $h=100$; ○ $h=280$; ● *sfd*; × *DEMON*; + *SDFAA*($n=-50, H_{max}=250$)

5. 결론

본 연구는 엔트로피의 개념을 이용하여 흐름분배 알고리즘들에 의해 계산된 습윤지수 분포함수들의 해석을 시도한 것으로서, 본 연구로 도출할 수 있는 결론은 다음과 같다.

1. 구조성과 임의성이 혼재한 유역을 구조성만으로 기술하는 단방향 흐름알고리즘과 DEMON 알고리즘은 다른 알고리즘에 비해 수치고도모형에서 습윤지수 분포함수로의 낮은 정보전달을 나타내었으며, 수로형성가능지역은 유역의 임의성 해석시 제외하는 것이 바람직하다.

2. 구배역급수 알고리즘과 SDFAA 알고리즘은 적절한 매개변수의 선정으로 습윤지수 분포함수가 가지는 정보량을 증대시킬 수 있었다.

감사의 글

본 연구는 21세기 프론티어 연구개발사업인 수자원의 지속적 확보기술 개발사업단의 연구비지원(과제번호 2-1-1, 2-2-1)에 의해 수행되었습니다.

6. 참고문헌

1. 김상현, 이학수, 강창용, 김남원 (2002) "자연하천 형상을 이용한 최적 흐름분배 알고리즘의 개발." 한국수자원학회 논문집, 게재확정
2. 한국건설기술연구원 (1998). 시험유역의 운영 및 수문특성 조사·연구 연구보고서, 건기연 98-077, pp. 5-24
3. Costa-Cabral, M.C., and Burges, S.J. (1994). "Digital elevation model networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas." *Water Resour. Res.*, Vol. 30, pp. 1681-1692.
4. Holmgren, P. (1994). "Multiple flow direction algorithms for runoff modeling in grid based elevation models and empirical elevation." *Hydrol. Process.*, vol. 8, pp. 327-334.
5. Mendicino, G., and A. Sole (1997). "The information content theory for the estimation of the topographic index distribution used in TOPMODEL." *Hydrol. Process.*, vol. II, pp. 1099-1114.
6. Nathabandu T. Kottegoda and Renzo Rosso. (1996). *Statistics, Probability, and Reliability for Civil and Environmental Engineers*, McGraw-Hill, pp. 115-116.
7. O'Callaghan, J.F., and Mark, D.M. (1984). "The extraction of drainage networks from digital elevation data." *Computer Vision, Graphics and Image Processing*, Vol. 28, pp. 323-344.
8. Quinn, P. F., K. Beven and R. Lamb (1995). "The $\ln(a/\tan\beta)$ index : How to calculate it and how to use it within the TOPMODEL framework." *Hydrol. Process.*, vol. 9, pp. 161-182.
9. Quinn, P., K. Beven, P. Chevallier, and O. Planchon (1991). "The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models." *Hydrol. Process.*, vol. 5, pp. 59-79.
10. Shannon, C.E., and Weaver, W. (1949). *The Mathematical Theory of Communication*, The University of Illinois Press, Urbana, pp. 117.
11. Western, A.W., Grayson, R.B., Blöschl, G., Willgoose, G.R., and McMahon, T.A. (1999). "Observed spatial organization of soil moisture and its relation to terrain indices." *Water Resour. Res.*, Vol. 35, No. 3, pp. 797-810.