



A point-scale gap filling of the flux-tower data using the artificial neural network

Jeon, Hyunho^a · Baik, Jongjin^b · Lee, Seulchan^c · Choi, Minha^{d*}

^aGraduate Student, Civil, Architectural and Environmental System Engineering, Sungkyunkwan University, Suwon, Korea

^bPh.D, Center for Built Environment, Sungkyunkwan University, Suwon, Korea

^cGraduate Student, Department of Water Resources, Sungkyunkwan University, Suwon, Korea

^dProfessor, Department of Water Resources, Sungkyunkwan University, Suwon, Korea

Paper number: 20-072

Received: 29 July 2020; Revised: 28 August 2020; Accepted: 28 August 2020

Abstract

In this study, we estimated missing evapotranspiration (ET) data at a eddy-covariance flux tower in the Cheongmicheon farmland site using the Artificial Neural Network (ANN). The ANN showed excellent performance in numerical analysis and is expanding in various fields. To evaluate the performance the ANN-based gap-filling, ET was calculated using the existing gap-filling methods of Mean Diagnostic Variation (MDV) and Food and Agriculture Organization Penman-Monteith (FAO-PM). Then ET was evaluated by time series method and statistical analysis (coefficient of determination, index of agreement (IOA), root mean squared error (RMSE) and mean absolute error (MAE). For the validation of each gap-filling model, we used 30 minutes of data in 2015. Of the 121 missing values, the ANN method showed the best performance by supplementing 70, 53 and 84 missing values, respectively, in the order of MDV, FAO-PM, and ANN methods. Analysis of the coefficient of determination (MDV, FAO-PM, and ANN methods followed by 0.673, 0.784, and 0.841, respectively.) and the IOA (The MDV, FAO-PM, and ANN methods followed by 0.899, 0.890, and 0.951 respectively.) indicated that, all three methods were highly correlated and considered to be fully utilized, and among them, ANN models showed the highest performance and suitability. Based on this study, it could be used more appropriately in the study of gap-filling method of flux tower data using machine learning method.

Keywords: Flux tower, Artificial neural network, Evapotranspiration, Mean diurnal variation, Food and agriculture Organization penman-monteith

인공신경망 기법을 이용한 청미천 유역 Flux tower 결측치 보정

전현호^a · 백종진^b · 이슬찬^c · 최민하^{d*}

^a성균관대학교 건설환경시스템공학과 석박통합과정, ^b성균관대학교 건설환경연구소 박사후연구원, ^c성균관대학교 수자원학과 석박통합과정,

^d성균관대학교 수자원학과 교수

요 지

본 연구에서는 청미천 유역에서의 플렉스타워에서 산출되는 증발산량의 결측값을 보완하기 위해 인공신경망(Artificial Neural Network, ANN)을 사용하였다. 비교 평가를 위해, Mean Diurnal Variation(MDV), Food and Agriculture Organization Penman-Monteith(FAO-PM) 방법들을 이용하여 증발산량을 산정하였고, ANN 방법을 이용한 결과와 비교하였다. 비교 평가 방법으로 시계열 방법 및 통계 분석(결정계수, IOA, RMSE, MAE)이 사용되었다. 각 gap-filling 모델의 검증에 위해 2015년의 30분 단위 데이터를 이용하였으며, 121개의 결측값 중 MDV, FAO-PM, ANN 방법 순으로 각각 70, 53, 54개의 결측값을 보완하여 모든 데이터가 관측되지 않은 36개의 데이터를 제외하면 각각 82.4%, 62.4%, 63.5%의 성능을 보였다. 결정계수(MDV, FAO-PM, ANN 방법 순으로 각각 0.673, 0.784, 0.841)와 IOA(MDV, FAO-PM, ANN 방법 순으로 각각 0.899, 0.890, 0.951)를 분석한 결과, 3가지 방법 모두 양질의 상관성을 보여 활용성이 충분하다고 판단되며, 이 중 ANN 모델이 가장 높은 적합도와 양질의 성능을 나타내었다. 본 연구를 기반으로 기계학습방법을 이용한 플렉스 타워 자료의 gap-filling 연구에 보다 적절하게 활용될 수 있을 것이다.

핵심용어: 플렉스 타워, 인공신경망, 증발산, 평균일변동법, FAO-PM

*Corresponding Author. Tel: +82-31-290-7527

E-mail: mhchoi@skku.edu (M. Choi)

1. 서론

지표 및 수면의 물이 공기 중으로 기화하는 증발(evaporation)과 식생의 기공을 통해 방출되는 증산(transpiration)을 아우르는 증발산(evapotranspiration, ET)은 에너지 및 물 순환을 파악하기 위한 핵심적인 인자이다(Allen et al., 2007). 지구온난화와 기상이변으로 인한 물 부족 해결을 위해, 증발산의 정확한 파악은 효율적인 수자원관리를 위해 필수적으로 고려되어야 하는 것으로 인지되고 있다.

증발산에 대한 관심이 높아지면서, 농업, 축산업 등의 실생활과 기타 연구에 이용하기 위해 증발산량을 측정하는 방법 중 가장 통상적으로 사용된 방법은 증발접시법이다(Jeong and Kang, 2009). 이는 금속재질의 여러 가지 규격의 증발 접시에 일정량의 물을 채운 후 그 변화량을 측정하는 방법으로 다른 방법들에 비해 측정하기에 간편하다. 그러나 증발접시법은 토양이 물에 포화됨을 가정하며, 식생의 증산을 반영하지 못하기에 실제 증발산량이라고 정의하기 어렵다는 이론적인 한계점을 가지고 있다(Peterson et al., 1995; Roderick and Farquhar, 2004). 과학이 발전함에 따라 지표면과 식생에서의 증발과 증산을 고려한 실제 증발산량을 관측하기 위해서, 생태계와 대기 사이 임의의 공간에서 난류로 인한 기체의 온도, 밀도 및 연직 풍속을 관측하여 에너지 및 다양한 물질을 정량화 하는 미기상학적 기술인 에디공분산(eddy covariance)을 기반으로 한 플럭스 타워(flux tower)가 사용되고 있다(Chen et al., 2009; Post et al., 2015). 플럭스 타워에는 대기온도, 습도 등과 같은 일반적인 수문기상인자들을 관측하는 장비와 에디공분산 기법을 이용하여 연속적으로 에너지 및 탄소에 관한 인자들을 관측하는 장비들이 탑재되어 있다(Kang et al., 2018). 세계 각지의 플럭스타워들은 다양한 생태계 및 토지피복에서 지속적으로 관측 및 관리되고 있으며, FLUXNET, AMERIFLUX, AsiaFlux 등의 네트워크를 통해 공유되고 있고, 우리나라의 경우 KoFlux를 출범하여 국가농림기상센터, 국립기상과학원, 한국수자원조사기술원 등에서 관측 및 운영하고 있다.

미기상학적 방법을 기반으로 다양한 가정 사항들을 고려하는 에디공분산 기법을 통해 산출된 관측값들은 폭설과 돌발성 강우 등의 극한기상상황에서의 환경적인 요인과 기기 오작동과 같은 기계적 결함 등에 의해 오차 값 및 결측값(플럭스 타워 자료의 수득률 65 ~ 70%)을 산출하게 된다(Kim et al., 2016). 위와 같은 오차 값들은 지형 및 기후적인 특성을 고려하여 각 기관의 플럭스 타워 품질 관리(Quality Control, QC) 프로세스를 토대로 제공되고 있다(Baldocchi et al., 2001; Papale et al., 2006). 국내에서는 KoFlux 표준화 프로그램을 통해 좌표 회

전, 밀도 효과 보정 및 튀는 자료 제거 과정을 거쳐 품질관리를 수행하고 있으며, 이 과정을 거치면서 불량자료(오차값)로 판별되는 flag 값들은 결측값이 된다. 그러나 실제적으로 수문순환 및 연구에 있어서 연속적인 자료가 요구되기 때문에, 이러한 결측치를 보완하기 위해 Penman-Monteith 복합방정식과 같은 물리식을 이용하여 증발산량을 보정하는 방법과, 수정된 색인 목록 방법과 같은 통계적 분석이 이용되고 있다(Kim et al., 2016; Park et al., 2015). 현재 KoFlux에서는 증발산 자료 결측 구간을 보완하기 위해서 수정된 색인 목록 방법(modified lookup table method, MLTM)과 Penman-Monteith (PM) 복합방정식에 갈만 필터(Kalman filter)를 적용하는 두가지 방법이 사용되고 있다(Hong et al., 2009).

기존 선행연구들을 살펴보면 플럭스 타워 기반 증발산 자료의 결측 구간 보완 방법(gap-filling)으로 일반적인 통계 기반 방법이 다수 시행되고 있으며, 최근 수치 예측 문제 해결을 위해 과학기술 전반에 도입되고 있는 머신러닝(machine learning) 방법의 우수함이 다수 검증되었음에도 불구하고(Kendall et al., 2018; Jean et al., 2016; Gao et al., 2018) 이를 이용한 gap-filling 방법은 아직 찾아보기 힘들다. 머신러닝은 컴퓨터를 활용하여 데이터를 자동적으로 학습하게 하는 연구로써 인공지능 분야에서 핵심적인 연구이다(Langley and Simon, 1995; Mitchell, 1997). 이 방법은 다양한 분야로 연구가 확대되고 있으며, 검색엔진, 게임, 음성인식 등의 일상생활과 밀접한 부분까지도 상용화가 되고 있다(Heo et al., 2018). 이에 본 연구에서는 머신러닝의 대표적인 방법 중 하나인 인공신경망 기법을 gap-filling 분야로 확대하였고, 기존 연구에서 활용되었던 Food and Agriculture Organization Penman-Monteith (FAO-PM) 방법 및 평균 일변동(Mean Diurnal Variation, MDV) 방법과 비교 검증 하였다. 검증방법으로는 시계열 분석 및 다양한 통계분석이 사용되었고, 성능 및 적합성이 가장 뛰어난 방법을 선정하였다.

2. 연구 지역

본 연구에서는 증발산의 측정이 가능한 플럭스 타워 자료를 활용하기 위해, 청미천(Cheongmicheon Farmland Site)을 연구지역으로 선정하였다(Fig. 1). 청미천은 남한강 하류(37° 9' 35"N, 127° 39' 10"E)에 위치하며, 그 주변은 41%의 임야지와 49%의 농경지로 이루어져 있다. 청미천은 경사가 거의 없는 평지이며 식생의 높이는 약 1 m 정도이다. 연평균 강수량은 1,220 mm이며 연평균 대기온도는 11.6°C, 해발고도는 141 m이다.

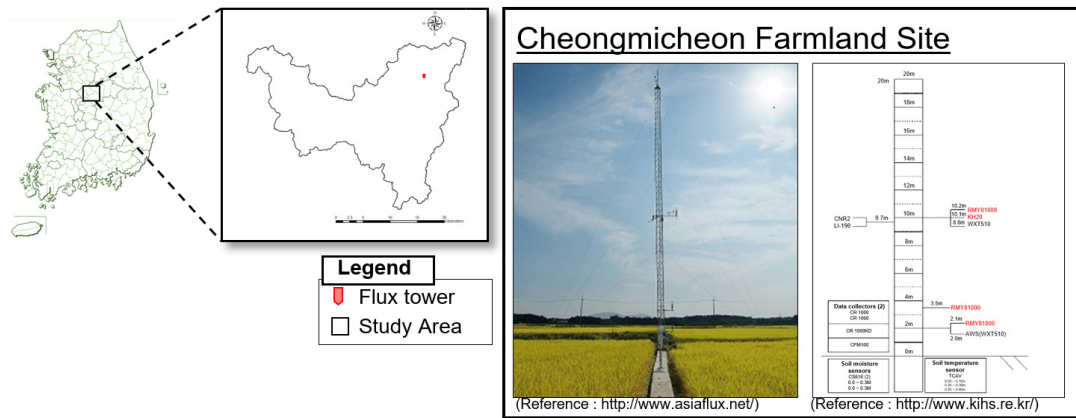


Fig. 1. Geographic information of Cheongmicheon Farmland Site

청미천 지역의 플렉스타워에는 미기상학적 방법을 기반으로 한 에다공분산 기법과 기타 기상관측 시스템, 순복사에너지 관측 시스템 등이 설치되어 있다. 사용되는 센서로는 순장파복사에너지, 순단파복사에너지, 순복사에너지를 관측하는 센서인 CNR2 Net Radiometer, 에다공분산 기법에 기반한 삼차원 초음파 풍향풍속계(CSAT3)와 고속반응 기체 분석기(EC-155) 및 기타 기상 및 플렉스를 관측하는 여러 센서들이 설치되어 있다. 플렉스타워의 자세한 설명은 다음과 같은 논문들에서 확인 할 수 있다(Park *et al.*, 2015; Yuan *et al.*, 2007; Kwon *et al.*, 2007).

3. 연구방법

3.1 Gap-filling Methods

플렉스타워의 증발산량을 gap-filling 하는 방법으로는 물리식을 기반으로 하여 증발산량을 추정하는 방법과, 플렉스타워에서 산출되었던 데이터들에 통계적인 기법을 활용하는 방법으로 크게 두 가지가 활용되고 있다. 본 연구에서는 각각의 대표적인 방법인 FAO-PM 방법, MDV방법과 인공지능을 활용한 ANN 기반 gap-filling 방법을 검증하였다.

3.1.1 FAO-PM

증발산량을 계산하기 위해 만들어진 다양한 경험적 식들은 새로운 환경적인 조건의 지역에서는 적절하지 않고, 이를 엄격하게 교정하기 위해서는 많은 시간과 비용이 든다. 이에 미국의 FAO에서는 증발산을 산정하기 위한 대상 표면이 반사율(albedo)이 0.23이며 수분공급이 원활하게 이루어지는 기준작물(알팔파 또는 잔디)이 있음을 가정하는 등 다양한 환경조

건을 아우르는 하나의 기준을 정하여 해당 조건에서의 기준증발산량을 산정할 수 있는 FAO-PM 식을 개발하였다. 이러한 기준증발산량은 아래 Eq. (1)의 식을 통하여 실제증발산으로 변환이 가능하다.

$$ET_a = K_c \times ET_0 \quad (1)$$

여기서, ET_a 는 실제 증발산량(mm/day), K_c 는 작물계수를 나타낸다. ET_0 는 기준 증발산량(mm/day)을 나타내며, Eq. (2)을 통해서 산정할 수 있다.

$$ET_0 = \frac{0.408 \Delta (R_N - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 u_2)} \quad (2)$$

여기서, ET_0 는 기준 증발산량(mm/day), Δ 는 증기압 곡선의 기울기(kPa/°C), R_N 은 순복사에너지(MJ/mday), G 는 지중열(MJ/m²day), γ 는 건습계 상수(kPa/°C), T 는 2 m 높이에서의 일평균대기온도(°C), u_2 는 2 m 높이에서의 풍속(m/s), e_s 는 포화증기압(kPa), e_a 는 실제증기압(kPa)이다. FAO-PM방정식을 계산하기 위해 필요한 입력자료의 종류는 풍속, 순복사량, 대기온도, 상대습도이다. 본 연구에서는 작물계수를 결측값이 발생한 시간의 인근 일주일 동안, 플렉스타워에서 관측된 실제 증발산량과 FAO-PM식을 이용해 추정된 기준 증발산량의 비를 이용해 계산하였고, 일반적인 작물계수의 값이 0.05 ~ 2의 범위에서 거동하기에 계산된 작물계수를 해당 범위에 포함되도록 보정하는 과정을 거쳤다(Kim *et al.*, 2016). 보정된 작물계수 값을 결측구간에서의 기준 증발산량에 곱하여 실제 증발산량을 산정하였다.

3.1.2 MDV

MDV 방법은 대표적인 통계적인 gap-filling 방법 중 하나로, 결측된 증발산량은 인근 기간의 동일한 시간대의 증발산량과 비슷한 경향성을 띠 것이라는 가정 하에 시행되는 선형 보간법이다. MDV 방법으로 높은 신뢰도의 증발산을 추정하기 위해서는, 선형보간법을 시행하기 위하여 연구자가 설정하는 결측구간에 인접한 날의 기간인 창 크기(window size)의 적절한 설정이 중요하다. 본 연구에서는 선행연구인 Baldocchi *et al.* (2001)에서와 같이 창 크기를 6으로 설정하여 결측구간 기준 3일 전후 총 6일의 증발산량에 선형보간법을 시행하였으며, 창 크기 내에 2개의 또 다른 결측값이 포함될 경우 해당 결측값은 보완이 불가능한 것으로 판단하였다. MDV 방법은 물리식을 이용한 FAO-PM 방법과 달리 기상인자에 대한 고려없이 인근 증발산량만을 이용하여 결측된 증발산량을 산정하기에 극한기상상황과 같은 환경적 요인에 인한 결측값들을 예측하기에는 문제가 있지만, FAO-PM과 같은 물리식 기반의 방법에 비해 계산 과정이 간편하다는 것에 가치가 있다. 본 방법에 대한 자세한 설명은 Kim *et al.* (2016)에서 확인할 수 있다.

3.2 Artificial Neural Network (ANN)

ANN은 인간이나 동물이 가지고 있는 뇌의 신경세포인 뉴런에서 착안하여 이들의 연결관계를 수학적으로 모델링 하여 인공적으로 학습하는 지능을 만드는 것을 목적으로 개발되었다. ANN의 장점은 데이터들을 이용하여 스스로 학습이 가능하며, 특히 회귀분석 모델에 비해 고려되는 가중치의 개수가 많아 비선형적인 모델을 예측, 분석하는데 효과적이다(Jensen *et al.*, 1999; Atkinson and Tatnall, 1997).

ANN은 다양한 구조의 모델이 있지만, 그중 예측모델로 자주 쓰이는 Multi-layer Perceptrons (MLP)의 구조는 Fig. 2와 같다. MLP는 입력층(input layer), 은닉층(hidden layer), 출

력층(output layer)의 3개 종류의 층으로 구성되었고, 각 층들은 여러 개의 노드(node)들로 이루어져 있다. 데이터들은 입력층의 노드에서 은닉층, 출력층 순으로 전달되며 각 전달 과정에서 가중치를 이용해 출력값이 변환된다. 출력층에서 나오는 출력값과 데이터의 출력값, 즉 원하는 출력값과의 차이를 이용해 가중치 값을 역으로 변화시켜 모델이 학습된다. MLP는 은닉층의 수와 node의 수를 증가시키면서 더욱 복잡한 비선형의 모델을 학습할 수 있다. ANN 모델의 학습에 있어 최적의 은닉층 수 및 노드 수를 선정하는 방법에 대해서는 많은 선행연구들이 진행되고 있으며, 은닉층 수와 노드 수를 부적절하게 선정할 시 과적합 또는 과소적합이 발생할 수 있다(Sheela and Deepa, 2013). Panchal *et al.* (2011)은 은닉층 수를 결정하는데 있어서 1개의 은닉층으로도 대부분의 모델들이 충분히 정상적으로 구동함을 밝혔다. 노드 수를 선정하는 방법에 대해서는 많은 연구들이 진행되었으나, 각 연구들마다 선정하는 방법 및 그 수가 상이하다(Sartori and Antsaklis, 1991; Tamura and Tateishi, 1997; Choi *et al.*, 2008; Hunter *et al.*, 2012).

본 연구에서는 2012년부터 2014년까지의 플렉스 타워 데이터를 사용하여 ANN 모델을 학습하였고, 2015년의 데이터를 활용하여 모델의 정확도를 검증하였다. 입력 자료로 사용되는 플렉스 타워 데이터 값의 범위가 각각 다르기에 정규화 과정을 거쳐 숫자의 차이로 인한 모델의 학습오류를 방지하였다. 모델 구축에 관해서는 1개의 은닉층을 설정하여 ANN을 시행하였고, 활성화 함수로는 linear, log-sigmoid, tangent-sigmoid로 3개를 고려하였다. 또한 노드수를 선정함에 있어서 앞에서 언급한 것과 같이 선행연구들마다 상이한 결과를 보이기에, 본 연구에서는 1개부터 600개의 범위에서 노드수를 변경하며 학습시켜, 최적 노드수를 찾을 수 있게 하였고, 또한 은닉층과 출력층에서의 2개의 구간에 대해 최적의 활성화함수를 고려 할 수 있게 하였다(Fig. 2).

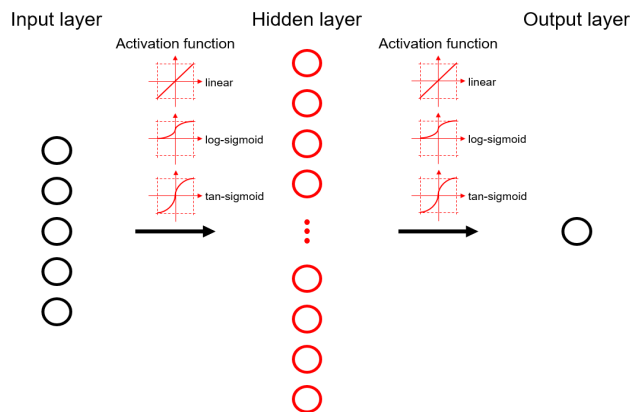


Fig. 2. Conceptual diagram of Artificial neural network

4. 결과 및 고찰

4.1 ANN 모델의 입력자료 선정을 위한 증발산량과 수문기상학적인자 간의 관계성 분석

Gap-filling 방법을 통해 산출된 증발산량의 다각적인 해석을 위하여, 청미천 유역에서의 플렉스타워를 이용해 관측한 증발산량과 다른 입력 자료들의 상관성 분석을 시행하였다 (Fig. 3). 분석결과를 확인하면, 모든 수문기상학적 인자들 간의 유의성($p < 0.01$)이 높게 나타나는 것으로 관찰되었다. 특히, 증발산량과 순 복사에너지간의 상관계수 값이 0.792로 가장 영향이 높은 것으로 나타났다. 이는 실제적으로 증발산량과 순 복사에너지가 아주 밀접한 관계를 가지고 있다는 것을 나타내고 있다.

Fig. 3에서 증발산을 기준으로 영향을 미치는 인자들과의 관계성을 파악하면 다음과 같이 순 복사에너지(Net radiation, R_N), 상대습도(Relative humidity, RH), 대기온도(Air temperature, T_a), 기압(Atmospheric pressure, P_a), 풍속(Wind speed, u), 강수량(Precipitation, P) 순으로 관계성이 크게 나타내는 것을 확인할 수 있다. 증발산은 순복사량과의 상관계수 값이 0.792로 제일 높은 것으로 관찰되었다. Amba and Baltas (2012)은 다양한 증발산을 산정하는 식들을 이용하여 증발산을 산정하였고, 증발을 산정하는데 영향을 미치는 민감도 분석의 결과 보통 R_N 이 가장 높은 민감도를 나타내는 것으로 확인되었으며, R_N , T , RH, u 순으로 민감도가 높게 나타났다. 이는 본 연구에서와 비교하였을 때 비슷한 경향을 나타내는 것으로 보이고 확연하게 증발산에 가장 영향을 미치는 인자가 R_N 임을 단편적으로 나타내고 있다. 그에 비하여, 풍속의 경우 증발산과의 상관계수 값은 0.224으로 강수량에 비해 비교적 높은 값이 나타났으나, 나머지 순 복사에너지, 대기온도, 기압, 상대습도에 비해 낮은 상관계수 값을 확인할 수 있었다.

2015년의 30분 단위 청미천 플렉스타워 자료의 결측값을 확인하였다. 총 121개의 증발산이 결측되는 시점 중 67개의 시점에서 풍속이 결측되었고 66개의 시점에서 대기온도가 결측되었으며, 나머지 인자들이 결측되는 시점은 37개 뿐이었다. 플렉스타워에서 증발산량이 결측되는 시점에 풍속과 대기온도가 동시에 결측되는 경우가 많았으며, 더 많은 결측값을 산출한 풍속을 인자를 제외한 순복사량, 대기온도, 상대습도, 강수량, 기압으로 증발산과 상관성이 있는 5개의 인자가 사용되었다. 이는 입력자료의 유무가 ANN 방법의 gap-filling 성능과 직결되는 중요한 요인이기에, 풍속을 제외하였으며 상대적으로 증발산량에 더 큰 상관성을 나타내는 대기온도는 포함하는 것이 적절하다고 판단하였다.

u	1						
T _a	0.109	1					
R _N	0.184	0.297	1				
RH	-0.414	-0.077	-0.476	1			
P	0.172	0.272	-0.101	0.141	1		
P _a	-0.234	-0.756	-0.116	-0.070	-0.166	1	
LE	0.224	0.456	0.792	-0.507	-0.126	-0.275	1
	u	T _a	R _N	RH	P	P _a	LE

Fig. 3. Multiple correlation coefficient between hydro-meteorological factors

본 연구에서는 최적의 ANN 모델을 선정하기 위해, 600개의 노드 수와 3개의 활성화 함수를 조건으로 5400개의 조합 중 가장 낮은 RMSE를 가지며, 높은 결정계수와 IOA를 가지는 모델을 선정하게 하였다. 먼저 변경 조건 중 노드 수를 기준으로 하여 각각 1개의 모델을 선정하였고, 이때 결정계수가 0.80 이하인 모델은 제외하였다. 선정된 598개 모델의 통계분석 결과 결정계수, IOA, RMSE, MAE의 평균값은 각각 0.825, 0.947, 1.287 mm/day, 0.6651 mm/day 이었으며, 이 중 결정계수와 IOA가 가장 높으며 RMSE와 MAE 값이 가장 작은 값을 출력하여 최적의 모델로 선정된 모델은 노드 수가 22개, 은닉층과 출력층에서의 활성화함수가 모두 tan-sigmoid인 모델이었다. 노드 수의 변화에 따른 모델의 적합도 차이가 크게 나타나지는 않았으나, 노드 수가 증가할수록 모델의 적합도가 낮아지는 경향을 보였으며, 최적의 모델로 선정된 모델을 이용하여 본 논문의 gap-filling 연구를 진행하였다.

4.2 Gap-filling 모델 성능 평가

2015년 청미천 플렉스타워의 30분 단위 데이터를 기반으로 하여 MDV, FAO-PM, ANN 방법을 이용하여 gap-filling을 시행하였다. 플렉스타워에서 어떠한 데이터도 생성되지 않은 시점 36개의 데이터를 제외한 17,484개의 데이터 중 결측값이 발생한 데이터는 85개이고, 각각 4월(71개), 10월(4개), 11월(4개), 12월(6개)에 발생하였다. 이 중 성공적으로 결측값을 보완할 수 있었던 데이터의 개수는 MDV방법과 FAO-PM은 각각 70개와, 53개의 결측값을 보완하였으며, 이에 비하여 ANN 방법은 54개의 결측값을 보완하였다. FAO-PM방법과 ANN 방법의 경우 입력자료의 부재가 모델의 성능에 큰 영향을 끼치며, 결측값이 발생했을 때의 모델 거동을 Fig. 4에 나타내었다.

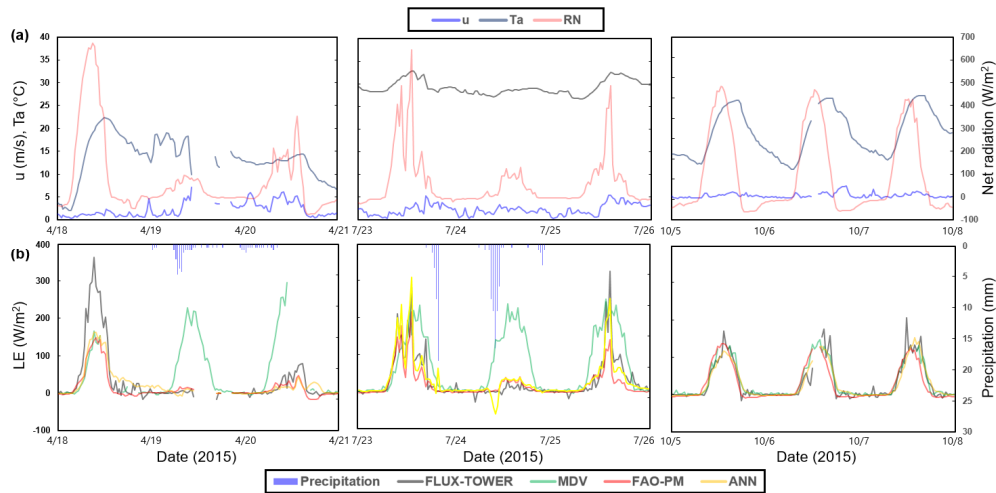


Fig. 4. Comparisons between (a) hydrometeorological factors and (b) latent heat fluxes when missing values occurred

Fig. 4는 결측값이 발생한 기간 중 2일(4/19, 10/6)과 결측값이 발생하지 않은 기간 중 1일(7/24)의 인근 30분 단위 데이터에 대해 각 수문기상인자 및 증발산량의 시계열을 나타내었다. Fig. 4(a)에서 나타낸 3개의 수문기상학적인자들은 각각 FAO-PM 방법과 ANN 방법에 큰 영향을 끼치는 인자들로, 각 인자들의 부재 및 오차가 결측값 보완 결과에 큰 영향을 미친다. Fig. 4(b)에서 4가지의 예측 및 관측된 증발산량의 거동을 확인 할 수 있다. Fig. 4(a)에서 수문기상학적인자들 중 결측값이 발생한 경우 Fig. 4(b)와 같이 결측값 보완을 하지 못하는 모습을 보인다. 또한 10월 6, 7일과 같이 환경적인 요인의 변동이 비교적 크지 않은 시점에는 3가지 gap-filling 방법 모두 관측값과 비슷한 거동을 보이지만, 4월 19일과 같이 환경적인 요인의 변동이 비교적 큰 시점에는 다른 두 가지 방법에 비해 MDV 방법이 그 경향성을 따라가지 못한다. 특히 7월 24일과 같이 비와 구름 등의 요인으로 R_N 값이 급격하게 줄어드는 시점에서 MDV 방법이 가장 성능이 떨어졌다. 7월 24일 근처 강수량이 집중되는 두 시점에 ANN 모델에서 산출된 증발산량이 전자는 소폭 상승하는 경향을 보이고, 후자는 하락하는 경향을 보인다. 이는 R_N 과 T_a 가 비슷한 값이 산출 되었을 때 풍속이 높은 값이 산출되는 경우 높은 증발산량이 산출되는 것으로 판단된다. 실제 두 시점의 R_N 과 T_a 의 차이가 각각 10 W/m^2 과 0.9°C 가량 나타나며 이는 정규화 과정($0 \sim 1$)을 거칠 시 0.1에 불과하나, 풍속의 경우 2 m/s 가량 차이이며 0.4에 해당하는 차이를 보여 그 영향력이 더욱 크게 나타난 것으로 보인다.

4.3 Gap-filling 모델 적합성 평가

각 결측값 보완 방법의 적합도 및 오차를 파악하기 위해 통계분석을 시행하였다(Fig. 5). 통계분석은 결정계수, IOA,

RMSE, MAE를 이용하였으며, 월별 분석과 1년 치 전체 데이터의 분석을 시행하였다. 각 gap-filling 방법 별 예측결과를 분석한 결과, IOA의 평균값은 MDV, FAO-PM, ANN 방법 순으로, 0.899, 0.890, 0.951, 결정계수의 평균값은 0.673, 0.784, 0.841으로, ANN 방법이 다른 두 가지 방법에 비해 높은 적합도를 보이는 것으로 확인되었다. 또한 기상학적 인자들을 반영하여 예측하는 ANN 방법과 FAO-PM 방법에 비해 인근 일자의 증발산량을 선형보간법을 이용하여 예측하는 MDV 방법은 적합도가 낮은 경향을 보인다(Fig. 5). 오차와 관련된 RMSE와 MAE는 ANN, FAO-PM, MDV 방법 순으로 좋은 결과를 보인다. 결측값이 발생하는 기간, 특히 4월의 인근에서 RMSE와 MAE값이 크게 증가하며 MDV 방법과 다른 두 방법들의 RMSE와 MAE가 역전되는 결과를 보인다(Fig. 5). 이는 계절이 바뀌면서 기상상황에서 많은 변수가 생기면서, 전체적인 플렉스타워 데이터의 질이 하락한 것으로 판단된다.

Fig. 6에서 각 3가지 방식의 일평균 증발산량을 산정한 결과를 나타내었다. MDV 방법은 window size 내의 값들을 선형보간 하는 방법으로 예측하는 만큼 그 결과 값들이 튀지 않는 반면, 각 기상인자들을 반영하는 FAO-PM과 ANN 방법은 그 결과 값들이 실제 관측 값을 따라 변동이 크게 나타났다. 이는 MDV 방법은 주변 환경이 급변할 때의 증발산량 값을 잘 예측하지 못한다는 것으로 볼 수 있다. Fig. 7은 월별 증발산량 및 누적 증발산량을 산정한 결과이다. MDV 방법의 경우 월 증발산량을 산정하는데 있어서 가장 정밀한 결과를 나타내었다. 이는 결측값을 보완하기 위해 창 크기 내의 실제 증발산량에 선형보간법을 이용하기에, 환경적인 변화가 비교적 큰 구간에서는 개별적인 데이터의 산정에 어려움을 겪지만, 전체적인 증발산량의 평균값은 경향을 잘 따르는 것으로 판단된

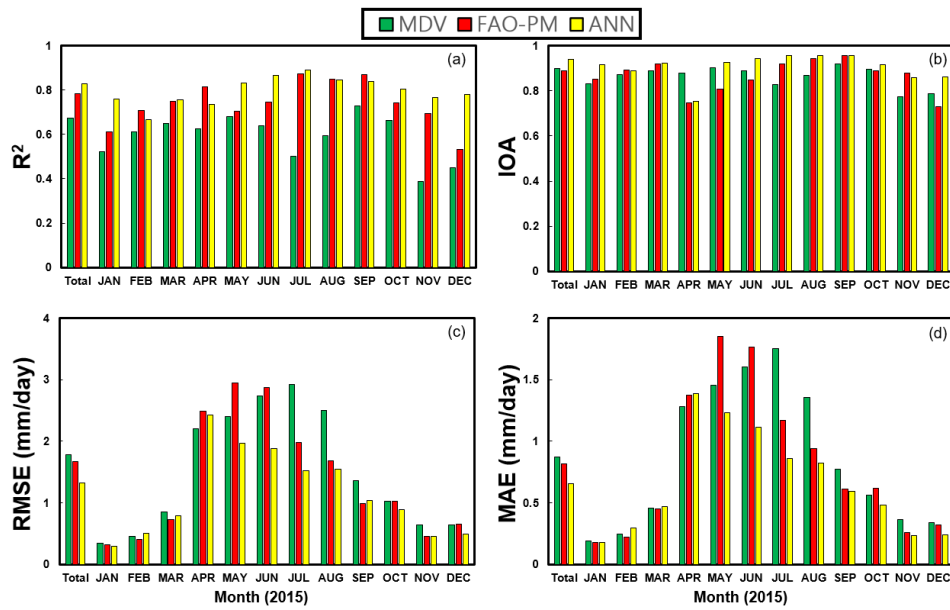


Fig. 5. The coefficient of determination, index of agreement (IOA), root mean squared error (RMSE) and mean absolute error (MAE) between estimated ET of each gap-filling method and observed ET

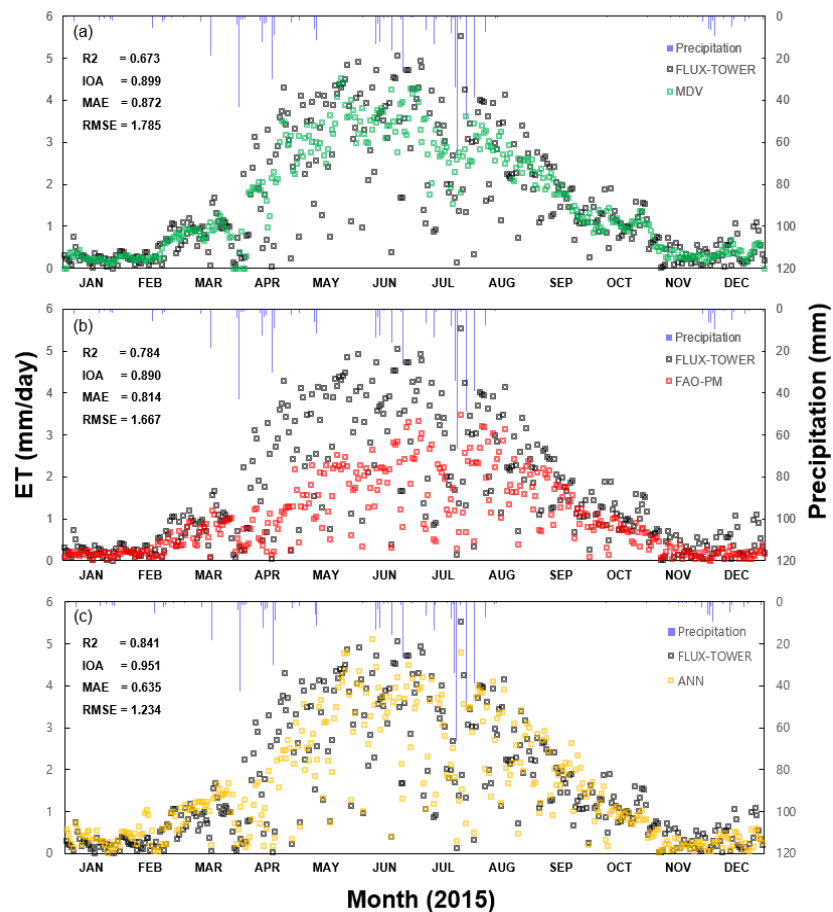


Fig. 6. Comparisons between observed daily ET and estimated daily ET of gap-filling methods. (a) : MDV, (b) : FAO-PM, (c) : ANN

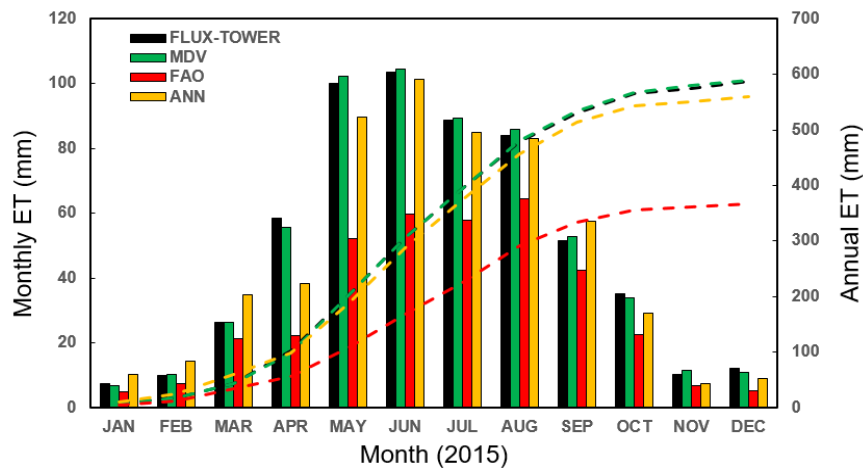


Fig. 7. Comparisons between observed monthly and annual accumulated ET and estimated monthly and annual accumulated ET of gap-filling methods

다. FAO-PM 방법의 경우에는 식물의 성장기에 과소 산정되는 경향을 보인다. 이는 Kc값을 산정할 때, 평균적인 농경지에서 Kc의 범위를 벗어난 값들을 보정해주는 과정에서 높게 나온 값들을 낮추면서 생긴 현상으로 보인다. 대략적인 토지의 종류 및 식생의 종류에 따른 Kc값을 FAO에서 제공하고 있으나, 이 값이 모든 지역에서 적절하다고 단정하기 어렵다. 이에 Kc를 산정하기 위해 많은 선행연구들이 시행되었으나, 대부분의 연구들이 한 특정지역에서의 Kc값을 산정하는데 그쳐 있어 범용적으로 사용하기에 무리가 있다(Marin *et al.*, 2016; Park *et al.*, 2017). 이에, gap-filling이 시행될 지역의 Kc 값을 산정하기 위한 후속연구가 시행되어, 엄밀한 Kc값이 산정되었을 시에 FAO-PM 방법이 더욱 좋은 적합도를 보일 것으로 판단된다. ANN 방법의 경우 여름철에는 과소산정 되는 경향을 보이며, 겨울철에는 과대산정 되는 경향을 보인다. 결측값이 발생하는 4월, 10월, 11월, 12월에는 예외적으로 예측값들이 관측값에 비해 과소 산정되는 경향을 보이며, 특히 결측값의 수가 가장 많았던 4월의 경우 그 차이가 크게 나타났다. 연 중 발산 누적 그래프를 보면, MDV 방법(591 mm), 플렉스타워의 관측값(587 mm), ANN 방법(559 mm), FAO-PM 방법(367 mm) 순으로 크게 나타났으며, 크게 과소산정된 FAO-PM 방법을 제외한 나머지 두 가지 방법의 결과 들을 살펴보면, 선행연구에서의 결과와 같이 연 누적 강수량(975 mm)의 약 60% 정도가 대기 중으로 증발했다는 것을 확인할 수 있다(Kim *et al.*, 2016). 이와 같은 결과들로 ANN 방법이 기존의 다양한 gap-filling 방법들보다 뛰어난 거동을 보인다는 것을 확인할 수 있으며, 이는 효율적인 수자원 관리에 큰 기여를 할 수 있을 것으로 보인다.

5. 결 론

본 연구에서는 청미천 유역의 플렉스타워에서 관측된 증발산량 데이터에서 인공신경망 기법을 활용한 gap-filling 방법의 적합성에 대해 연구하였다. 이에 비교 대상으로 기존에 gap-filling 방법으로 사용되고 있던 MDV 방법과 FAO-PM 방법을 사용하였고 그 결론은 다음과 같다.

- 1) 2015년 청미천유역 플렉스타워의 결측값을 보완한 결과, 121개의 결측값 중 MDV, FAO-PM, ANN 방법 각각 70(57.9%), 53(43.8%), 54(44.6%)개의 결측값을 보완할 수 있었으며, 모든 데이터가 관측되지 않은 36개의 데이터를 제외하면 ANN 방법이 63.5%의 결측값을 보완할 수 있었다.
- 2) 본 연구에서 활용한 세 가지 gap-filling 방법인 MDV, FAO-PM, ANN 방법 각각 평균 결정계수는 0.673, 0.784, 0.841로 나타났으며, IOA는 0.899, 0.890, 0.951로 나타났다. ANN 방법의 적합성이 가장 우수한 것으로 판단되며, FAO-PM 방법은 MDV 방법에 비해 각 시점에서의 증발산량을 잘 모의하지만 전체적으로 과소산정이 되어 IOA 값이 더 낮은 것으로 판단된다.
- 3) MDV, FAO-PM, ANN 방법 각각 평균 RMSE 값은 1.785 mm/day, 1.667 mm/day, 1.234 mm/day로 나타났으며, MAE 값은 0.872 mm/day, 0.714 mm/day, 0.635 mm/day로 나타났다. 이는 MDV 방법이 기상변화에 따른 증발산량의 변화를 반영하지 못하여 오차값들이 커진 것으로 판단된다.

- 4) Gap-filling 방법 중 과소산정이 심한 FAO-PM 방법을 제외한 ANN, MDV 두 가지 방법 모두 2015년의 연 누적 증발산량이 559 mm와 590 mm로 플릭스 타워의 관측량과 비슷하며, 선행연구와 같이 연 강수량의 약 60%를 차지한다.

전반적으로 MDV 방법은 앞서 언급한 것과 같이 주변 날짜에 대한 선형보간법을 적용함으로써 외부조건을 고려하지 않고 결측값을 보완할 수 있으나, 그 반면에 다양한 기상자료를 활용함으로써 결측값을 보완하는 FAO-PM 방법은 전반적인 경향성이 기존의 MDV 방법에 비해서는 높으나, 요구되는 기상인자가 관측되지 않는 시점에서 결측값을 보완하지 못하는 결과를 나타내었다. 그에 비하여 본 연구에서 활용한 ANN 방법은 증발산량이 관측되지 않을 때 동시에 관측되지 않는 기상인자들을 최대한 제외하고 머신러닝을 시행하여 그 성능이 상대적으로 높게 나타났고, 적합성 또한 다른 방법들에 비해 가장 높게 나타났다. 이러한 결측값을 보완하는 방법으로 머신러닝 방법들 중 하나인 ANN을 활용함으로써 효율적인 수자원관리를 위한 양질의 수문기상학적 자료의 생성 및 축적에 큰 도움이 될 것이다.

본 연구에서는 현재 청미천의 2012-2014년간의 3년 자료를 ANN 모형 구축을 위해 활용하였으나, 향후 장기간의 자료를 확보하고 보다 다양한 머신러닝 기법을 적용한다면 양질의 gap-filling 방법을 파악 및 개발할 수 있을 것으로 사료된다. 또한, 식물의 생장기와 비 생장기에 각각 과소 산정 및 과대 산정되는 경향을 보여 기간을 나누어 학습을 진행할 시 더 양질의 모델 적합도를 얻을 수 있을 것으로 판단된다.

감사의 글

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(NRF-2019R1A2B5B01070196)입니다. 또한 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(NRF-2018R1D1A1B07049029)에 의해 수행되었습니다.

References

- Allen, R.G., Tasumi, M., and Trezza, R. (2007). "Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC)-Model." *Journal of Irrigation and Drainage Engineering*, Vol. 133, No. 4, pp. 380-394.
- Ambas, V.T., and Baltas, E. (2012). "Sensitivity analysis of different evapotranspiration methods using a new sensitivity coefficient." *Global NEST Journal*, Vol. 14, No. 3, pp. 335-343.
- Atkinson, P.M., and Tatnall, A.R. (1997). "Introduction neural networks in remote sensing." *International Journal of remote sensing*, Vol. 18, No. 4, pp. 699-709.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H.P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S. (2001). "FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities." *Bulletin of the American Meteorological Society*, Vol. 82, No. 11, pp. 2415-2434.
- Chen, B., Black, T.A., Coops, N.C., Hilker, T., Trofymow, J.T., and Morgenstern, K. (2009). "Assessing tower flux footprint climatology and scaling between remotely sensed and eddy covariance measurements." *Boundary-Layer Meteorology*, Vol. 130, No. 2, pp. 137-167.
- Choi, B., Lee, J.H., and Kim, D.H. (2008). "Solving local minima problem with large number of hidden nodes on two-layered feed-forward artificial neural networks." *Neurocomputing*, Vol. 71, No. 16-18, pp. 3640-3643.
- Gao, H., Struble, T.J., Coley, C.W., Wang, Y., Green, W.H., and Jensen, K.F. (2018). "Using machine learning to predict suitable conditions for organic reactions." *ACS central science*, Vol. 4, No. 11, pp. 1465-1476.
- Heo, S., Kim, J., and Moon, T. (2018). "Predicting crime risky area using machine learning." *Journal of the Korean Association of Geographic Information Studies*, Vol. 21, No. 4, pp. 64-80.
- Hong, J.K., Kwon, H.J., Lim, J.H., Byun, Y.H., Lee, J.H., and Kim, J. (2009). "Standardization of KoFlux eddy-covariance data processing." *Korean Journal of Agricultural and Forest Meteorology*, Vol. 11, No. 1, pp. 19-26.
- Hunter, D., Yu, H., Pukish III, M.S., Kolbusz, J., and Wilamowski, B.M. (2012). "Selection of proper neural network sizes and architectures-A comparative study." *IEEE Transactions on Industrial Informatics*, Vol. 8, No. 2, pp. 228-240.
- Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., and Ermon, S. (2016). "Combining satellite imagery and machine learning to predict poverty." *Science*, Vol. 353, No. 6301, pp. 790-794.
- Jensen, J.R., Qiu, F., and Ji, M. (1999). "Predictive modelling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensor data." *International Journal of Remote Sensing*, Vol. 20, No. 14, pp. 2805-2822.
- Jeong, D., and Kang, J. (2009). "An analysis of changes in pan evaporation and climate values related to actual evaporation." *Journal of Korea Water Resources Association*, Vol. 42, No. 2, pp. 117-129.
- Kang, M., Kwon, H., Kim, J., Kim, H.S., Ryu, Y., Lee, S.J., and Choi, T. (2018). "Korean flux monitoring network's past, present, and future." *Korean Journal of Agricultural and Forest Meteorology*,

- Vol. 20, No. 1, pp. 1-4.
- Kendale, S., Kulkarni, P., Rosenberg, A.D., and Wang, J. (2018). "Supervised machine-learning predictive analytics for prediction of postinduction hypotension." *Anesthesiology*, Vol. 129, No. 4, pp. 675-688.
- Kim, K., Baik, J., Lee, J., Lee, Y., Jung, S., and Choi, M. (2016). "An assessment and analysis of the gap-filling techniques for revising missing data of flux tower based evapotranspiration-FAO-PM, MDV, and Kalman filter." *Journal of the Korean Society of Hazard Mitigation*, Vol. 16, No. 6, pp. 95-107.
- Kwon, H.J., Park, S.B., Kang, M.S., Yoo, J.I., Yuan, R., and Kim, J. (2007). "Quality control and assurance of eddy covariance data at the two KoFlux sites." *Korean Journal of Agricultural and Forest Meteorology*, Vol. 9, No. 4, pp. 260-267.
- Langley, P., and Simon H. (1995). "Applications of machine learning and rule induction." *Communications of the ACM*, Vol. 38, No. 11, pp. 54-64.
- Marin, F.R., Angelocci, L.R., Nassif, D.S., Costa, L.G., Vianna, M. S., and Carvalho, K.S. (2016). "Crop coefficient changes with reference evapotranspiration for highly canopy-atmosphere coupled crops." *Agricultural Water Management*, Vol. 163, pp. 139-145.
- Mitchell, T.M. (1997). *Machine learning*. The McGraw-Hill Companies Inc., Boston, U.S., pp. 52-78.
- Panchal, G., Ganatra, A., Kosta, Y.P., and Panchal, D. (2011). "Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers." *International Journal of Computer Theory and Engineering*, Vol. 3, No. 2, pp. 332-337.
- Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D. (2006). "Towards a standardized processing of net ecosystem exchange measured with eddy covariance technique: Algorithms and uncertainty estimation." *Biogeosciences*, Vol. 3, No. 4, pp. 571-583.
- Park, J., Baik, J., and Choi, M. (2017). "Satellite-based crop coefficient and evapotranspiration using surface soil moisture and vegetation indices in Northeast Asia." *Catena*, Vol. 156, pp. 305-314.
- Park, J., Byun, K., Choi, M., Jang, E., Lee, J., Lee, Y., and Jung, S. (2015). "Evaluation of statistical gap fillings for continuous energy flux (evapotranspiration) measurements for two different land cover types." *Stochastic Environmental Research And Risk Assessment*, Vol. 29, No. 8, pp. 2021-2035.
- Peterson, T.C., Golubev, V.S., and Groisman, P.Y. (1995). "Evaporation losing its strength." *Nature*, Vol. 377, No. 6551, pp. 687-688.
- Post, H., Franssen, H.H., Graf, A., Schmidt, M., and Vereecken, H. (2015). "Uncertainty analysis of eddy covariance CO₂ flux measurements for different EC tower distances using an extended two-tower approach." *Biogeosciences*, Vol. 12, No. 4, p. 1205.
- Roderick, M.L., and Farquhar, G.D. (2004). "Changes in Australian pan evaporation from 1970 to 2002." *International Journal of Climatology*, Vol. 24, No. 9, pp. 1077-1090.
- Sartori, M.A., and Antsaklis, P.J. (1991). "A simple method to derive bounds on the size and to train multilayer neural networks." *IEEE transactions on neural networks*, Vol. 2, No. 4, pp. 467-471.
- Sheela, K.G., and Deepa, S.N. (2013). "Review on methods to fix number of hidden neurons in neural networks." *Mathematical Problems in Engineering*, 2013, Article 425740.
- Tamura, S.I., and Tateishi, M. (1997). "Capabilities of a four-layered feedforward neural network: Four layers versus three." *IEEE Transactions on Neural Networks*, Vol. 8, No. 2, pp. 251-255.
- Yuan, R., Kang, M.S., Park, S.B., Hong, J.K., Lee, D.H., and Kim, J. (2007). "The effect of coordinate rotation on the eddy covariance flux estimation in a hilly KoFlux forest catchment." *Korean Journal of Agricultural and Forest Meteorology*, Vol. 9, No. 2, pp. 100-108.