



## 식생여과대 유사 저감 효율 산정을 위한 정규화 방안

### A Study on Regularization Methods to Evaluate the Sediment Trapping Efficiency of Vegetative Filter Strips

배주현<sup>a</sup> · 한정호<sup>b</sup> · 양재익<sup>c</sup> · 김종건<sup>d</sup> · 임경재<sup>e</sup> · 장원석<sup>f, †</sup>

Bae, JooHyun · Han, Jeongho · Yang, Jae E · Kim, Jonggun · Lim, Kyoung Jae · Jang, Won Seok

#### ABSTRACT

Vegetative Filter Strip (VFS) is the best management practice which has been widely used to mitigate water pollutants from agricultural fields by alleviating runoff and sediment. This study was conducted to improve an equation for estimating sediment trapping efficiency of VFS using several different regularization methods (i.e., ordinary least squares analysis, LASSO, ridge regression analysis and elastic net). The four different regularization methods were employed to develop the sediment trapping efficiency equation of VFS. Each regularization method indicated high accuracy in estimating the sediment trapping efficiency of VFS. Among the four regularization methods, the ridge method showed the most accurate results according to  $R^2$ , RMSE and MAPE which were 0.94, 7.31% and 14.63%, respectively. The equation developed in this study can be applied in watershed-scale hydrological models in order to estimate the sediment trapping efficiency of VFS in agricultural fields for an effective watershed management in Korea.

**Keywords:** Elastic net; LASSO; regularization methods; ridge regression; vegetative filter strips

#### 1. 서 론

토지 재활과 복원은 토양이 인간에게 제공해주는 생태계 서비스를 회복하기 위해서 무엇보다도 중요하기에 최근 들어 토양관리에 더욱 관심을 가지게 되었다 (Keesstra et al., 2018; Zhang et al., 2016). 토양과 그에 수반되는 영양소 손실은 토양의 질을 떨어뜨리고 지구 기후 변화의 맥락에서 증가할 것으로 예상된다 (Borrelli et al., 2017). 토양유실은 주로 농경지에서 발생하고 있으며, 이를 저감하기 위해 구조적, 비구조적 최적관리기법들이 개발되고 적용되어 왔다. 이 중 식생여과

대 (vegetative filter strips, VFS)와 같은 일부 자연 기반 솔루션은 토양 및 양분 유출저감에 효과적이고 친환경적이며 지속 가능하면서도 경제적이기 때문에 많은 관심과 연구가 이루어지고 있다 (Golkowska et al., 2016; Keesstra et al., 2018). 식생여과대란 농경지에서 강우 발생 시에 일어날 수 있는 유사를 저감하기 위한 기법으로 경작지 말단부에 띠 모양으로 설치된 식생대를 말한다 (Schmitt et al., 1999).

최적의 식생여과대 면적을 산정하는 것은 효율적인 식생여과대 설계를 위해서 무엇보다도 중요하다 (Park and Hyun, 2014; Park et al., 2007). 식생여과대가 필요 이상으로 크게 설계된다면 작물 재배 면적이 감소하여 생산량이 감소하고 반대로 식생여과대를 너무 작은 규모로 설치한다면 기대 이하의 유사저감 효율을 나타내게 되어 수질 향상에 큰 도움이 되지 못한다 (Han et al., 2015).

최근 비점 오염원 연구에서 다방면으로 활용되고 있는 SWAT (Soil and Water Assessment Tool) 모형 내에서는 식생대에 의한 유사저감 효율을 모의하는데 있어서 식생여과대의 폭만 고려하기 때문에 유사 저감 효율 모의에 있어서 다소 한계가 있을 것으로 판단되어 보다 개선된 식을 구하기 위한 여러 연구들이 이루어졌다. Park et al. (2008)은 식생여과대의 폭과 유출량을 이용하여 유사저감 효율을 산정하였으며, 더 확장하여 유출수량, 투수계수, 단위 식생여과대 면적당 유입 유사량을 이용하여 효율식을 유도하였다 (White and Arnold,

<sup>a</sup> Senior Researcher, Korea Water Environment Research Institute

<sup>b</sup> PhD Student, Department of Regional Infrastructure Engineering, Kangwon National University

<sup>c</sup> Professor, Department of Biological Environment, Kangwon National University

<sup>d</sup> Professor, Department of Regional Infrastructure Engineering, Kangwon National University

<sup>e</sup> Professor, Department of Regional Infrastructure Engineering, Kangwon National University

<sup>f</sup> SILC Fellow (PhD), Sustainability Innovation Lab at Colorado (SILC), University of Colorado at Boulder

† **Corresponding author**

Tel.: +1-303-735-7439 Fax: +1-303-492-6187

E-mail: won.jang@colorado.edu

Received: October 24, 2019

Revised: October 25, 2019

Accepted: October 29, 2019

2009). Han et al. (2015)은 단위 시험포 면적 당 발생하는 유사량과 시험포 면적에 대한 식생여과대의 면적비를 활용하여 식생여과대의 유사저감 효율을 산출한 바가 있다. 또한 다양한 토양 유형과 강수 시나리오에 대한 유사저감 효율을 계산함으로써 인해 실제 현장에서 유사량을 제어할 수 있는 식생여과대를 설계하기에 최적화된 모델을 개발하기도 하였다 (Khatavkar and Mays, 2017).

본 연구에서는 국내 경작지를 대상으로 하는 식생여과대의 유사저감 효율을 모의하기 위한 효과적인 식을 개발하기 위하여, 일반적인 최소제곱법 (Ordinary Least Squares, OLS) 뿐만 아니라 라쏘 (Least Absolute Shrinkage and Selection Operator, LASSO), 능형 회귀분석 (ridge regression analysis), 엘라스틱 넷 (Elastic Net)의 정규화 방법론들을 이용하여 토양 침식 방지와 통제, 비점오염원 오염 통제 및 유역의 지역 관리 평가에 사용되고 있는 SWAT에서 사용되는 유사저감 효율식 보다 한층 더 현실성 높은 식을 도출하고자 한다.

이를 위해 국내 경작지 특성을 고려한 여러 가지 시나리오를 설정하여 VFSSMOD-W (Vegetative Filter Strip Modeling System) 모형을 실행하고 그 입력장과 실행 결과 중 유사저감 효율식을 산출하는 데 민감한 인자들을 추가로 선정하여 식생여과대 유사저감 효율을 계산할 수 있도록 학습을 시키고 테스트 자료들을 통한 검증 후 우수성이 보이는 회귀방법을 적용하였다.

단순히 식생여과대의 폭, 시험포로부터의 유출량 등 특정 요인에 의해 계산된 유사저감 효율이 아닌, 식생여과대 유사저감 효율에 영향을 주는 다양한 인자들을 고려한 개선된 식을 개발하고자 한다. 본 연구에서와 같은 시도는 과적합 (over fitting)을 방지해 줄 수 있는 정규화 과정을 거친 회귀식을 활용하여 국내 각 지역에 설치될 식생여과대의 현장 적용에 도움을 주고 SWAT 모형에서 모의되고 있는 식생여과대 유사저감 효율 예측의 한계를 극복하고 다양한 사례에 적용할 수 있고자 함이다.

## II. 재료 및 방법

### 1. 모델 개요 및 입력자료 구축

#### 가. SWAT 모형 내 식생여과대 모듈

SWAT 모형은 미국 농무성 농업연구소 (USDA Agriculture Research Service, ARS)에서 개발한 유역단위 모형으로 기상 자료 (일 강수량, 기온, 풍속, 일조량, 상대습도)와 토지이용도, 토양 속성 그리고 수치표고모델 (DEM)을 입력자료로 하여 장기 유출, 유사 및 비점오염원의 거동 등을 모의한다. SWAT 모형의 활용은 다양하며, 식생여과대의 유사저감 효율을 모

의하기 위해서도 많이 사용되고 있으며, 기존 SWAT 모형 내 VFS 모듈에 사용되었던 유사저감 효율 산정식을 식 (1)에 제시하였다 (Bosch et al., 2013). 살펴보는 바와 같이, 식생여과대의 유사저감 효율에 영향을 미치는 다양한 인자에 대한 고려 없이 단순히 식생여과대의 폭만을 이용하여 유사저감 효율을 산정한다는 문제점이 발생한다.

$$trap_{ef} = 0.367(width_{filtstrip})^{0.2967} \quad (1)$$

여기서,  $trap_{ef}$ 는 유사 저감 효율 (%)이며,  $width_{filtstrip}$ 는 식생여과대의 폭으로 VFS모듈에서는 이 값으로 저감효율을 산정하고 있다.

#### 나. VFSSMOD-W 모형의 개요

VFSSMOD-W 모형은 시험포 단위 모형으로 식생여과대에서의 유출, 침투, 유사저감 효율을 모의할 수 있으며, 이미 다양한 현장실험을 통해 모형의 정확도가 검증되었다 (Muñoz-Carpena et al., 1999).

VFSSMOD-W 모형은 설계 강우에 의한 단위 유출량을 이용하며, 다양한 조건의 식생여과대를 모의할 수 있도록 식생의 간격, 길이, 경사도, 침투 특성 등의 식생여과대 관련 인자뿐만 아니라 토성과 토양 입자 크기 등 다양한 인자를 고려한 분석이 가능하다. 이 모형은 시험포 단위 모형으로 식생여과대에서의 유출, 침투, 유사저감 효율을 모의할 수 있으며, 다양한 현장 실험을 통해 모형의 정확도를 검증한 바가 있다.

Fig. 1에서 시험포와 식생여과대의 간략한 모식도를 살펴볼 수 있으며, VFSSMOD-W 모형의 입력자료를 살펴보면, 시험포의 유출 및 유사 발생량 모의를 위해서는 강우량, 강우지속시간, CN (Curve Number) 등 강우사상 및 유출에 영향을 주는 인자, 시험포의 길이, 경사도, 면적 등의 시험포의 지형적 특성 인자, 토성, 입자크기 등 시험포의 토양관련 인자가 있다. 식생여과대로 유입된 유출수와 유사에 대한 저감효율을 모의하기 위한 입력자료는 식생여과대의 폭, 면적, 경사도, 식생의 간격과 길이, 조도계수 등이 있으며, 이를 활용하여 식생여과대에서 유출되는 유출수와 유사량을 모의한다.

VFSSMOD-W 모형은 Infiltration 모듈, Kinematic overland flow 모듈, Sediment filtration 모듈로 구성되어있다 (Fig. 2). Infiltration 모듈은 Green-Ampt 침투 방정식을 이용하여 침투량 및 초과유출을 산정하며, Kinematic overland flow 모듈은 침투층의 유출심과 유출량을 계산한다. Sediment filtration 모듈은 식생여과대로 유입된 유사의 이동과 퇴적을 모의한다 (Muñoz-Carpena and Parsons, 2014).

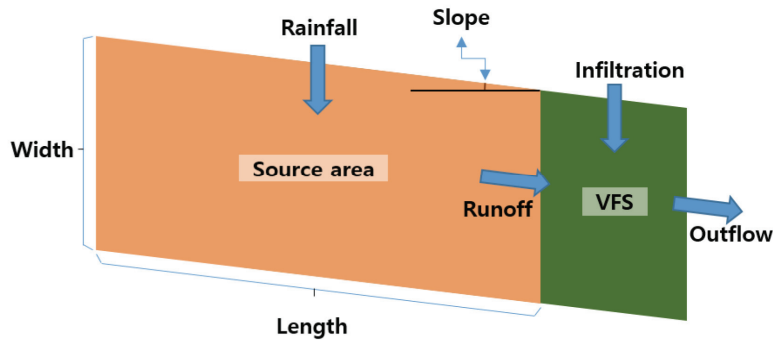


Fig. 1 Source area and vegetative filter strip in VFSMOD-W

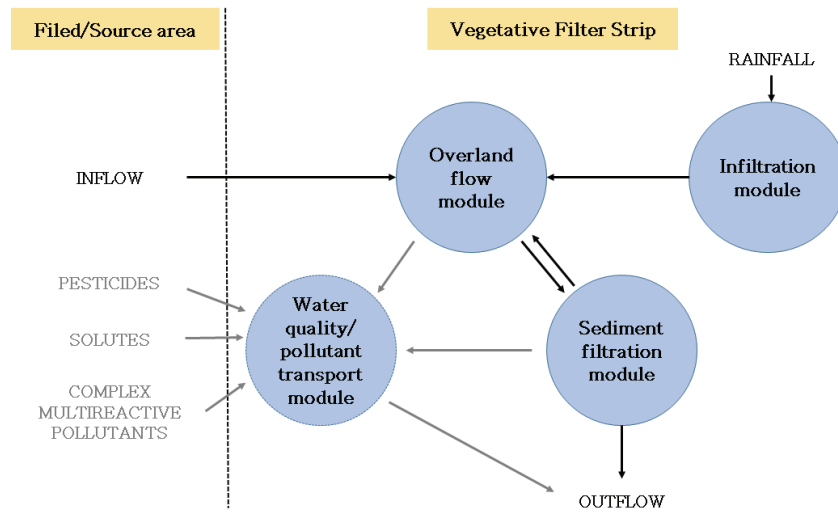


Fig. 2 Schematic diagram of VFSMOD-W

다. 회귀분석 입력자료

국내 경작지 특성 조사 결과를 반영하여 VFSMOD-W 모의를 위한 시나리오를 설정하였다 (Table 1). 강우량과 강우 지속시간은 한국 확률강우량정보에서 제공하는 확률강우량을 사용하여 60분 강우 지속시간의 10년, 20년, 30년 확률강우량에 해당하는 31 mm, 57 mm, 67 mm로 설정하였으며, 토양도, 토지이용도 그리고 식생피복 처리 상태 및 수문학적 조건 등 다양한 조건으로 유효우량 발생에 미치는 영향을 나타내는 CN는 60, 74, 86으로 선정하였다. 이때, CN 값은 국내 경작지 배수특성과 선형강우조건에 따른 작물별 CN값에 대한 기존의 문헌을 참고하여 설정하였다 (Han et al., 2015; Woodward et al., 2003). 경사도는 국내 경작지 경사도 분포 현황을 고려하여 완만한 경작지에서 급경사지 조건까지 고려하기 위하여 2~20% 넓은 범위를 설정하였다. 또한, 추후 SWAT 모형 적용 시 본 연구에서 회귀분석을 위해 선정한 입력자료들이 SWAT 모형의 어떠한 매개변수와 관련되었는

지 참고할 수 있도록 Table 1에 SWAT 모형 내 관련 매개변수명을 함께 제시하였다.

토양 관련 인자 중 하나인 토성은 국내 경작지에서 가장 많은 양을 나타낸 양토 (loam)와 사양토 (sandy loam)를 고려하였으며 회귀 모델 생성을 위한 학습에서는 각각 1과 0의 정수값으로 변환하여 입력하였다. 시험포에서 유사 발생 산정을 위한 인자인 USLE 토양보존대책인자 (P-factor)는 1.0을, 다양한 작물을 고려할 수 있도록 토양피복인자 (C-factor)는 0.1, 0.3, 0.5를 사용하였으며, 시험포 면적에 대한 식생여과대 면적 비 (AR)는 0.5~10%를 입력 인자로 사용하였다. 초종은 국내 경작지를 대상으로 잔디를 사용하였다 (Choi and Jang, 2014; Han et al., 2015).

그 외 회귀식 도출을 위한 입력변수로 VFSMOD-W 실행 후 계산되는 식생여과대로 유입되는 유수량 (runoff loading) 과 유사량 (sediment loading)도 회귀 분석 모델의 입력자료로 설정하였다. Table 1에 회귀식 도출을 위한 학습 속성 이름

**Table 1** Description of inputs in the VFSMOD-W and regression models

Model parameter	Attribute name (Notation)	Parameters of SWAT	Used values	Number of model parameter phase	Total scenario number
Rainfall (mm/60min)	rf	PRECIPITATION	31, 57, 67	3	3×3×2×10×1 ×3×11×9×1 = 53,460
CN	cn	CN2	60, 74, 86	3	
Soil type	st	TEXTURE	0 (Sandy loam), 1 (Loam)	2	
Slope (%)	sp	HRU_SLP	2, 4, 6, 8, 10, 12, 14, 16, 18, 20	10	
USLE P-factor	pf	USLE_P	1	1	
USLE C-factor	cf	USLE_C	0.1, 0.3, 0.5	3	
Ratio of VFS area to source area (%)	rv	VFSRATIO	0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	11	
Source area (ha)	sa	HRU_FR	0.05, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2	9	
Vegetation	vg	-	Turf grass	1	
Runoff loading (m <sup>3</sup> )	rl	-	0-12.6	VFSMOD-W outputs	
Sediment loading (g)	sl	-	16.8-5,932,000		

(attribute name)을 같이 제시하였으며 토양보존대책인자와 식생의 경우는 한가지 값을 가지기 때문에 설명변수로는 의미가 없고 나머지 9개의 변수를 이용하여 입력자료를 구성하였다.

정리하면, 식생여과대의 유사 저감 효율을 계산하는 회귀식을 개발하기 위해 VFSMOD-W 모형의 입력과 출력자료를 활용하였으며, 선형회귀식 산정을 위한 입력자료로는 확률강우량 (rf), CN (cn), 경사도 (sp), 토성 (st), 토양피복인자 (cf), 시험포 면적 (sa), 시험포 면적에 대한 식생여과대 면적 비 (rv), 유수량 (rl), 유사량 (sl)으로 총 9개이다.

## 2. 정규화 선형회귀 분석

회귀분석은 데이터에 포함된 한 변수와 또 다른 하나 이상의 변수들 사이의 관계를 함수식으로 표현하여 데이터를 간편하게 분석하는 기법이며, 예측값과 실제값 오차의 제곱합을 최소화 하는 일반적인 회귀분석 방법인 최소제곱법의 문제점을 해결하는 방법으로 회귀계수 수축 (coefficient shrinkage)과 변수선택 (subset selection)의 방법이 있다.

정규화(regularized) 선형회귀 방법은 선형회귀 계수(weight)에 대한 제약 조건을 추가함으로써 모형이 과도하게 최적화되는 현상을 막는 방법이다. 모형이 과도하게 최적화되면 모형 계수의 크기도 과도하게 증가하는 경향이 나타난다. 따라서 정규화 방법에서 추가하는 제약 조건은 일반적으로 계수의 크기를 제한하는 방법이며, 라쏘, 능형 회귀분석, 엘라스틱 넷 방법론의 3가지 방법이 사용된다.

본 연구에서는 파이썬 (python)의 사이킷-런 (scikit-learn) 패키지에서 제공하는 linear\_model 클래스 (class) 내 4가지 회

귀 함수인 LinearRegression, Lasso, Ridge, ElasticNet을 사용하였으며, 학습과 테스트를 위한 자료 분리를 위해서는 사이킷-런에서 제공하는 model\_selection 클래스의 train\_test\_split 함수를 이용하여 랜덤 (random) 자료의 80%를 추출한 후, 학습하고 모델을 생성하여 나머지 20% 자료로 테스트하고 검증한다. 테스트 세트에서 모델을 평가함으로써 오차에 대한 추정값을 얻으며, 이 값은 이전에 본 적이 없는 새로운 샘플에 모델이 얼마나 잘 작동할지 알 수 있다.

### 가. 라쏘의 개요

라쏘는 가중치의 절대값 합을 최소화하는 것을 추가적인 제약 조건으로 하며, 회귀계수를 전체적으로 축소하여 예측 모형의 정확도를 높이고, 동시에 변수선택의 예측에 필요한 중요한 변수만을 선택해 차원축소를 통한 예측 모형의 해석력을 증가시켜준다 (Park, 2013; Robert, 1996).

라쏘의 회귀계수 추정량을 구하는 식 (2)는 다음과 같다.

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^P \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^P |\beta_j| \quad (2)$$

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^P \beta_j x_{ij})^2 \text{ subject to } \sum_{j=1}^P |\beta_j| \leq t$$

식 (2)에서  $n$ 은 관측 개체의 수,  $P$ 는 설명 변수의 수,  $y_i$ 는 설명 변수의  $i$  번째 관찰치,  $\beta_1, \dots, \beta_p$ 는 회귀계수를 의미하며 설명 변수  $x_{ij}$ 의 척도 (scale)에 의존하기 때문에 이 문제를 해결하기 위해  $x_{ij}$ 는 표준화된 값을 사용한다.  $\beta_0$ 의 추정량은  $y$ 절편이기 때문에 의미를 두지 않는다.  $\lambda_1 \geq 0$ 은 회귀 계수

크기의 축소량을 조절하는 조율 모수 (tuning parameter)이다. 위 식에서  $\lambda_1$ 은  $t$ 로 표현되는데, 제약조건인  $t$  값이 작아질수록 영향력이 없는 변수의 회귀계수를 0으로 만들어 차원을 축소하여 변수 선택이 가능하며,  $t$  값이 커질수록 회귀계수에 대한 제약 (penalty)을 주지 않게 되어 선형 회귀분석과 같은 결과를 갖게 된다 (Kim, 2016; Hastie et al., 2007).

나. 능형회귀의 개요

능형회귀는 가중치들의 제곱합 (squared sum of weights)을 최소화하는 것을 추가적인 제약 조건으로 하였다. 즉 설명 변수들 사이에 높은 상관관계가 존재할 경우에는 분산, 공분산 행렬의 행렬식이 0에 가까운 값이 되어 회귀계수의 추정량이 매우 나빠지게 된다. 능형 회귀분석은 그런 경우 편향 (bias) 오류를 어느 정도 허용하여 전체적인 회귀계수의 크기를 축소함으로써 다중공선성 (multicollinearity)의 문제를 해결한 방법이다 (Hoerl and Kennard, 2000).

능형 회귀분석의 회귀계수 추정량을 구하는 식 (3)은 다음과 같다.

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^P \beta_j x_{ij})^2 + \lambda_2 \sum_{j=1}^P |\beta_j|^2 \quad (3)$$

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^P \beta_j x_{ij})^2 \text{ subject to } \sum_{j=1}^P |\beta_j|^2 \leq t$$

식 (3)에서  $n, P, y_i, x_{ij}, \beta_0$ 은 라쏘와 마찬가지로  $\lambda_2 \geq 0$ 은 회귀 계수 크기의 축소량을 조절하는 조율 모수이다. 위 식에서  $\lambda_2$ 는  $t$ 로 표현되는데,  $t$  값이 커질수록 회귀계수의 축소량이 많아지며 회귀계수는 0에 가까이 도달한다. 이와 같이 능형 회귀분석은  $\lambda_2$  제약 조건을 이용하여 회귀계수의 크기를 축소함으로써 예측정확도를 높인 방법이다. 하지만 능형 회귀분석은 차원 축소가 불가능하기 때문에 모든 변수가 모형에 포함되어 모형에 대한 해석력이 떨어진다는 단점이 존재한다.

다. 엘라스틱 넷의 개요

엘라스틱 넷은 라쏘의 문제점을 극복하기 위하여 능형 회귀분석의 좋은 특징을 결합하여 고안된 정규화 회귀 방법이다. 즉, 가중치의 절대값 합과 제곱합을 동시에 제약 조건으로 가지는 모형이다 (Kyung, 2014; Zou and Hastie, 2005).

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^P \beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^P |\beta_j| + \lambda_2 \sum_{j=1}^P |\beta_j|^2 \quad (4)$$

엘라스틱 넷은 식 (4)에서처럼 라쏘의 제약과 능형 회귀분석의 제약을 결합한 방법이다. 라쏘의 영향력이 없는 변수의 회귀계수를 0으로 만들어 차원을 축소하여 변수선택이 가능한 장점과 능형회귀 분석의 전체적인 회귀계수의 크기를 축소함으로써 관련성이 높은 설명 변수가 있을 때 변수들을 그룹화하여 다중공선성의 문제를 해결한 장점 두 가지를 동시에 만족하는 방법이다.

$$(1 - \alpha) \sum_{j=1}^P |\beta_j| + \alpha \sum_{j=1}^P |\beta_j|^2 \leq t \quad (5)$$

식 (4)에서  $\lambda$ 를  $\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$ 로 치환하면, 엘라스틱 넷의 페널티 항은 식 (5)처럼 표현될 수 있다. 따라서 엘라스틱 넷에서  $\alpha=0$ 이면 능형회귀,  $\alpha=1$ 이면 라쏘와 같으므로, 엘라스틱 넷은  $0 \leq \alpha \leq 1$ 에 해당하며, 라쏘와 능형 회귀분석의 볼록 결합 (convex combination)에 해당하기 때문에 능형 회귀분석과 라쏘의 비율을 조정할 수 있다 (Zou and Hastie, 2005).

III. 결과 및 고찰

1. 민감도 실험 결과

먼저 회귀식 산출을 위해 선정된 9개 속성들간의 상관관계 (correlation)를 히트 맵 (heat map)으로 살펴보았다 (Fig. 3). 그림에 제시된 수치는 각 속성들간의 상관관계 값으로 식생 여과대 유입 유수량 (sl)이 -0.56으로 유사저감 효율 (ste)에 가장 큰 영향력을 주었고, 다음으로 식생여과대 면적 비 (rv)는 0.39로 양의 상관관계를, 유입 유수량 (rl)은 -0.38, 경사도 (sp)는 -0.36의 음의 상관관계를 나타내었다.

유사저감 효율 (ste) 외에 각 설명 변수들간의 상관관계를 살펴보면, 식생 여과대 유입 유수량 (sl)과 유입 유수량 (rl)이 서로 상관성이 0.58로 높을 뿐만 아니라 시험포 면적 (sa)과 높은 상관성을 보였고 그 값은 각각 0.41, 0.66으로 나타났다. 또한 CN (cn)과는 0.27, 0.48의 값을 보임으로써 회귀식에서 독립변수들 간에 강한 상관관계가 나타나는 문제인 다중공선성이 존재한다고 볼 수 있다. 그 외 각 속성에 대한 자세한 설명은 Table 1을 참고하면 되고 Fig. 3에서 각 속성들간의 상관성을 동시에 분석할 수 있었다.

각 속성들이 유사저감 효율에 주는 영향력은 크지 않지만, 속성에 따른 민감도 분석을 통해 모든 변수들을 입력변수로 사용하는 것이 유사저감 효율식을 산출하는데 유의함을 알 수 있었으며, 유사저감 효율을 더 잘 계산해 낼 수 있는 식을 얻기 위해서 각 속성값의 로그 변환 및 다차항으로의 변환을

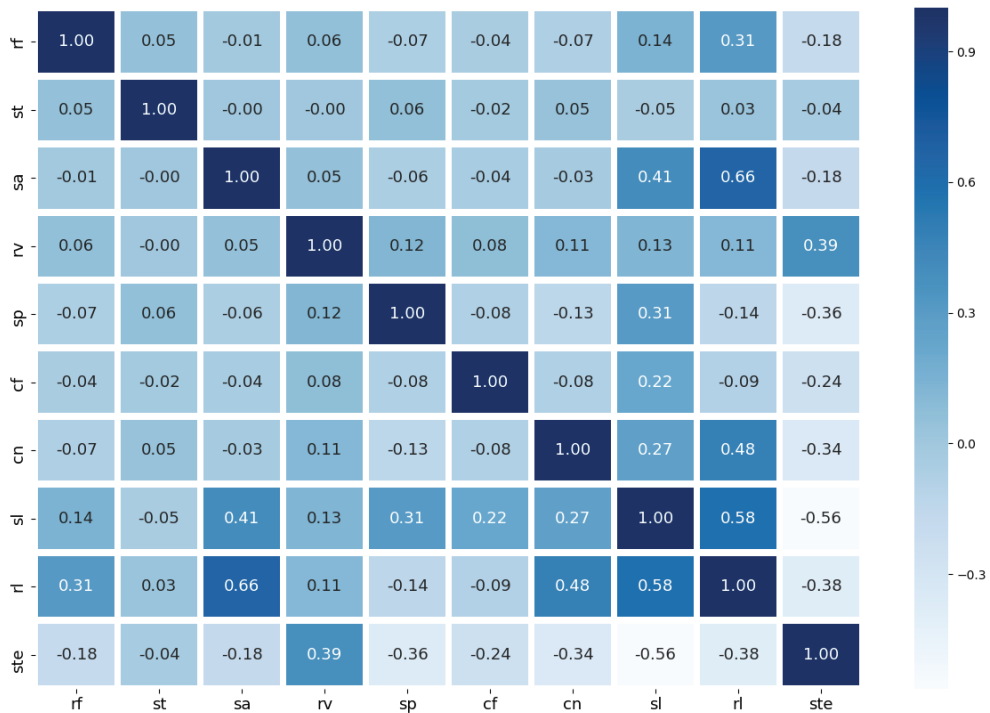


Fig. 3 Heat map to analyze correlation coefficients of model attributes and sediment trapping efficiency

Table 2 Regression accuracy of training data sets using converted variables

Converted Variable	OLS	LASSO	Ridge Regression	Elastic Net
All	0.942	0.941	0.942	0.912
Remove $rf^3$	0.941	0.940	0.941	0.911
Remove $st$	0.941	0.939	0.940	0.911
Remove $sp^3$	0.941	0.940	0.941	0.912
Remove $cf^3$	0.942	0.940	0.941	0.911
Remove $[\log(rv)]^5$	0.942	0.940	0.942	0.911
Remove $[\log(sa)]^3$	0.941	0.940	0.941	0.911
Remove $rl^5$	0.941	0.940	0.941	0.911
Remove $sl$	0.936	0.935	0.936	0.904
Add $rf^4$	0.941	0.940	0.941	0.911
Add $sp^4$	0.941	0.940	0.941	0.912
Add $cf^4$	0.941	0.940	0.941	0.912
Add $[\log(rv)]^6$	0.941	0.940	0.941	0.912
Add $[\log(sa)]^4$	0.942	0.940	0.941	0.912
Add $rl^6$	0.942	0.941	0.942	0.912
$[\log(rv)] \rightarrow rv$	0.941	0.939	0.941	0.911
$[\log(sa)] \rightarrow sa$	0.941	0.939	0.940	0.910

통하여 설명변수 개수를 25개로 확장하고 최종식인 식 (9)에서와 같은 선형 회귀식을 산정하였다. 25개에 대한 변수는 식에서 살펴볼 수 있다.

선정된 설명변수에 대해서 차수와 로그변환 등에 따른 결과에 대한 민감성을 실험하여 그 변수들을 선정하였고 그 결

과에 대한 간략한 요약은 Table 2에 나타내었다. 변환된 변수를 제거하거나 추가항을 설정하고 나머지 변수들은 기존과 동일하게 입력하여 학습자료에서의 회귀식 정확도를 살펴봄으로써 본 연구에서 선정한 변수들의 타당성을 입증하고자 하였다.

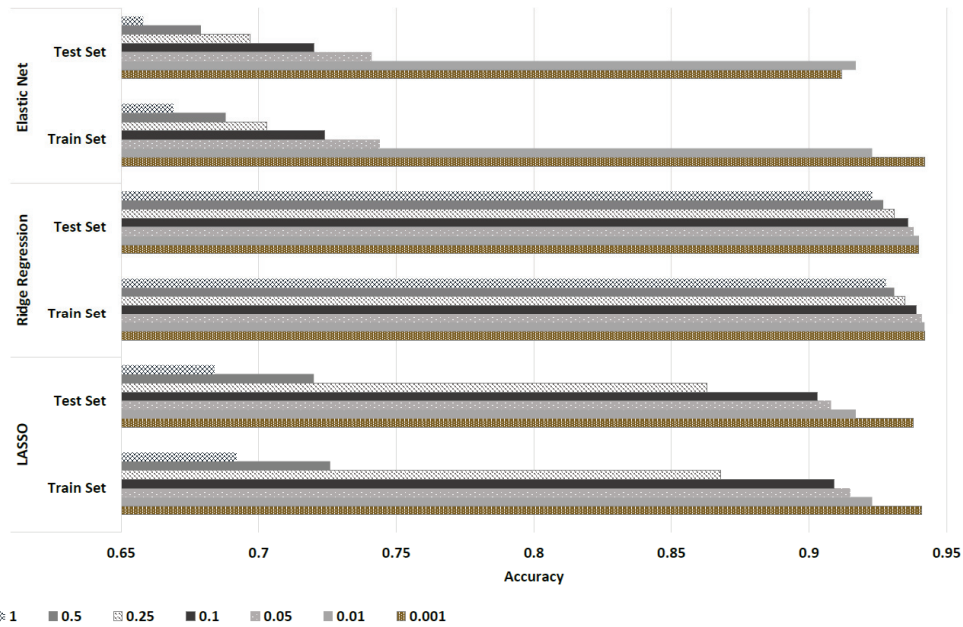


Fig. 4 Verification of regression equations with variations in alpha values using regularization methods

본 연구에서 선형 회귀모형에서 중요한 초매개변수 (hyperparameter)인 알파 ( $\alpha$ )를 조정하는 실험이 이루어졌고, 모델을 생성하기 위해서 모델들을 조정 (tuning)할 때 설정해야 되는 초매개변수의 경우는 매개변수 (parameter)와 달리 알고리즘 구현 과정에서 사용되는 값으로 경험에 의해 정해지며 알고리즘을 여러 번 수행해보면서 최적의 값을 찾게 된다. 초매개변수인 알파는 모델을 얼마나 많이 규제할지 조절하는 값으로 알파가 아주 크게 되면 모든 가중치가 거의 0에 가까워지고, 결국 데이터의 평균을 지나는 수평선이 되는 값으로 Fig. 4에 학습과 테스트 자료들을 이용하여 알파값의 설정에 따른 정규화 회귀 분석 결과를 살펴보았다.

라쏘와 능형회귀에서는 알파값이 0.001을 선택하였을 때 가장 높은 정확도를 나타내었다. 하지만 엘라스틱 넷의 테스트 자료에 대한 검증의 경우는 알파값이 0.01일 때 정확도가 알파값 0.001 보다도 더 높은 정확도인 0.917를 보임으로 인해 정규화 방식으로 엘라스틱 넷을 사용할 경우에는 초매개변수인 알파값을 0.01로 설정함으로 인해 테스트 자료에 있어서 더 좋은 결과를 얻을 수 있을 것이다. 하지만 본 연구에서는 학습자료에서 가장 좋은 결과들을 얻었고 테스트 자료에서도 엘라스틱을 제외한 경우에는 모두 알파값 0.001에서 높은 정확도를 보임으로 인해, 정규화 모형 비교를 위해 동일한 알파 값 지정을 하고자 모두 알파값을 0.001로 지정하고 학습하였다.

## 2. 모의 결과

라쏘 결과에서 기존 변수들 중 2개만 제거되면서 분석결과 의 큰 변화는 찾아볼 수 없었으며 마찬가지로 엘라스틱 넷의 경우도 최소제곱법 회귀나 능형 회귀 분석 보다 더 개선된 결과를 나타내지는 못하였다 (Table 3~4). 회귀식의 검증을 위해, 먼저 정확도는 각 모델의 score 함수를 이용하여 결정계수 (coefficient of determination,  $R^2$ )를 구했고, 모형 간에 예측 오차의 비교가 용이하며 그 신뢰도가 높아서 많이 사용되는 평균오차제곱근 (Root Mean Squared Error, RMSE), 절대평균 오차비율 (Mean Absolute Percentage Error, MAPE)을 사용하여 모형들간의 오차 정도를 비교하였다.

결정계수 정확도는 1에 가까울수록, 오차값들은 그 값이 작을수록 보다 정확한 예측력을 갖는 것으로 평가되며, 특히 절대평균오차비율은 모형 간에 예측오차의 비교가 용이하고, 그 신뢰도가 높아서 많이 사용되는 통계량으로 10% 미만이면 매우 정확한 예측을 나타내고, 10~20% 이면 비교적 정확한 예측을 나타내며, 20~50%이면 비교적 합리적인 예측임을 알 수 있고, 50% 이상이면 부정확한 예측이라고 할 수 있다 (Son and Park, 2012).

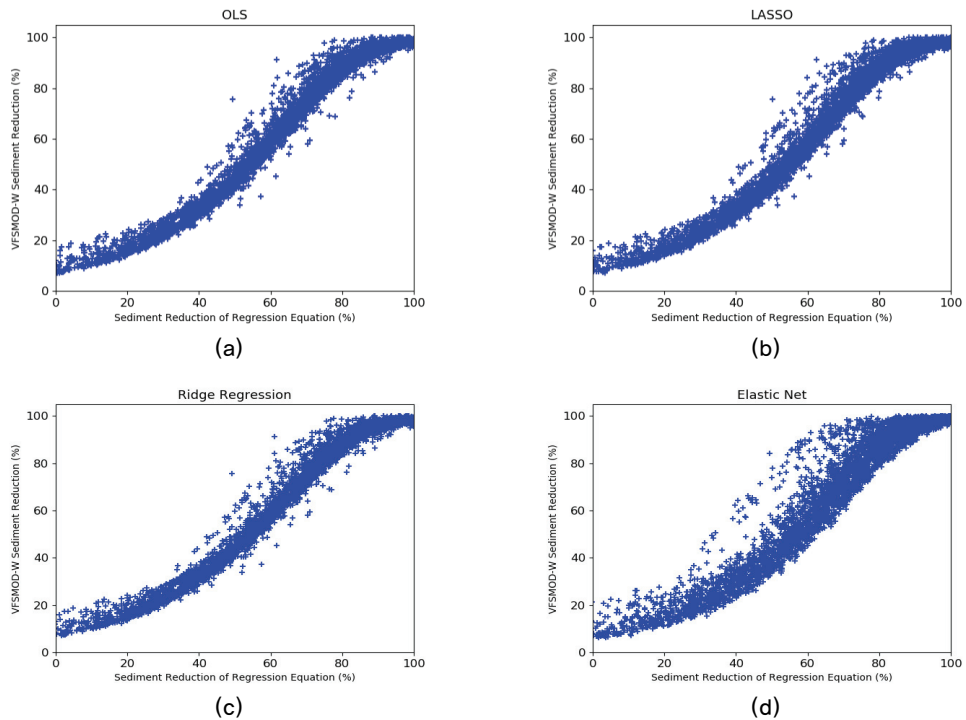
평가를 위한 정확도 결정계수와 오차값들은 식 (6)~(8)에서 유도될 수 있으며, 계산을 위하여 사이킷-런 metrics 라이브러리의 mean\_squared\_error 함수를 이용하여 평균오차제곱근과 절대평균오차비율을 얻었다.

**Table 3** The performance of regression methods

Method		OLS	LASSO	Ridge Regression	Elastic Net
R <sup>2</sup>	Train set	0.942	0.941	0.942	0.912
	Test set	0.940	0.938	0.940	0.908
Number of attributes used		25	23	25	25

**Table 4** Error results of regression methods

Method		OLS	LASSO	Ridge Regression	Elastic Net
Train set	RMSE (%)	7.24	7.31	7.25	8.91
	MAPE (%)	14.86	15.04	14.89	18.07
Test set	RMSE (%)	7.29	7.39	7.31	9.02
	MAPE (%)	14.62	14.82	14.63	17.77



**Fig. 5** Comparison of sediment trapping efficiencies by (a) OLS analysis, (b) LASSO, (c) ridge regression analysis, (d) Elastic Net and VFSMOD-W model with test data

$$R^2 = 1 - \frac{\sum(Y_t - F_t)^2}{\sum(Y_t - \bar{Y}_t)^2} \quad (6)$$

$$SE = \sqrt{\frac{\sum(Y_t - F_t)^2}{n}} \quad (7)$$

$$MAPE = \frac{1}{n} \left| \sum \left( \frac{Y_t - F_t}{Y_t} \right) \times 100 \right| \quad (8)$$

식에서  $Y_t$ 는 t번째의 실제값,  $F_t$ 는 t번째의 추정값이며  $n$ 은

자료의 총 개수이다.

Table 3에서 회귀 모형들의 결정계수 ( $R^2$ ) 정확도와 사용된 변수들의 결과를 정리하였으며, Table 4에서는 각 모형 추정 값들과 VFSMOD-W 유사저감 효율값들간의 오차를 나타내고 있다. 본 연구에서는 테스트 자료 검증에서 정확도 결정계수 0.94, 평균오차제곱근 7.31%, 절대평균오차비율 14.63%를 보인 능형회귀식을 추천하고자 하며, 능형회귀 분석 방법을 사용한다면 식생여과대의 유사저감 효율을 잘 추정해 낼 수 있으리라 기대된다.



Fig. 5에서는 각 모형들의 테스트 자료들을 이용하여 유사저감 효율을 계산한 값과 실제 VFSMOD-W에서 계산된 효율값들을 비교하여 그래프로 나타내었다. 최소제곱법과 능형회귀분석이 가장 VFSMOD-W의 효율과 유사한 추세를 나타내고 있다.

정규화 방안들 중 라쏘는 설명 변수 간에 높은 상관관계가 존재할 때는 극단적으로 큰 값을 가지는 변수 하나만 선택을 하고 나머지 값은 제거하게 되는 문제가 발생하고, 관측개체 수가 설명 변수의 수보다 작을 때에는 회귀계수의 최소자승 추정량을 구할 수 없어 분석이 불가하다는 단점이 존재한다 (Zou and Hastie, 2005). 그로 인해 사용된 변수는 타 회귀식보다 2개 적은 변수들을 사용하였다.

### 3. 식생여과대 유사저감 효율 산정식 개발

본 연구에서는 초기 설계와는 달리 식생여과대 유사저감 효율을 산정하는 속성들의 수가 적은 편이며 이로 인해 학습한 자료들을 과적합하는 현상이 발생하지 않음으로 인해 3가지 정규화 과정을 통한 페널티 부과 회귀 분석 과정이 크게 의미가 있지는 않았다. 하지만 차후 현장 적용 등에 있어서의 유용성을 위해, 정규화 회귀 방안들 중 가장 우수한 결과를 보인 능형 회귀식 모형을 이용하여 100% 모든 자료를 학습시켜서 아래와 같은 회귀식인 식 (9)를 도출하였다. 식 내부의 변수 즉 속성들은 확률강우량 (rf), 경사도 (sp), 토성 (st), 토양피복인자 (cf), 시험포 면적 (sa), 시험포 면적에 대한 식생여과대 면적비 (rv), 유수량 (rl), 유사량 (sl)이다. 토성의 경우는 양토와 사양토를 각 각 1과 0의 범주값을 부여해서 생성한 식이기에 차후 다른 토성의 값을 입력할 경우에는 이 사이값을 중심으로 분포할 수 있는 표준화과정의 전처리라 필요할 것으로 보인다.

차후, SWAT 모형의 VFS 모듈의 식생여과대 유사저감 효율식으로 본 연구에서 도출한 식을 사용한다면 효율적인 식생여과대의 현장 적용 및 국내 경작지를 대상으로 보다 정확한 식생여과대 유사저감 효율 모의가 가능할 것으로 판단된다. Fig. 6은 최종적으로 식생여과대 유사저감 효율을 위해 선정된 능형 회귀 모형으로 자료 전체를 모두 검증한 결과를 나타내었다.

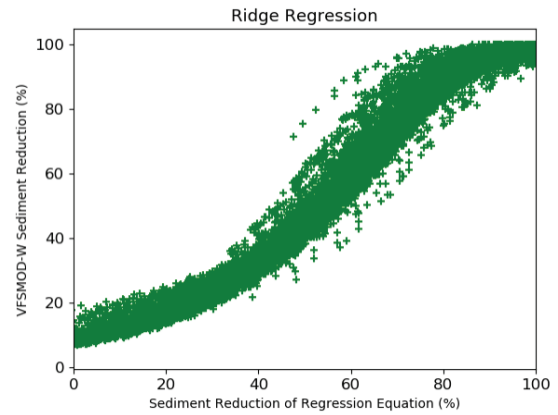


Fig. 6 Comparison of sediment trapping efficiencies by ridge regression analysis and VFSMOD-W model

## IV. 결 론

최근 들어 토양 오염 문제에 관심이 높아지고 있고, 보다 환경 친화적인 방법으로 비점원오염을 저감시키는 방법에 관심이 많아졌으며 식생여과대에 의한 토사 유출 저감과 오염저감 방안에 대한 연구 또한 그 중 하나라고 볼 수 있다. 이에 본 연구에서는 국내 경작지를 대상으로 식생여과대의 유사저감 효율을 모의할 수 있도록 SWAT 모형 내 VFS 모듈을 개선하기 위한 새로운 유사저감 효율 산정식을 개발하고자 하였다. 이를 위해 국내 경작지 특성을 반영하여 VFSMOD-W 모형 모의를 위한 다양한 조건을 고려한 시나리오를 설정하였으며, 모의한 결과를 바탕으로 식생여과대의 유사저감 효율 예측을 위한 새로운 식을 제시하였다.

식생여과대의 유사저감 효율에 영향을 주는 속성들인 확률강우량, CN, 토성, 경사도, 토양피복인자, 시험포 면적에 대한 식생여과대 면적비, 시험포 면적 그리고 VFSMOD-W에서 계산된 시험포로부터 식생대로의 유출 유수량과 유출 유사량을 변수로 설정하고 정규화 회귀식들을 유추하였으며 최소제곱법 회귀 분석, 능형 회귀 분석, 라쏘, 엘라스틱 넷 방법론들을 모두 적용해 보았다.

결론적으로 다양한 속성 선택 등의 민감도 실험에서도 최

Trapping Efficiency

$$\begin{aligned}
 = & -0.000129719198 \times rf^3 + 0.00622813827 \times rf^2 + 0.0000050454 \times rf \\
 & - 0.029934396 \times [\log(rv)]^5 - 0.118135582 \times [\log(rv)]^4 - 0.222686453 \times [\log(rv)]^3 \\
 & + 0.251341952 \times [\log(rv)]^2 + 19.9072392 \times [\log(rv)] + 34.0229978 \times st \\
 & - 0.458029222 \times [\log(sa)]^3 + 1.05890397 \times [\log(sa)]^2 + 5.34233173 \times [\log(sa)] \\
 & - 768.838609 \times sp^3 + 1644.7148 \times sp^2 - 638.878389 \times sp + 81.7548766 \times cf^3 \\
 & + 57.4707160 \times cf^2 - 130.570575 \times cf - 0.943700502 \times cn \\
 & - 0.00000608238593 \times sl - 0.00208462159 \times rl^5 + 0.0743983596 \times rl^4 \\
 & - 1.01861350 \times rl^3 + 6.88709084 \times rl^2 - 24.7987299 \times rl + 214.95001574223
 \end{aligned}$$

(9)

수제곱법 방식이 과대최적화를 발생시키지 않음으로 인해 정규화 방안들에 의한 패널티를 부과할 필요성이 없었다. 하지만 보다 일반성을 보장할 수 있는 회귀식을 찾는 시도는 중요하였으며, 더 우수한 회귀 결과를 얻기 위해서는 비선형 회귀 등의 개선된 방안이 추가적으로 진행되어야 될 것이지만 선형 회귀식을 제시할 수 있는 최선의 연구였다고 판단된다.

이에 본 연구를 바탕으로 차후 다른 자료들의 적용함에 있어서의 일반성을 보장하기 위하여 정규화 방안들 중 가장 우수한 결과를 보인 능형회귀 분석 모델을 사용하고자 하며, 최종적으로 전 자료를 사용하여 생성된 능형회귀 모형식을 제시하였다. 그 검증결과는 정확도 0.94 정도였지만 이는 정규화 방안으로 과적합되지 않고 어떠한 사례에서도 식생여과대에서의 유사 저감 효율을 잘 예측해 주리라 기대된다.

차후 연구로, 앞서 산정된 능형 회귀 모형식을 SWAT 내 VFS 모듈의 유사 저감효율 산정식으로 SWAT을 개선하고 이에 대한 적용성 평가 연구가 필요하다. 또한 개선된 SWAT으로 산출된 결과들의 정확도 검증을 위해서는 많은 현장 실험들의 관측값들을 이용하여 타당성을 입증해야 할 것이다.

## 감사의 글

본 연구는 환경부 “포토보전관리기술개발사업; 20190028 20002”으로 지원받은 과제임.

## REFERENCES

- Borrelli, P., D. A. Robinson, L. R. Fleischer, E. Lugato, C. Ballabio, C. Alewell, K. Meusburger, S. Modugno, B. Schütt, and V. Ferro, 2017. An assessment of the global impact of 21st century land use change on soil erosion. *Nature Communications* 8(1): 1-13. doi:10.1038/s41467-017-02142-7.
- Bosch, N. S., J. D. Allan, J. P. Selegean, and D. Scavia, 2013. Scenario-testing of agricultural best management practices in Lake Erie Watersheds. *Journal of Great Lakes Research* 39(3): 429-436. doi:10.1016/j.jglr.2013.06.004.
- Choi, K. S., and J. R. Jang, 2014. Selection of appropriate plant species of VFS (Vegetative Filter Strip) for reducing NPS pollution of uplands. *Journal of Korea Water Resources Association* 47(10): 913-983 (in Korean). doi: 10.3741/JKWRA.2014.47.10.973.
- Golkowska, K., B. Rugani, D. Koster, and C. Van Oers, 2016. Environmental and economic assessment of biomass sourcing from extensively cultivated buffer strips along water bodies. *Environmental Science & Policy* 57(1): 31-39. doi:10.1016/j.envsci.2015.11.014.
- Han, J. H., Y. Park, D. Kum, Y. G. Jung, G. C. Jung, K. S. Kim, and K. J. Lim, 2015. Modification of sediment trapping efficiency equation of VFS in SWAT considering the characteristics of the agricultural land in Korea. *Journal of Korean Society on Water Environment* 31(5): 482-490 (in Korean). doi:10.15681/KSWE.2015.31.5.482.
- Hastie, T., J. Taylor, R. Tibshirani, and G. Walther, 2007. Forward stagewise regression and the monotone lasso. *Electronic Journal of Statistics* 1: 1-29. doi:10.1214/07-EJS004.
- Hoerl, A. E., and R. W. Kennard, 2000. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 42(1): 80-86. doi:10.1080/00401706.2000.10485983.
- Keesstra, S., J. Nunes, A. Novara, D. Finger, D. Avelar, Z. Kalantari, and A. Cerdà, 2018. The superior effect of nature based solutions in land management for enhancing ecosystem services. *Science of The Total Environment* 610: 997-1009. doi:10.1016/j.scitotenv.2017.08.077.
- Khatavkar, P., and L. W. Mays, 2017. Optimization models for the design of vegetative filter strips for stormwater runoff and sediment control. *Water Resources Management* 31(9): 2545-2560. doi:10.1007/s11269-016-1552-y.
- Kim, S. B., 2016. Sentiment dictionary and polarity classification of blog review by using elastic net. Master. Seoul, Korea: Dongguk University (in Korean).
- Kyung, M. J., 2014. Comparison of laplace and double pareto penalty: LASSO and elastic net. *Korean Journal of Applied Statistics* 27(6): 975-989 (in Korean). doi:10.5351/KJAS.2014.27.6.975.
- Muñoz-Carpena, R., J. E. Parsons, and J. W. Gilliam, 1999. Modeling hydrology and sediment transport in vegetative filter strips. *Journal of Hydrology* 214(1-4): 111-129. doi:10.1016/S0022-1694(98)00272-8.
- Muñoz-Carpena, R., and J. E. Parsons, 2014. Vegetative filter strips modeling system: Model Document and User's Manual Version 6. Gainesville, FL, USA: Press of University of Florida
- Park, C. Y., 2013. Simple principal component analysis using Lasso. *Journal of the Korean data & Information Science Society* 24(3): 533-541 (in Korean).
- Park, Y. S. and G. W. Hyun, 2014. Optimization of vegetative filter strip using VFSSMOD-w model and

- Genetic-Algorithm. *Journal of Korean Society on Water Environment* 30(2): 159-165 (in Korean). doi:10.15681/KSWE.2014.30.2.159
16. Park, Y. S., J. G. Kim, N. W. Kim, K. S. Kim, J. D. Choi, and K. J. Lim, 2007. Analysis of sediment yields at watershed scale using area/slope-based sediment delivery ratio in SATEEC. *Journal of Korean Society on Water Environment* 23(5): 650-658 (in Korean).
  17. Park, Y., J. Kim, N. Kim, J. Park, W. S. Jang, J. Choi, and K. J. Lim, 2008. Improvement of sediment trapping efficiency module in SWAT using VFSSMOD-W model. *Journal of Korean Society on Water Environment* 24(4): 473-479 (in Korean).
  18. Robert, T., 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B Statistical Methodology* 58(1): 267-288.
  19. Schmitt, T. J., M. G. Dosskey, and K. D. Hoagland, 1999. Filter strip performance and processes for different vegetation, widths, and contaminants. *Journal of Environmental Quality* 28(5): 1479-1489. doi:10.2134/jeq1999.00472425002800050013x.
  20. Son, E. H., and D. B. Park, 2012. Forecasting of Yeongdeok tourist by seasonal ARIMA model. *Journal of Agricultural Extension & Community Development* 19(2): 301-320 (in Korean). doi:10.12653/jecd.2012.19.2.301.
  21. White, M. J., and J. G. Arnold, 2009. Development of a simplistic vegetative filter strip model for sediment and nutrient retention at the field scale. *Hydrological Processes* 23(11): 1602-1616. doi:10.1002/hyp.7291.
  22. Woodward, D. E., R. H. Hawkins, R. Jiang, Jr, A. T. Hjelmfelt, J. A. Van Mullem, and Q. D. Quan, 2003. Runoff curve number method: examination of the initial abstraction ratio. In *World Water & Environmental Resources Congress 2003*: 1-10. doi:10.1061/40685(2003)308.
  23. Zhang, B., C. He, M. Burnham, and L. Zhang, 2016. Evaluating the coupling effects of climate aridity and vegetation restoration on soil erosion over the Loess Plateau in China. *Science of the Total Environment* 539: 436-449. doi:10.1016/j.scitotenv.2015.08.132.
  24. Zou, H., and T. Hastie, 2005. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B Statistical Methodology* 67(2): 301-320. doi:10.1111/j.1467-9868.2005.00503.x.