J. Korea Water Resour. Assoc. Vol. 53, No. 3 (2020), pp. 215-223
doi: 10.3741/JKWRA.2020.53.3.215
pISSN 1226-6280
eISSN 2287-6138

Model development for the estimation of specific degradation using classification and prediction of data mining

Jang, Eun-kyung[a] · Kang, Woochul[b]*

[a]Research Specialist, Department of Land, Water and Environment Research, Korea Institute of Civil Engineering and Building Technology, Goyang, Korea
[b]Post-doctorate Researcher, Department of Land, Water and Environment Research, Korea Institute of Civil Engineering and Building Technology, Goyang, Korea

Paper number: 20-014
Received: 19 February 2020; Revised: 10 March 2020; Accepted: 10 March 2020

Abstract

The objective of this study is to develop a prediction model of specific degradation using data mining classification especially for the rivers in South Korea river. A number of critical predictors such as erosion and sediment transport were extracted for the prediction model considering watershed morphometric characteristics, rainfall, land cover, land use, and bed material. The suggested model includes the elevations at the mid relative area of the hypsometric curve of watershed morphomeric characteristics, the urbanization ratio, and the wetland and water ratio of land cover factors as the condition factors. The proposed model describes well the measured specific degradation of the rivers in South Korea. In addition, the development model was compared with the existing models, since the existing models based on different conditions and purposes show low predictability, they have a limit about the application of Korean River. Therefore, this study is focusing on improving the applicability of the existing model

Keywords: Data mining, Factors extraction, Model development, Specific degradation, Watershed

데이터 마이닝의 분류 및 예측 기법을 적용한 비유사량 추정 모델 개발

장은경[a] · 강우철[b]*

[a]한국건설기술연구원 국토보전연구본부 전임연구원, [b]한국건설기술연구원 하천연구센터 국토보전연구본부 박사후연구원

요 지

본 연구의 목적은 국내 하천을 대상으로 데이터 마이닝의 분류 및 예측 기법을 활용하여 비유사량 추정 모델을 개발하는 것이다. 이를 위해 유사이송에 영향을 미치는 요소들을 전반적으로 고려하여 유역인자를 추출하였으며, 유역의 지형학적 요소, 강우, 토지 피복, 토지 이용, 하상 재료 등이 고려되었다. 추출된 인자를 활용하여 모델을 도출한 결과 유역 형태학적 특성인자 중 평균 면적비에서 유역고도 및 토지피복인자 중 도시화 비율과 전체 유역 중 습지와 수역의 비율이 조건인자로 활용되었다. 도출된 모델은 실측값과의 비교를 통해 실측 비유사량의 발생 패턴이 유사하게 재현됨을 확인하였다. 또한 기존의 사용되던 산정 공식과 비교하였으며, 국외의 데이터를 기반으로 도출된 모델은 개발 배경 및 국내 하천 환경과의 차이로 인해 국내 하천 데이터 적용에 한계가 있는 것으로 나타났다. 이에 본 연구에서는 개발 및 적용 환경, 데이터 범위의 차이 등으로 인해 발생하던 기존 공식의 한계를 개선하고자 하였다.

핵심용어: 데이터 마이닝, 모델 개발, 비유사량, 유역, 인자 추출

***Corresponding Author.** Tel: +82-54-843-1826
E-mail: kang@kict.re.kr (W. Kang)

© 2020 Korea Water Resources Association. All rights reserved.

1. 서 론

토양의 침식은 지상류(overland flow)의 전단력이 토양의 응집력보다 큰 경우 일어나며, 강우 이벤트 중 토양입자의 분리는 강우에너지와 강우로 발생된 지표수의 흐름으로 인한 전단력에 의해 발생한다. 전단력의 정도는 빗방울의 크기와 면상류의 깊이(sheet flow depth) 및 토양 입자의 크기에 영향을 받으며, 분리된 입자는 보통 흐름에 의해 이송된다. 이송되는 유사의 양과 질의 정도는 하천의 수리학적인 특성, 하도의 특성, 유역형태, 강우강도, 강우량, 토지이용 등 다양한 유역특성의 영향을 받는다. 지상류에 의해 이송되는 분리된 토양입자는 하천에 도달하기 전이나 하천에 도달이후 부유사가 되어 하천에 퇴적이 된다. 이렇듯 유사의 생성, 이송, 퇴적의 과정은 하천과 유역의 지리학, 지형학, 지질학, 토양학, 기후학, 임학등과 관련된 유역특성들의 영향을 받는다(Julien, 2018). 따라서 다양하고 복잡한 유사특성을 현재의 지식으로 명확하게 규명하고 정량적인 양을 정확하게 예측하는 것은 매우 어려운 일이다. 그럼에도 불구하고 대략적으로 유사량 혹은 유사이송률을 추정할 수 있는 수많은 공식들이 제안되어 왔다. 이들의 원론적인 개념은 부유사와 소류사를 구분한 후 이송-확산모형으로 접근하는 방법, 유사이송에 필요한 에너지는 물 흐름 에너지로부터 공급받는다는 에너지 개념의 모형으로 접근하는 방법, 마지막으로 기존의 실측 혹은 실험으로 분석된 유사량과 흐름 자료를 통계적으로 분석하는 경험적 접근방법으로 구분된다. 이러한 유사량의 추정은 하천을 설계하거나 계획하는 경우 필수적일 뿐만 아니라 운영 및 관리에도 중요한 요소이다. 특히 단위기간에 단위면적당 상류유역에서 유출되어 유역출구의 한 지점을 통화하는 유사량을 의미하는 비유사량의 경우 하천의 댐을 계획하거나 그 계획댐의 저수지 퇴사관계를 설명하고 규명하기 위한 필수요소이다(KICT, 1992).

비유사량 추정을 위해 Ryu and Min (1975)의 삽교천내 9개 저수지 퇴사자료를 이용한 다중회귀분석에 의한 비퇴사량 추정식을 제안한 연구를 시작으로 Ryu and Kim (1976), Yoon (1982), Ahn and Lee (1984), Suh *et al*. (1988) 등이 측정 유사량을 바탕으로한 회귀분석식을 제시하였다. 특히 비유사량 추정을 위해 경험공식, 유사운송비 방법, 도표식 방법 등의 통계적 경험방법에 의해 개발된 공식(KICT, 1992)이 현재까지 가장 일반적으로 활용되고 있다. 그러나 어떤 방법을 사용하여 개발되었건 지금까지의 유사 이송 공식들은 동일한 지점, 동일한 수리조건일 지라도 공식개발에 기초한 이론적 배경과 사용 자료의 적용 범위에 따라 큰 차이를 보이는 경우가 있다. 따라서 최근에는 존재하는 변수들로 이루어진 대규모 데이터 안에서 체계적이고 자동적으로 통계적 규칙이나 패턴을 도출하는 데이터 마이닝 기법을 적용한 비유사량 추정 모델개발 연구가 진행되고 있다. Jain (2001)의 연구에서는 인공신경망 모델을 활용하여 유사농도를 추정하였으며, 이를 위해 수위, 유량, 유사량과 시간의 관계를 규명하였으며, Lin and Namin (2005)은 부유사 이송을 분석하기 위해 인공신경망 모델과 수치해석을 통합적으로 사용하였다. 또한 Nagy *et al*. (2002)는 유사농도 산정을 위해 인공신경망 모델 활용한 후 기존의 공식들과 비교하였으며, Bhattacharya *et al*. (2007)은 유사량 산정을 위해 인공신경망과 Model Tree를 사용하여 비교 및 평가하였다. 그러나 이러한 데이터 마이닝 기법의 적용을 위해서는 대규모의 데이터 수집이 우선되어야 하며, 비유사량 발생에 영향을 미치는 다양한 인자들이 반드시 구축되어야 한다.

따라서 본 연구에서는 국내 하천을 대상으로 현장에서 측정된 비유사량을 활용하여 유사의 생성, 이송, 퇴적에 영향을 미치는 유역특성인자를 도출한 후 데이터 마이닝의 분류 및 예측 기법을 적용하여 새로운 비유사량 산정 모델을 제시하고자 한다. 또한 기존에 사용 중인 비유사량 산정공식과의 비교를 통해 국내 하천에서의 적용 가능성 및 한계성을 검토하는 것이 본 논문의 목적이다.

2. Model Tree의 개요 및 분석 조건

2.1 Model Tree의 개요

데이터 마이닝의 가장 대표적인 예측과 분류 분석 기법은 인공신경망(Artificial Neural Network, ANN), 회귀분석(Regression Analysis) 혹은 로지스틱 회귀분석(Logistic Regression Analysis, LRA), 의사결정나무(Decision Tree) 등이 있다(Jang, 2017). 이 중 의사결정나무의 Model Tree 기법은 데이터로부터 모델을 만드는 여러 가지 기법들 중에서도 가장 명료하고 결과에 대한 해석이 용이하고 상대적으로 이상치의 영향을 덜 받는 것으로 알려져 있다(Jang, 2017). 특히 Model Tree는 경험식과 같은 명시적 공식을 도출하는데 간단하면서도 탁월하여, 효과적으로 사용할 수 있는 기법으로, 여러 개의 물리량들이 복잡하게 서로 연계되어 유사가 이송되는 유사량 산정에는 이 기법이 제일 적합할 것으로 판단된다(Jang, 2017). Model Tree 기법은 데이터 그룹을 유사한 성격의 데이터끼리 분류하여 일정한 조건에 의해 하위그룹으로 분리해 나가는 절차를 기반으로 한다. Model Tree의 구조는 tree의 성장(growing)과 전지(pruning) 그리고 다듬는 작업(smoothing)으로 구성되어 있으며, Eq. (1)의 표준편차감소

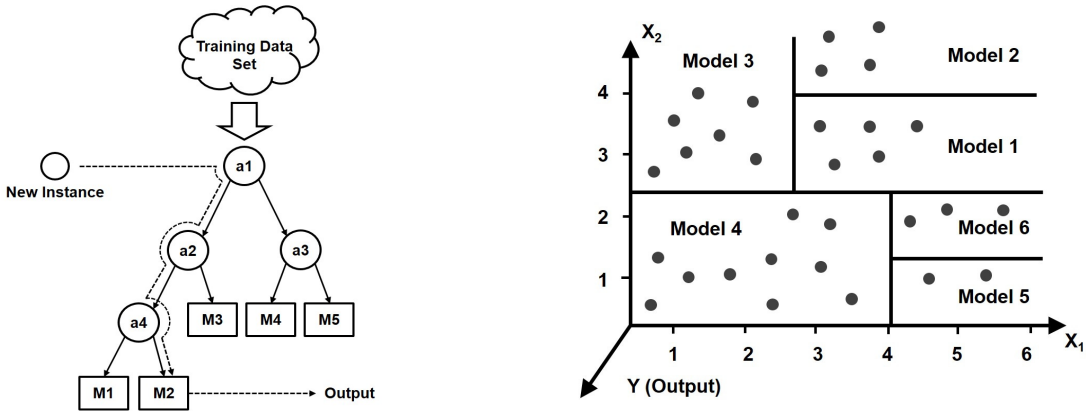


Fig. 1. Diagram for process and calculation of model tree (Bhattacharya *et al.*, 2007)

율(Standard Deviation Reduction, SDR)이 최대치가 될 때 큰 가지에서 작은 가지로 전지된다(Quinlan, 1992; Wang and Witten, 1996).

$$SDR = \sigma(T) - \sum_i \frac{|T_i|}{|T|} \times \sigma(T_i) \quad (1)$$

여기서, T는 종속변수의 전체 표본 집합이며, T_i는 세부 구간으로 나누어진 종속변수의 하위 표본 집합, σ는 표준편차(standard deviation), $|T|$와 $|T_i|$는 원소의 개수로 표현되는 집합의 크기이다. 각 독립변수의 그룹화 구분은 종속변수의 표준편차감소율로 판단된다. 즉, 불필요한 가지를 포함한 전체 표본 집합 T에서 표준편차 $\sigma(T)$를 계산 한 후 각 독립변수 별 임의 구간에 따라 임의의 세부 구간으로 나누어진다. 임의로 구분된 세부 구간 중 표준편차감소율이 가장 큰 구간을 선택하여 전지되며, 이 후 하부 tree가 성립한다. 전지된 하위 그룹의 성장 및 전지조건도 똑같은 과정을 거치며, 표준편차감소율이 원하는 수치에 도달했을 경우, 혹은 그룹화 후 남은 자료의 개수가 선택한 기준보다 작아졌을 때 tree의 성장이 종료된다. Model Tree의 생성 및 전지, 종료 과정은 Jang (2017)에 설명되어 있으며, 이와 같은 과정을 거쳐 최종적인 세부 구간이 결정되면 각 구간의 다중회귀함수가 결정된다. 일정한 조건을 만족하여 성장이 끝난 최종 Model Tree의 모식도는 Fig. 1과 같다.

모든 분류 과정이 끝난 후 동일 그룹에 속한 데이터들의 대표식은 다중회귀분석을 통해 제시된다. 본 기법의 장점중에 하나는 일정한 규칙이나 통일성이 없는 자료의 그룹안에서 하위 그룹을 통해 동질한 데이터 그룹을 생성하고 이 과정을 통해 최적의 다중 회귀식을 제시한다는 것이다(Jang, 2017). 본 연구에서는 현장에서 측정된 대용량의 유사량 자료를 유사의 생성 및 이송에 영향을 미치는 유역특성인자와 함께 Model Tree를 적용하여 분석하고, 단위 유역당 연간 총비유사량 산정 모델을 도출하고자 하였다.

2.2 유사량 측정 자료 및 유역인자 추출

수자원조사기술원(Korea Institute of Hydrological Survey, KIHS)은 2005년부터 국내 하천의 여러 지점에 대해 유사량 측정을 수행해 왔다. 본 연구에서 활용한 유사량 자료는 한강유역의 여주, 흥천, 율극, 청미, 남한강, 흑천지점, 낙동강유역의 선산, 동촌, 구미, 낙동, 왜관, 일선교, 진동, 정암, 향석, 동문, 점촌, 용곡, 죽고, 개진2지점 금강유역의 회덕, 공주, 합강, 우성, 구룡지점, 영산강유역의 학교, 나주, 마륵, 남평, 선암지점, 섬진강유역의 죽곡, 곡성, 구례2, 용서지점이다. 총유사량은 2005년부터 2014년 까지 해당 지점에서 측정된 유사량 자료(HSC, 2007; 2008; 2009; 2010; 2011; 2012; 2013; 2014a; 2014b)와 수정 아인슈타인 공식(Einstein, 1950)을 이용하여 계산하였으며, 계산된 총유사량과 10년간 유량 자료를 이용하여 Flood Duration and Sediment Rating Curve (FD-SRC)를 도출한 후 연평균 총유사량을 산정하였다. 또한 해당 유역에서 생성되는 연간 총유사량을 단위면적당 비교하기 위해 비유사량을 계산하였다. 비유사량을 계산하기 위한 상세 과정은 Kang *et al.*, 2019을 참고하였다. 해당지점에서 측정된 비유사량에 영향을 주는 유역 관련 특성인자들은 GIS 분석을 이용하여 추출하였으며 이를 Model Tree 분석에 활용되었다.

또한 기상청에서 제공하는 60개 지점의 30년간 강우량 자료를 이용하여 연평균 강우량을 계산한 후 Original Krigging Method (OKM)를 이용하여 우리나라 전역의 연평균 강우량을 산정하였다. OKM 결과에서 35개 유역을 대상으로 유역평균 결과와 유사량 측정 지점의 결과를 획득하였으며, 같은 방법

으로 강우침식도에 대한 결과를 추출하였다. 유역의 토양학적인 특성을 고려하기 위하여 국립농업과학원에서 제공하는 전국 토양통 정보를 가진 세부정밀 토양도를 이용하여 유효 토심 0 *cm*에서 10 *cm*까지 진흙($d_{clay} \leq 0.002\ mm$), 실트($0.002\ mm < d_{silt} \leq 0.05\ mm$), 모래($0.05\ mm < d_{sand} \leq 2\ mm$), 자갈($2\ mm < d_{rock}$)의 비율을 계산하여 토양종류를 구분하였다. 유역의 토지이용과 관련된 특성은 환경부에서 제공하는 22개 항목 중분류 토지피복도를 이용하여 토지이용 비율을 대분류 기준(시가화 건조지역, 농업지역, 산림지역, 초지, 습지, 나지, 수역)으로 분류하였다. 또한, 유사량 측정 과정에서 측정된 하상재료의 중앙입경도 고려되었다. 유역의 자연적인 지리, 지형학적인 특성을 고려하기 위하여 5 m Digital Elevation Model (DEM) 자료를 활용하여 유역의 형태와 관련된 인자들을 획득 하였다. 이 과정에서 하천의 길이, 측정지점 하천차수, 경사도를 고려하기 위하여 유역 추출 과정에서 획득한 하천자료와 환경부에서 제공하는 Korea Reach File (KRF, version 3.0.)의 자료를 이용하여 산정하였다. 유역의 지형 및 형상등과 연관이 있는 유역면적(km^2), 유역밀도(km / km^2), Form Factor (km^2 / km^2), Shape Factor(km^2 / km)를 계산하였으며, 유역의 경사와 연관된 유역평균경사(%), 최소 및 최대 유역고도(m)가 계산 되었다. 마지막으로 Strahler (1952)가 제안한 고도분포곡선을 이용여 고도분포곡선으로부터 해당유역의 곡선면적비적분 인자와 평균 면적비에서 유역고도를 획득하여 모델의 인자로 고려되었다. 본 연구에서 활용한 유역특성인자를 정리하면 Table 1과 같다.

3. 비유사량 모델 도출

본 절에서는 현장에서 측정된 유사량 자료로부터 산정된 비유사량과 추출된 유역인자를 활용하여 비유사량 산정 모델을 도출하였다. 이를 위해 2.1절에서 설명한 데이터 마이닝의 Model tree 기법을 활용하였으며, 산정결과 유역 형태학적 특성인자 중 고도분포 지수(Hypsometric Index, *Hyp*) 및 토지피복인자 중 도시화 비율(Percentage of Urban, *Urban*)과 전체 유역 중 습지와 수역의 비율(Percentage of wetland and water, *WW*)이 조건인자로 사용되었다. 도출된 모델은 Eq. (2)와 같다.

$$SD = a \times Urban^b \times WW^c \times Hyp^d \tag{2}$$

여기서, $SD\ [tons/km^2 \cdot year]$는 단위 유역당 생성되는 연간 비유사량이다. 또한 이 식의 사용 조건 및 지수와 계수는 Table 2와 같다. 다음의 사용조건은 Model Tree의 성장과 전지 등에 의해 성립된 구조로써 최종적으로 성립된 최하위 그

Table 1. Parameters for developed model

Classification (Number of Parameters)	Parameters
Watershed Morphometry (16)	Main Stream Length, Tributary Length, Total Stream Length, Strahler Stream Order (KRF), Strahler Stream Order (Delineated Stream), Shreve Stream Order (Delineated Stream), Watershed Area, Drainage Density, Shape Factor, Form Factor, Watershed Average Slope, Elevation At Mid Relative Area, Low Elevation, Mean Elevation, Hypsometric Index, River Slope
Precipitation (4)	Mean Annual Precipitation At Gauging Point, Mean Annual Precipitation At Watershed, Rainfall Erosivity At Gauging Point, Rainfall Erosivity At Watershed
Land Cover (8)	Percentage of Urban, Agriculture, Forest, Pasture, Bareland, Wetland, Water, Wetland And Water
Soil Type at From 0 *cm* to 10 *cm* Effective Soil Depth (4)	Percentage of Clay, Silt, Sand, Rock
Bed Material Size (2)	Minimum of d_{50}, Maximum of d_{50}

Table 2. Model tree application equations by conditions and determination coefficients

Hypsometric Index, *Hyp*	Main Stream Length, L_{main}	Percentage of Wetland and Water, *WW*	Percentage of Urban, *Urban*	Eq	Constant Coefficient	Exponent for Variables		
					a	b	c	d
<= 187	-	-	-	Eq. 1	661.10	0.551	-0.342	-0.442
> 187	<= 265	<= 2.63	<= 3.11	Eq. 2	109.16	0.639	-0.264	-0.231
			> 3.11	Eq. 3	111.61	0.642	-0.264	-0.231
		> 2.63	-	Eq. 4	113.10	0.580	-0.271	-0.231
	> 265	-	-	Eq. 5	100.83	0.567	-0.179	-0.231

룹에서의 다중 회귀된 모델을 의미한다. 사용되는 조건은 평균 면적비에서 유역고도, 주하도 길이, 도시화 비율, 습지와 수역의 비율로 각각의 조건에 의해 총 5개 모델이 성립된다. 제안된 식의 지수(a~d)가 0보다 작은 경우 각 지수에 해당하는 인자가 증가하면 생성되는 유사량이 감소 및 유역에서 생성되는 유사량이 작음을 의미하며 반면에 0 보다 큰 경우 반대의 상황을 의미한다. 본 연구에서 제안된 식의 경우 지수들의 유사한 음수의 값 또는 양수의 값을 가진다. 이러한 유사 범위의 지수 값은 비유사량과 조건인자들의 관계가 일관성을 가지고 있으며, 제안된 식이 미계측 유역에 이용될 수 있음을 의미한다.

제안된 식에서 사용된 인자들의 의미를 살펴보면 도시화 비율의 경우 지수의 값이 양의 값을 가지며 도시화 비율이 높은 경우 총유사량이 늘어남을 의미한다. 이미 다른 연구에서도 유사한 경향을 보인바 있으며, 건설현장과 같은 도시화가 진행되는 과정 중 많은 유사량이 생성될 수 있기 때문에 도시화의 비율이 유사량 증가에 영향을 미친다. 습지와 수역의 비율이 높은 경우에는 생성되는 유사량이 감소한다. 내륙습지의 경우 대부분 하천의 범람원을 이루고 있으며 습지와 수역의 비율이 높으면 생성된 유사량이 퇴적이 될 기회가 많음을 의미한다. 또한 평균 면적비에서 고도가 낮은 경우 유역에 산지보다 평지가 많음을 의미한다. 제안된 식에서는 산지하천과 같이 고도가 높은 경우 유역의 유사량이 감소한다. 대부분의 충적하천이며 산지가 산림으로 이루어진 한국의 경우 경사도가 큰 영향을 미치지 않으며, 유역 경사가 큰 경우 비유사량이 감소하는 경향은 Kang *et al*. (2019)과 Kane and Julien (2007)의 연구에서도 보여준바 있다. 그룹의 분류에 이용된 주하도 길이는 값이 큰 경우 유역의 면적이 크기 때문에 유사량의 이송되며 퇴적되는 양이 많아질 수 있음을 의미한다.

도출된 단위유역당 비유사량 산정 모델에 의한 비유사량은 Fig. 2와 같이 실제 비유사량과 비교하였다. 최적 적합도를 의미하는 $y=x$를 기준으로 아래쪽에 분포된 데이터는 측정값에 비해 과소 산정, 위쪽에 분포된 데이터는 측정값에 비해 과대 산정됨을 의미하고 적합도(Discrepancy Ratio, *R*)가 1에 가까울 경우 측정값과 산정값이 유사함을 의미한다. Model Tree에 의한 비유사량 산정 모델 도출 결과, 평균 1.26으로 다소 과대 산정되는 경향이 있음에도 불구하고 도출된 모델은 최적 적합도에 매우 근접하여 실측값과 상당히 유사한 것으로 나타났다. 특히 대부분의 실측 데이터가 해당하는 단위 유역당 비유사량 범위 100 $tons/km^2 \cdot yr$ 이상 1,000 $tons/km^2 \cdot yr$ 이하의 범위에서는 평균 적합도가 1에 근접한 것으로 나타났다. 또한 개별 관측치에 대한 오차의 정도를 판단하는 평균 제곱근 오차(Root Mean Square Error, RMSE)를 산정하였다.

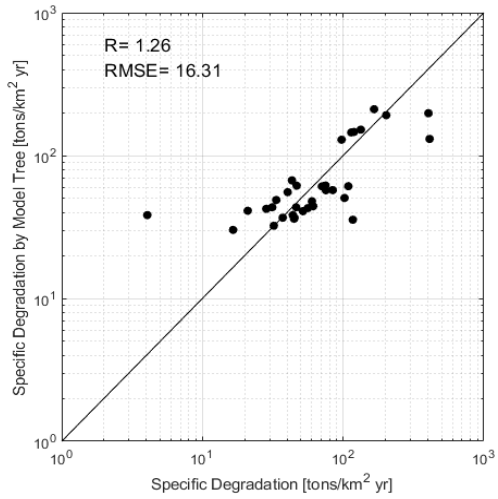


Fig. 2. Comparison of model tree with measured specific degradation

평균 제곱근 오차는 값이 작을수록 정밀도가 높음을 의미하며, 산정결과, 16.31로 나타났다. 이를 통해 곧 본 연구를 통해 도출된 모델이 국내 하천 범위에서 활용 가능성이 높을 것으로 판단되며, 국내 하천에서의 비유사량을 결정짓는 인자로 도시화비율, 습지와 수역의 비율, 고도 분포에 의한 변수가 가장 결정적인 영향이 있음을 보여주었다.

4. 기존 비유사량 산정 공식에 의한 비유사량과 도출된 모델 비교

본 절에서는 유사량 추정을 위해 사용되고 있는 기존의 여러 경험식들에 의한 산정값을 도출된 모델과 비교해보았다. 한국건설기술연구원은 한국형 댐 유사량 산정을 위해 유사량 산정 공식을 제안한바 있다(KICT, 1992). 이 후 두 차례 걸쳐 해당 공식이 보완되었으며, 본 논문에서는 한국 댐 설계기준 해설(MLTMA, 2011)의 공식을 적용하여 비교하고자 하였다. 해당 공식은 유역 면적 조건에 따라 다른 공식이 제안되고 있으며, 본 논문에서는 유역면적 200 km^2 이하인 경우와 유역면적이 200 km^2에서 2,000 km^2인 경우의 조건을 적용하였다.

$$SD = 8{,}668 \times A^{-0.896} \ (\text{if } A < 200\,km^2) \tag{3}$$

$$SD = 972 \times D^{1.039} \times M^{-0.825}$$
$$(\text{if } 200\,km^2 \le A < 2{,}000\,km^2)$$

여기서, *A*는 유역면적(km^2), *D*는 하천 밀도(km^2 / km^2), *M*은 하상 재료의 중앙입경(*mm*)을 의미한다.

한국 댐 설계기준 해설(MLTMA, 2011)에 제안된 Yoon (2011)의 공식의 경우 한국의 10개 다목적댐의 퇴사량 자료를 기반으로 다중회귀분석을 활용하여 식을 제안하였다. 본 절에서 검토된 Yoon (2011)공식은 다음과 같다.

$$SD = 43{,}954\,A^{0.464} \times S^{-2} \times M^{-0.855} \tag{4}$$

여기서, A는 유역면적(km^2), S는 유역경사(km/km), M은 하상재료의 중앙입경(mm)을 의미한다.

또한 Kane (2003)은 미국의 1,463개 저수지의 퇴사량 자료를 기반으로 비유사량의 산정을 위해 유역면적, 평균 강우량, 평균 유역경사를 이용한 회귀분석에 의한 공식을 다음과 같이 제안하였다. Eq. (3)은 연평균 강우량과 비유사량의 관계에 의한 식이고, Eq. (4)는 유역면적과의 관계에 의한 식이다.

$$SD = 0.02 \times R^{1.7} \times e^{-0.0017R} \tag{5}$$

$$SD = 410A^{-0.09} \tag{6}$$

여기서, R는 연평균강수량(mm), A는 유역면적(km^2)을 의미한다.

마지막으로 Kang *et al*. (2019)은 한국 하천의 35개 지점을 대상으로 유사량 측정자료에 의해 비유사량을 계산한 후, 38개의 유역 특성을 이용한 회귀분석식을 제안하였다. Kang *et al*. (2019)은 사용자 편의를 고려하려 접근성이 쉬운 인자로부터 회귀분석을 통한 모델을 개발하였으며, 다음과 같은 6개 식들을 제안하였다.

$$SD = 357.16 \times A^{-0.204} \tag{7}$$

$$SD = (3.35 \times 10^{-7}) \times A^{-0.16} \times P^{2.864} \tag{8}$$

$$SD = (3 \times 10^{-4}) \times A^{-0.08} \times P^{1.65} \times U^{0.75} \tag{9}$$

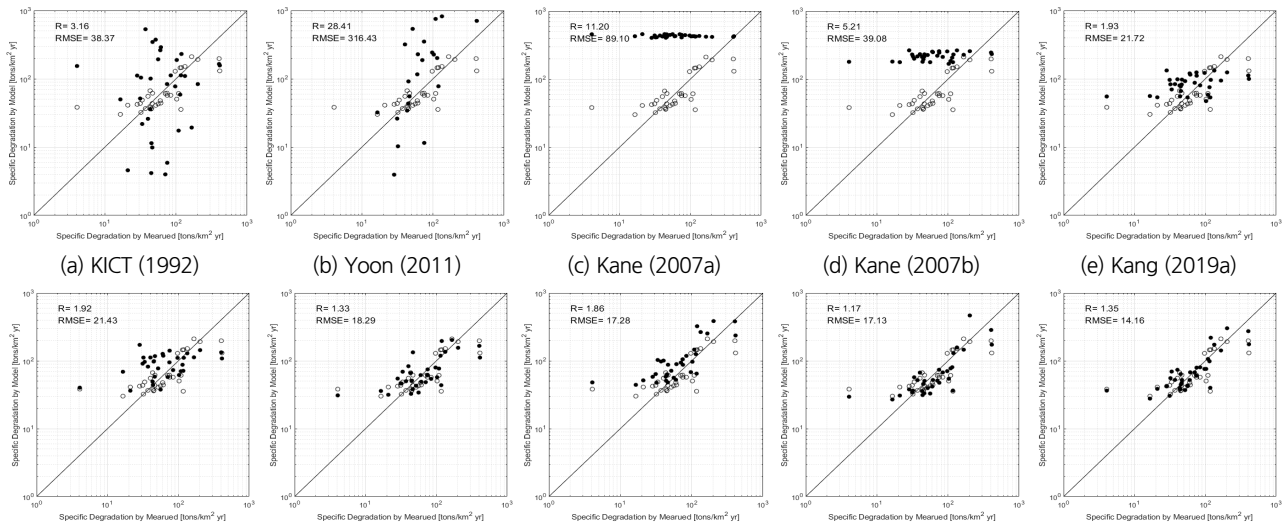
$$SD = (1.75 \times 10^{-7}) \times A^{-0.05} \times P^{1.89} \times U^{0.89} \times Sand^{1.931} \tag{10}$$

$$SD = (1.77 \times 10^{-5}) \times A^{-0.009} \times P^{1.91} \times U^{0.53} \times Sand^{1.09} \times S^{-0.93} \tag{11}$$

$$SD = (2.45 \times 10^{-7}) \times A^{-0.04} \times P^{1.94} \times U^{0.61} \times W^{-0.64} \times Sand^{1.51} \times Hyp^{1.84} \tag{12}$$

여기서, A는 유역면적(km^2), P는 연평균강수량(mm), U는 유역내 도시화면적비(%), $Sand$는 유효 토심 0 cm에서 50 cm까지 구간에서의 모래비율(%), S는 유역평균경사(%), W는 유역내 습지 비율(%), Hyp는 고도분포곡선의 기울기를 의미한다.

위에서 설명한 기존의 비유사량 산정공식과 Model Tree에 의한 비유사량 추정 모델의 비교를 위해 실제 측정된 비유사량과 함께 Fig. 3과 Table 3과 같이 비교하였다. 비교 결과, 국내 하천 데이터를 기반으로 개발된 KICT (1992) 공식과 Yoon (2011) 공식의 경우 최적 적합도로부터 산포도가 매우 큰 것을 알 수 있으며, 두 공식의 적합도는 각각 3.158과 28.411인 것으로 나타났다. 이는 비유사량 범위에 따라 과대 산정 폭이 매우 크다는 것을 의미한다. 또한 외국의 데이터를 활용하여 개발된 Kane (2007) 공식의 경우 연평균 강우량에 관계한 식과 유역면적과 관계한 식 모두 국내 하천에 적용하는데 한계가 있는 것으로 나타났다. 그러나 국내 하천의 데이터를 기반으로 개발된 Kang (2019)의 모형의 경우 평균 적합도가 모두 2 이하였으며, 평균 제곱근 오차 역시 평균 18.34로 우수하였다. 특히 Eq. (11)을 적용한 Fig. 3(i)과 Eq. (12)을 적용한 Fig. 3(j)의 결과의 평균 적합도 및 평균 제곱근 오차가 매우 우수함을 알 수 있다. 또한 비유사량 100 $tons/km^2 \cdot yr$에서 1,000 $tons/km^2 \cdot yr$까지의 범위에서는 최적 적합도가 1 초반으로 매우 우수하였다. 이는 Kang (2019)의 모든 모델이 국내 하천의 데이트를 기반으로 도출된 공식이며, 기존 데이터의 회귀분석을 통해 도출된 공식으로 이와 유사한 국내 데이터베이스 내에서 매우 우수한 결과를 도출한 것으로 판단된다. 또한 Mode Tree에 의해 도출된 본 논문의 비유사량 추정 모델과도 매우 유사함을 알 수 있다. 그러나 Kang (2019)의 두 공식은 필요한 변수로 최소 5가지가 요구되는 공식으로 미계측 유역 혹은 여러 가지 요인에 의해 변수 도출이 어려운 구간의 경우 해당 공식 활용에 제약이 있을 수 있다. 따라서 도출하고자 하는 구간에서의 환경적인 요인 및 수집 가능한 유역 변수, 모형 개발에 활용한 데이터의 범위 등을 판단한 후 비유사량 추정을 위한 모델을 활용하여야 할 것이다. 본 논문에서 활용한 Model Tree 기법의 경우, 제한적인 데이터 범위 내에서도 우수한 결과를 도출할 수 있다는 장점을 확인하였다. 그러나 이와 같은 장점에도 불구하고 Table 3의 결과 중 일선교 구간과 같이 일반적인 유사 발생의 패턴을 벗어나는 이상적인 수치의 실측값이 존재하는 경우 모델 정확도에 영향을 미칠 수 있음을 유의하여야 할 것이다.



(a) KICT (1992) (b) Yoon (2011) (c) Kane (2007a) (d) Kane (2007b) (e) Kang (2019a)

(f) Kang (2019b) (g) Kang (2019c) (h) Kang (2019d) (i) Kang (2019e) (j) Kang (2019f)

Fig. 3. Comparison of model tree (○) and existing model (●) with measured specific degradation

Table 3. Comparison of discrepancy ratio and correlation coefficient (Unit: $tons/km^2 \cdot yr$)

Name	Measured	Model Tree	KICT (1992)	Yoon (2011)	Kane (2007a)	Kane (2007b)	Kang (2019a)	Kang (2019b)	Kang (2019c)	Kang (2019d)	Kang (2019e)	Kang (2019f)
Yeoju	117.19	35.56	112.48	1307.82	420.47	177.28	53.40	71.31	43.76	265.13	36.55	39.28
Heungcheon	403.57	197.78	164.99	1662.67	416.17	245.77	111.97	133.54	166.24	1689.52	285.22	274.60
Yulgeuk	202.52	191.84	83.75	4249.13	416.80	257.27	124.20	143.89	155.87	1747.57	466.08	305.02
Cheongmi	411.74	130.74	157.03	705.85	426.69	233.55	99.74	108.35	111.77	1037.19	173.18	175.47
Namhanriver	16.51	30.17	49.98	32.21	426.48	180.97	55.95	69.02	35.98	175.64	26.94	27.13
Heukcheon	75.29	57.19	5.89	11.52	410.00	244.84	111.01	141.29	74.84	344.94	47.59	79.05
Seonsan	70.69	60.88	3.99	187.90	456.00	220.43	87.49	57.74	48.44	413.94	51.06	67.15
Dongchon	43.30	66.99	35.80	34.27	458.02	211.82	79.94	49.45	65.89	211.82	35.76	39.66
Gumi	20.99	41.08	4.58	1568.95	457.92	177.57	53.59	36.31	31.65	216.74	30.80	38.34
Nakdong	43.90	38.31	100.69	2805.76	452.92	179.97	55.25	44.19	32.64	212.07	31.29	37.33
Waegwan	56.08	42.87	193.03	2503.39	457.04	177.29	53.41	37.67	34.13	237.07	32.98	42.28
Ilseon bridge	4.06	38.30	153.98	2156.62	455.91	179.74	55.09	40.30	30.98	200.46	29.60	36.04
Jindong	102.38	50.48	188.64	226.58	424.52	167.88	47.19	61.82	57.32	401.72	54.42	75.71
Jeongam	33.45	48.86	21.82	1108.62	411.53	199.51	69.79	96.70	69.13	419.87	57.29	68.78
Hyangseok	84.28	57.31	112.12	1511.56	442.58	212.11	80.19	73.11	54.27	531.53	67.47	79.46
Dongmun	74.81	61.70	83.80	350.94	447.12	257.29	124.21	94.99	62.33	454.23	73.93	59.84
Jeomchon	31.34	43.47	51.43	26.13	437.97	230.22	96.55	90.91	46.70	262.64	34.45	29.80
Yonggok	46.33	43.46	9.87	1977.58	454.22	214.86	82.56	58.41	38.12	194.06	33.26	47.15
Jukgo	37.22	36.73	532.21	1843.83	437.07	215.95	83.51	82.22	50.03	431.70	51.94	51.59
Gaejin2	51.76	40.80	374.86	538.74	445.70	225.96	92.55	77.47	42.98	306.91	39.62	36.50
Hoedeok	119.12	146.16	230.94	77.74	422.44	230.22	96.55	111.00	196.03	1312.14	129.93	219.08
Gongju	108.76	61.03	17.44	753.98	427.51	184.31	58.32	70.42	76.96	620.80	77.45	106.42
Hapgang	133.54	151.82	109.83	821.74	430.39	264.37	132.10	129.00	135.85	1150.69	157.42	173.04
Useong	61.23	44.15	290.17	228.75	428.22	248.79	115.11	118.99	49.25	395.51	52.23	67.27
Guryong	60.08	47.87	262.07	116.34	425.69	253.50	120.10	126.85	61.55	195.54	41.80	52.29
Hakgyo	97.44	129.13	77.48	242.73	437.07	255.27	122.02	110.69	79.71	250.25	70.46	76.55
Naju	114.35	145.15	58.72	201.45	426.12	206.36	75.34	87.55	118.94	465.74	80.50	98.65
Mareuk	165.99	211.13	19.32	1712.53	419.46	227.85	94.31	112.68	202.81	961.98	145.99	142.13
Nampyeong	46.72	61.53	346.14	55.31	418.09	231.07	97.35	117.25	133.50	294.51	50.72	52.41
Seonam	40.34	55.44	25.94	318.90	422.94	232.16	98.40	112.03	83.32	395.50	67.07	73.55
Jukgok	32.15	32.27	104.12	10.26	410.74	215.44	83.07	111.72	46.39	183.56	36.47	55.33
Gokseong	44.85	36.12	4.16	43.47	418.55	208.88	77.45	97.50	50.37	241.78	45.61	63.93
Gurye2	45.15	36.63	11.41	91.63	405.23	195.13	66.37	98.86	51.19	249.75	45.72	44.36
Yongseo	28.43	42.34	111.60	3.95	404.04	264.80	132.58	172.03	54.78	229.53	37.93	41.72

5. 결 론

본 연구에서는 국내 하천을 대상으로 유사에 영향을 미치는 전반적인 유역특성인자를 도출한 후 데이터 마이닝의 분류 및 예측 기법을 적용하여 새로운 비유사량 산정 모델을 제시하였다. 또한 도출된 모델은 기존에 사용 중인 비유사량 산정 공식과의 비교분석을 통해 적용성을 검토하였으며, 주요 결과를 다음과 같이 정리할 수 있다.

첫째, 국내에서 2005년부터 측정된 유사량 측정 데이터를 기반으로 GIS 분석을 통해 유역인자를 추출하고자 하였으며, 강우자료, 토지이용, 토지피복 등을 종합하여 총 34개의 유역인자를 도출하였다. 도출된 인자는 데이터 마이닝의 예측과 분류 기법 중 Model Tree 기법을 적용하여 비유사량 추정 모델을 도출하였으며, 도출된 모델은 유역 형태학적 특성인자 중 고도분포 지수와 토지피복자 중 도시화 비율 및 전체 유역 중 습지와 수역의 비율이 조건인자로 활용되었다. 또한 도출된 모델과 측정값을 비교한 결과 최적 적합도를 기준으로 매우 근접하게 산포되어 도출된 모델이 국내 하천의 비유사량값을 산정하는 모델로서 매우 적합함을 알 수 있었다.

둘째, 기존에 사용되던 비유사량 산정공식과 도출된 모델의 비교분석을 수행하였다. 그 결과, 국외의 데이터를 기반으로 도출된 모델은 개발 배경 및 국내 하천 환경과의 차이로 인해 국내 하천 데이터 적용에 한계가 있었다. 이는 기존의 국외 공식 개발에 활용한 데이터의 유량범위와 국내 데이터의 유량범위가 상이하기 때문에 나타나는 결과라고 판단되며 또한 단순히 홍수사상의 크기뿐만 아니라 홍수 발생기간과 빈도가 유사 공급과 유사 이송 능력에 영향이 있음을 추정할 수 있다. 국내 하천의 데이터를 활용하여 도출된 기존의 공식의 경우도 활용 데이터 범위와 하천 환경 변화 등으로 인해 실측값과 큰 차이가 있음을 알 수 있었으며, 따라서 본 논문에서 도출된 모델이 국내 하천의 비유사량 추정을 위한 모델로써 매우 적합함을 입증하였다. 또한 비교적 최근에 개발된 모델의 경우 도출된 모델을 활용한 결과와 유사함을 알 수 있었으며, 이에 따라 환경적인 요인, 데이터 수집 가능 여부 등을 판단한 후 모델을 활용하여야 할 것으로 판단된다. 또한 Model Tree 기법 적용 시 일반적인 패턴을 벗어나는 이상적인 수치의 실측값이 존재하는 경우 모델 정확도에 영향을 미칠 수 있음을 반드시 유의하여야 한다.

최종적으로 본 연구에서 제안한 데이터 마이닝 기법에 의한 비유사량 산정 방법은 새로운 유사량 추정 방식으로 활용이 가능 할 것이며, 현재 사용하고 있는 기존의 방법들이 가지고 있는 낮은 신뢰성이나 불확실성 또는 한계성을 상당부분 개선시킬 수 있을 것으로 기대된다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2019R1C1C1009719)을 받아 수행된 연구임.

References

Ahn, S.J., and Lee, J.H. (1984). “A correlation of reservoir sedimentation and watershed factors.” *Water for Future*, KWRA, Vol. 17, No. 2, pp. 107-112.

Bhattacharya, B., Price, R., and Solomatine, D. (2007). “Machine learning approach to modeling sediment transport.” *Journal of Hydraulic Engineering*, Vol. 133, No. 4, pp. 440-450.

Einstein, H.A. (1950). *The bed-load function for sediment transportation in open channel flows.* US Department of Agriculture Washington DC.

Hydrological Survey Center (HSC) (2007). *Hydrological annual report in 2006*. Ministry of Maritime Affairs and Fisheries.

Hydrological Survey Center (HSC) (2008). *Hydrological annual report in 2007*. The Ministry of Land, Transport and Maritime Affairs.

Hydrological Survey Center (HSC) (2009). *Hydrological annual report in 2008*. The Ministry of Land, Transport and Maritime Affairs.

Hydrological Survey Center (HSC) (2010). *Hydrological annual report in 2009*. The Ministry of Land, Transport and Maritime Affairs.

Hydrological Survey Center (HSC) (2011). *Hydrological annual report in 2010*. The Ministry of Land, Transport and Maritime Affairs.

Hydrological Survey Center (HSC) (2012). *Hydrological annual report in 2011*. The Ministry of Land, Transport and Maritime Affairs.

Hydrological Survey Center (HSC) (2013). *Hydrological annual report in 2012*. Ministry of Land, Infrastructure and Transport.

Hydrological Survey Center (HSC) (2014a). *Hydrological annual report in 2013*. Ministry of Land, Infrastructure and Transport.

Hydrological Survey Center (HSC) (2014b). *Hydrological annual report in 2014*. Ministry of Land, Infrastructure and Transport.

Jain, S.K. (2001). “Development of integrated sediment rating curves using ANNs.” *Journal of hydraulic engineering*, Vol. 127, No. 1, pp. 30-37.

Jang, E.K. (2017). *Sediment discharge assessment for rivers using model tree in data mining*. Ph. D. dissertation, University of Myongji, pp. 17-30.

Julien, P.Y. (2018). *River mechanics (2nd ed.)*. Cambridge University Press, C.O., U.S..

Kang, W. (2019). *Geospatial Analysis of Specific Degradation In South Korea*. Ph. D. dissertation, Colorado State University,

C.O., U.S., pp. 48-72.

Kane, B. (2003). *Specific degradation as function of watershed characteristics and climatic parameters*. Ph. D. dissertation, Colorado State University, C.O., U.S.

Kane, B., and Julien, P. (2007). "Specific degradation of watersheds." *International Journal of Sediment Research*, Vol. 22, No. 2, pp. 114-119.

Kang, W., Yang, C.Y., Lee, J., and Julien, P.Y. (2019). "Sediment yield for ungauged watersheds in South Korea." *Journal of Civil Engineering*, KSCE, Vol. 23, No. 12, pp. 5109-5120.

Korea Institute of Civil Engineering and Building Technology (KICT) (1992). *Research of sediment yield rate for dam design*. Ministry of Transportation.

Lin, B., and Namin, M.M. (2005). "Modelling suspended sediment transport using an integrated numerical and ANNs model." *Journal of Hydraulic Research*, Vol. 43, No. 3, pp. 302-310.

Minister of Land, Transport and Maritime Affairs (MLTMA) (2011). *Dam design manual and analysis*. The Ministry of Land, Transport and Maritime Affairs.

Nagy, H.M., Watanabe, K.A.N.D., and Hirano, M. (2002). "Prediction of sediment load concentration in rivers using artificial neural network model." *Journal of Hydraulic Engineering*, Vol. 128, No. 6, pp. 588-595.

Quinlan, J.R. (1992). "Learning with continuous classes." *Proceedings, 5th Australian joint conference on artificial intelligence*, Singapore, pp. 343-348.

Ryu, H.J., and Kim, S.W. (1976). "Study on sedimentaion in reservoir." *Journal of Hydrological Sciences*, Korean Association of Hydrological Sciences, Vol. 9, No. 2, pp. 67-75.

Ryu, S.C., and Min, B.H. (1975). "A study for sedimentation in reservoir-on district of Chin Young-." *Magazine of the Korean Society of Agricultural Engineers*, The Korean society of Agricultural Engineers, Vol. 17, No. 3, pp. 46-53.

Strahler, A.N. (1952). "Hypsometric (area-altitude curve) analysis of erosional topography." *Bulletin of the Geological Society of America*, Vol. 63, No. 11, pp. 1117-1141.

Suh, S.D., Lim, H.I., Cheon, M.B., and Yoon, K.D. (1988). "Regression analysis between specific sediments of reservoirs and physiographic factors of watersheds." *Magazine of the Korean Society of Agricultural Engineers*, The Korean society of Agricultural Engineers, Vol. 30, No. 4, pp. 45-61.

Wang, Y., and Witten, I.H. (1996). *Induction of model trees for predicting continuous classes*. Working Papaers, University of Waikato, Hamilton, New zealand.

Yoon, Y.J. (2011). *Devleopment of Prediction Formulae for Sediment Rate of Multipurpose Dams*. Ph. D. dissertation, Inha University, pp. 91-111.

Yoon, Y.N. (1982). "Estimation of silting load and capacity loss rater of irrigation reservoirs in Korea." *Fourteenth International Congress on Large Dam*, Rio de Janeiro, Brazil, pp. 175-188.