

토양에 살포된 축산 분뇨로부터 암모니아 방출량 예측을 위한 인공신경망의 초매개변수 최적화와 데이터 증식

정평곤 · 임영일[†]

한경대학교 화학공학과 CoSPE 연구센터
17579 경기도 안성시 중앙로 327
(2022년 5월 30일 접수, 2022년 7월 14일 수정본 접수, 2022년 7월 14일 채택)

Hyperparameter Optimization and Data Augmentation of Artificial Neural Networks for Prediction of Ammonia Emission Amount from Field-applied Manure

Pyeong-Gon Jung and Young-Il Lim[†]

CoSPE, Dept. Chemical Engineering, Hankyong National University, Anseong, 17579, Korea
(Received 30 May 2022; Received in revised form 14 July 2022; Accepted 14 July 2022)

요 약

인공신경망을 이용한 모델 개발에서 데이터의 품질은 모델 성능에 큰 영향을 주고, 양질의 충분한 데이터가 인공신경망 훈련을 위해 필요하다. 하지만, 공학 분야에서는 적은 양의 데이터로 모델을 개발해야 하는 경우가 자주 발생한다. 본 논문은 토양에 살포된 축산 분뇨로부터 암모니아 방출량에 대한 적은 수의 데이터(83 개)를 사용하여 인공신경망 모델의 예측 성능을 향상할 수 있는 방안을 제시하였다. Michaelis-Menten 식으로 표현되는 암모니아 방출량 문제는 11개 입력변수에 대하여 2개 출력변수로 구성되었다. 출력변수는 최대 질소 발생량(N_{max} , kg/ha)과 N_{max} 의 절반에 도달하는 시간(K_m , h)이다. 범주형 입력변수에 대해 다차원 등간격 기법인 one-hot encoding 을 이용하여 데이터 전처리를 수행하였고, 훈련데이터 66개에 대하여 generative adversarial network (GAN)을 이용하여 13개 데이터를 추가로 보강하였다. 또한, 인공신경망의 초매개변수인 은닉층 수, 각 은닉층 내 뉴런 수, 활성화 함수의 최적 조합을 찾기 위하여 Gaussian process (GP)를 사용하였다. 기존의 인공신경망 구조(Lim *et al.*, 2007) 는 17개 평가데이터에 대하여 mean absolute error (MAE)는 K_m 에서 0.0668, N_{max} 에서 0.1860이었다. 본 연구에서 제시된 인공신경망 모델은 K_m 에서 0.0414, N_{max} 에서 0.0818로 MAE 가 기존 모델 대비 각각 38%, 56% 감소하였다. 본 연구에서 제시된 방법은 적은 양의 데이터를 갖는 문제에서 인공신경망 성능을 향상하기 위하여 활용할 수 있을 것이다.

Abstract – A sufficient amount of data with quality is needed for training artificial neural networks (ANNs). However, developing ANN models with a small amount of data often appears in engineering fields. This paper presented an ANN model to improve prediction performance of the ammonia emission amount with 83 data. The ammonia emission rate included eleven inputs and two outputs (maximum ammonia loss, N_{max} and time to reach half of N_{max} , K_m). Categorical input variables were transformed into multi-dimensional equal-distance variables, and 13 data were added into 66 training data using a generative adversarial network. Hyperparameters (number of layers, number of neurons, and activation function) of ANN were optimized using Gaussian process. Using 17 test data, the previous ANN model (Lim *et al.*, 2007) showed the mean absolute error (MAE) of K_m and N_{max} to 0.0668 and 0.1860, respectively. The present ANN outperformed the previous model, reducing MAE by 38% and 56%.

Key words: Livestock manure, Ammonia emission amount, Artificial neural network (ANN), Hyperparameter optimization, Data augmentation

1. Introduction

대형 농기계 및 화학비료 사용이 일반화되고 농경지와 주변 생태계 파괴, 그리고 토양침식으로 인한 지하수 오염이 심각한 수준으로 거론되면서 이에 대한 유기 농업의 필요성이 제기되어 왔다[1]. 유기 비료의 질소, 인 및 칼륨은 작물 생산을 위한 중요한 자원이며, 지속

[†]To whom correspondence should be addressed.

E-mail: limyi@hknu.ac.kr

‡이 논문은 광운대학교 교재교육 교수님의 정년을 기념하여 투고되었습니다.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

가능한 농업은 이들 물질에 대해 대기 및 지하수로의 손실을 최소화해야 한다[2]. 현대 농업에서 암모니아(NH_3), 온실가스, 휘발성유기화합물, 악취를 발생시키는 화합물 및 공기 중 미세 입자 등의 배출로 대기오염이 유발된다[3]. 유럽에서 질소화합물 대기 배출량의 90% 이상은 농업 활동에서 발생하고 있다고 예측된다[2].

토양에 살포된 축산 분뇨에서 배출되는 암모니아는 질소 침착, 산성화 및 대기 중 미세 입자 형성에 영향을 미친다[3]. 산성비를 유발하고, 가축 분뇨에 잠재되어 있는 질소성분의 손실이 되는 암모니아 방출량을 예측하고 감소시키려는 많은 연구가 진행되었다[1,3,4]. 살포된 분뇨의 암모니아 방출량은 토양상태, 날씨, 분뇨 특성, 시비 방법 등의 복잡한 상호작용의 영향을 받는다[1]. Lim et al.(2007)[1]은 축산 분뇨로부터 암모니아 방출량의 예측을 위하여 83개 ALFAM project 데이터를 활용한 인공신경망(artificial neural network, ANN) 기법을 제안하였다. 하지만, 이 인공신경망 기법에서는 83개 모든 데이터를 훈련데이터로 활용함으로써 평가 데이터 안에 훈련데이터를 포함하고 있어 제시된 인공신경망 구조가 과적합(over-fitting)의 가능성을 갖고 있으며, 새로운 입력조건에 대한 예측 성능을 저하시킬 우려가 있다.

최근 기계학습 알고리즘은 비약적인 발전을 거듭해 왔고, 주어진 데이터의 전처리를 위한 정규화 기법[5], 주성분분석(principal component analysis, PCA)[6], 불연속 이산값에 대한 다차원 중간격 코딩 기법인 one-hot encoding[7] 등이 개발되었다. 또한, 인공신경망의 초매개변수인 은닉층(hidden layer) 수, 각 은닉층의 뉴런 수, 활성화 함수 등의 최적화를 위한 Gaussian process (GP) [8]을 통하여 효과적으로 인공신경망 구조를 구축할 수 있다.

데이터 기반 실제 공학 문제에서는 데이터 수가 충분하지 않거나, 불균형 분포를 보일 수 있다[9-12]. 기계학습 모델의 성능은 주어진 데이터의 품질의 영향을 받기 때문에[13], 편향되거나 너무 적은 데이터는 기계학습 모델의 예측 성능을 떨어뜨리는 요인 중 하나이다[9]. 실제 데이터는 인공신경망이 학습하기에 충분치 않거나, 일부 목표 값에 치중되어 있어, 특정 목표 값이 훨씬 적은 불균형 분포를 나타내는 경우가 많다[12]. 불균형 데이터는 인공신경망 모델에 특정 목표 값에 대한 과적합이나 과소적합(under-fitting)과 같은 문제를 야기시킴으로 데이터 불균형을 해결하기 위한 연구가 필요하다. 불균형 분류(classification) data set의 보강을 위해 자주 사용되는 synthetic minority over-sampling technique (SMOT)과 k -nearest neighbor (KNN) 기법은 적은 데이터 수로 구성된 그룹에 대하여 데이터간 선형으로 보간된 가상 데이터를 생성한다[14,15]. 이를 통하여 적은 데이터 수로 구성된 그룹의 데이터가 보강된다. 하지만, 본 연구의 암모니아 방출량 데이터는 분류가 아닌 연속형 데이터의 회귀(regression)를 위하여 사용됨으로, 불균형 연속형 데이터에 대한 보강법이 요구된다[12,16]. 원래 데이터의 통계적 특성을 가진 가상 데이터 생성을 위한 generative adversarial network (GAN) 기법은 불균형 데이터의 증식을 위하여 사용되었다[17,18].

인공신경망 모델의 성능은 연구자가 설정해야 하는 초매개변수(은닉층 수, 각 은닉층의 뉴런 수, 활성화 함수 등)에 영향을 받으며, 여러 초매개변수의 조합을 모두 고려하기란 사실상 불가능에 가깝다[19]. 실제로 많은 연구에서 시행착오법을 이용하여 최적의 매개변수 조합을 선정한다[19,20]. 그러나 이러한 방법은 시간과 자원을 많이 소모하며, 시행착오법으로 얻어진 모델은 최적의 모델이 아닐 수도 있다[21]. 따라서 초매개변수를 연구자가 직접 조정하는 것이

아닌 최적화 기법을 활용하여 초매개변수를 설정하는 연구가 진행되었다[19-22].

초매개변수 최적화 방식 중 하나인 실험 계획법(design of experiment)은 1920년대 주로 농업용으로 개발했지만, 통계적 방법으로서 과학 및 산업의 다양한 분야에서 활용되고 있다[23]. 실험 계획법은 적은 수의 관측 혹은 실험으로 최적의 변수 조합을 찾는 것을 목적으로 하며, Bergstra et al. [20]는 초매개변수 최적화를 위하여 많이 사용되는 기법들, Gaussian process, tree-structured Parzen estimator, random search 등을 소개하였다. Yang et al. [22]는 기계학습 모델의 최적화에 대한 이론과 활용을 중점적으로 설명하였고, Du et al. [19]은 4가지 다른 데이터에 대하여 초매개변수 최적화를 진행하여 최적화 전후의 모델 성능을 비교하였다. Yang et al. [21]는 순차적인 방법을 이용하여 초매개변수 공간을 탐색하여 같은 비용에서 더욱 우수한 성능을 보이는 실험 계획법을 제안하였다.

본 연구의 목표는 데이터 수가 적고, 불균형 분포를 보이는 암모니아 방출량 문제에 대하여 과적합이 최소화된 인공신경망 기법을 제시하는 것이다. 과적합을 최소화하기 위하여 학습 데이터는 서로 다른 훈련 데이터와 평가 데이터로 구분하고, 인공신경망의 초매개변수는 GP 를 이용하여 최적화되며, 불균형 데이터의 보강을 위해 GAN 기법을 활용한다. 본 논문의 2장에서는 축산분뇨로부터 암모니아 방출량 예측 모델을 소개하고, ALFAM 데이터 구조를 설명한다. 3장에서는 본 연구에서 사용될 인공신경망 기법을 기술하고, 4개의 비교 연구를 제시한다. 4장에서는 4개의 인공신경망 모델에 따른 암모니아 방출량 예측 성능 결과를 비교하고, 평가한다. 5장은 본 연구의 결론이다.

2. Ammonia emission rates from field-applied manure

Michaelis-Menten 방정식은 암모니아 방출량을 예측하는 데 사용되는 모델 중 하나로서[24], 다음과 같이 표현된다[1]:

$$N_{\text{NH}_3}(t) = N_{\text{max}} \frac{t}{t + K_m} \quad (1)$$

$N_{\text{NH}_3} \left[\frac{\text{kg}}{\text{ha}} \right]$ 는 시간 t 동안 누적된 질소 방출량을 뜻하고, $N_{\text{max}} \left[\frac{\text{kg}}{\text{ha}} \right]$ 는 최대 질소 방출량이다. $K_m[\text{h}]$ 는 최대 암모니아 방출량의 절반($N_{\text{max}}/2$)에 도달하기까지 걸리는 시간이다. 두 개의 변수(N_{max} 와 K_m)는 실험 데이터로부터 추정되며, 입력변수인 토양 조건, 기후, 시비법 그리고 경작지 조건에 의존한다[1]. 본 연구에서는 11개 입력변수에 따른 Michaelis-Menten 모델의 매개변수(N_{max} 와 K_m)에 대한 83개 ALFAM 데이터를 사용하였다(Appendix A1 참조)[1].

2-1. Input variables of ammonia emission rates

Table 1은 11개 입력변수의 범위와 범주형 데이터 값들을 보여준다. 83개의 데이터에 대한 평균 값은 $N_{\text{max}} = 25.9 \text{ kg/ha}$ 및 $K_m = 8.4 \text{ h}$ 이다. 여러 유럽 국가에서 서로 다른 조건에서 실험이 수행되었기 때문에 두 매개변수의 표준 편차는 높고, 입력변수의 값에 따라 불균형한 분포를 보여준다(Appendix A1 참조).

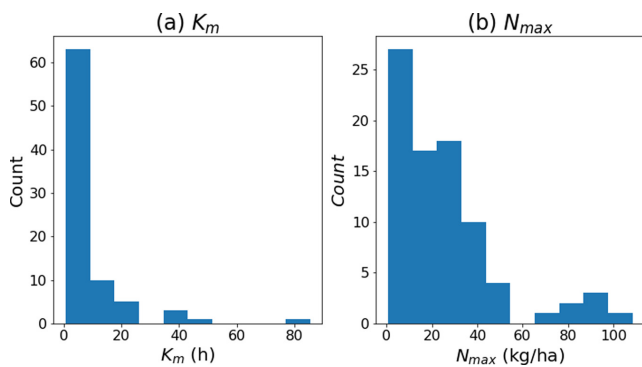
2-2. Characteristics of given data sets

본 논문에서 사용되는 83개 데이터의 입력변수 중 soil type, manure application method, crops type와 measuring technique 4개 입력 변

Table 1. Eleven input variables influencing ammonia emission rates from field-applied manure [1]

| Input variable | Index | Range | Category description or units |
|----------------------------|----------|-------------|---|
| Soil type | p_1 | [1, 3] | 1=sandy, 2=clay, 3=loam |
| pH of soil | p_2 | [5, 8] | |
| Air temperature | p_3 | [0, 28] | °C |
| Wind speed | p_4 | [0, 6.5] | m/s |
| Dry matter in manure | p_5 | [1.0, 11.5] | % |
| TAN ^a in manure | p_6 | [1.5, 6.5] | (g-nitrogen)/(kg-manure) |
| pH of manure | p_7 | [6.5, 8] | |
| Manure application method | p_8 | [0, 3] | 0=broad spread, 1=band spread, 2=trailing shoe, 3=open-slot injection |
| Manure application rate | p_9 | [7.5, 60] | (ton-slurry)/(ha-field) |
| Crops type | p_{10} | [1, 4] | 1=grass, 2=stubble, 3=bare soil, 4=growing crops |
| Measuring technique | p_{11} | [1, 3] | 1=wind tunnel, 2=micrometeorological mass balance technique, 3=equilibrium concentration method |

^aTAN: total ammoniacal nitrogen ($=\text{NH}_3 + \text{NH}_4^+$)

**Fig. 1. Histogram for distribution of target values (K_m and N_{max}).**

수가 정수형이며, 범주형 데이터이다. 이러한 범주형 데이터는 상호 간 균등한 지위에 있음에도 불구하고, 정수형으로 표현됨으로 서로 간의 거리가 다르게 해석될 여지가 있다[25]. 또한, 다음 Fig. 1에서 보듯이 본 연구에서 사용된 데이터의 목표값(K_m , N_{max})은 정규분포를 따르지 않고 한 쪽에 편향되어 있으며, 데이터가 없는 구간도 존재한다. 이러한 두가지 문제점을 해결하기 위하여 다차원 등간격 one-hot encoding [7] 및 불균형 데이터 보강을 위한 GAN 기법[18]을 사용할 것이다.

3. Artificial neural network (ANN) model

인공신경망은 비선형 기계 학습 모델로서 입력층, 여러 개의 은닉층과 출력층으로 이루어져 있다. 인공신경망의 각 층은 상호 연결되어 있는 뉴런들로 구성된다. 각 뉴런들은 이전 뉴런의 출력값(p)에 가중치(w)를 곱하고, 편향(b)을 더한 후, 활성화 함수(f)로 계산된 값(q)을 출력한다[26].

$$q = f(b + \mathbf{w}^T \mathbf{p}) \quad (2)$$

활성화함수는 rectified linear unit (ReLU)와 sigmoid, tanh 함수 등이 있다[27]. 이러한 활성화함수는 뉴런의 출력에 비선형성을 추가하는 역할을 한다[26]. 입력층과 출력층의 뉴런 수는 각각 입력과 출력 변수의 수와 동일하다[1]. 주어진 인공신경망의 가중치와 편향은 훈련데이터를 사용하여 학습된다. 학습 방법은 무작위로 초기화

된 가중치와 편향에서 출발하여 인공신경망으로 계산된 출력값과 실제값과의 차이값이 최소화 되도록 가중치와 편향을 조정하는 것이다[28]. 본 연구는 평균 절대 오차(Mean absolute error; MAE)와 상관계수(R^2)를 인공신경망의 성능 지표로 사용하여 인공신경망 구조 및 인자값들을 최적화 한다.

$$\min_{w,b} \text{MAE} = \frac{\sum_{i=1}^n |X_i - Y_i|}{n} \quad (3)$$

$$R^2 = \left(\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \right)^2 \quad (4)$$

위 식에서 X_i 와 Y_i 는 K_m 과 N_{max} 에 대한 실제값과 인공신경망 계산값이다. n 은 데이터 수이며, \bar{X} 와 \bar{Y} 는 각각 X_i 과 Y_i 의 평균이다.

83개 데이터를 66개 훈련데이터와 17개의 평가데이터로 나누었고, 66개 훈련데이터는 5-fold cross validation [29]을 이용하여 인공신경망을 학습시켰다. 인공신경망의 초매개변수는 MAE를 손실함수로 사용하여 Cases 1-2에서는 Lim et al. (2007)에서 사용된 stochastic gradient descent (SGD)를, 그리고 Cases 3-4에서는 resilient backpropagation (Rprop) 최적화 해석기(optimizer)를 이용하여 학습을 진행하였다. 11개 입력변수에 대하여 출력변수인 K_m 과 N_{max} 의 83개 데이터 분포는 Appendix A1에 있는 Fig. A1.1에서 보여준다. 평가데이터(붉은색 점)는 비교적 균일하게 분포되어 있지만, 너무 적은 데이터수로 인하여 입력변수에 따라서 비균일하게 샘플링되었다.

3-1. Data preprocessing

본 논문에서는 인공신경망 모델의 성능 향상을 위하여 4가지 전처리 기법을 사용하였다: (1) Maximum normalization(최대 정규화), (2) PCA(주성분 분석), (3) One-hot encoding, (4) 데이터 보강을 위한 GAN.

3-1-1. Maximum normalization

입력 데이터에 대한 정규화는 학습 시간을 단축하는 것에 도움이 된다[30]. 주로 사용되는 정규화 기법에는 최대 정규화, 최소-최대 정규화와 Z-점수 정규화(Z-score normalization)[5]가 있으며, 본 논문에서는 데이터의 최대값을 1로 조정하는 최대 정규화를 사용하였다.

$$\bar{x}_i = \frac{x_i}{x_{\max}}, x_{\max} = \max_{1 \leq i \leq n} x_i \quad (5)$$

위 식에서 x_i 는 입력 변수이고, x_{\max} 는 x_i 값의 최대값이다. 최소-최대 정규화가 아닌 최대 정규화를 사용한 이유는 입력 매개변수 중 soil pH (p_2)와 manure pH (p_7)의 경우, 이들의 범위는 83개의 데이터에서 $5 \leq p_2 \leq 8$, $6.5 \leq p_7 \leq 8$ 이지만, 일반적인 pH는 0과 14 사이임으로, 새로운 데이터가 입력될 때 음수값으로 정규화되는 것을 방지하기 위함이다.

3-1-2. Principal component analysis (PCA)

기계학습에서 입력변수의 종류가 많을 수록 학습 시간이 증대하고, 입력변수간 상관 정도(다중 공선성)가 높을 수록 출력변수에 대한 예측 정확도가 저하된다[1]. 주성분 분석은 p 개의 입력변수로 구성된 데이터에 대하여 공분산도를 계산하고, 데이터의 분산을 최대한 유지하는 새로운 축을 찾아 p 보다 작은 수의 입력변수 축에 데이터를 투영하는 기법이다. 따라서 PCA의 주된 목적은 입력변수의 차원을 축소시키는 데 있다[31]. 하지만, 데이터의 수가 적은 경우, 좋은 성능을 보이지 않고, 데이터가 정규분포를 갖지 않을 경우, PCA를 적용하기가 어렵다 [32]. 본 논문에서는 데이터수가 적은 암모니아 방출량 예측문제에서 주성분분석의 사용 유무에 따른 인공지능망 성능을 비교한다.

3-1-3. One-hot encoding

One-hot encoding (OHE)은 어떤 범주형 입력 변수내 데이터 값을 데이터 값의 수(m)에 해당하는 다차원 이진수 벡터로 변형하는 기법이다. 범주형 값은 각 범주 사이의 간격이 동일하게 취급되어야 하지만, 정수로 변환할 경우 각 범주 사이의 간격이 달라진다. 즉, 1, 2, ..., m 의 형태로 범주형 값을 나타내면 첫 번째 특징과 두 번째 특징 사이의 거리는 1이지만, 첫 번째와 세 번째 특징 사이의 거리는 2가 된다. OHE의 경우 m 차원 공간에서 각 차원별로 1의 거리에 위치하도록 해당 입력변수의 데이터 값을 변환한다.

Table 2는 범주형 입력변수인 soil type (Table 1에서 p_1)에 대하여

Table 2. One-hot encoding for Soil type

| Soil type | Integer encoding | One-hot encoding |
|-----------|------------------|------------------|
| Sandy | 1 | [1, 0, 0] |
| Clay | 2 | [0, 1, 0] |
| Loam | 3 | [0, 0, 1] |

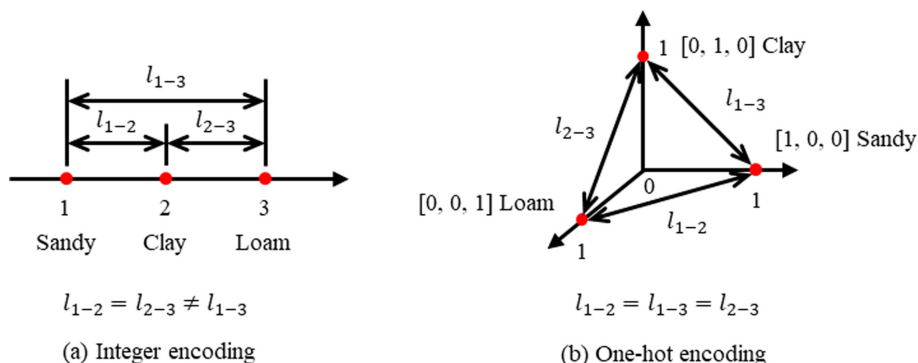


Fig. 2. Diagram of Euclidean distances for soil type by integer and one-hot encodings [7].

integer encoding과 OHE의 표현 방식을 비교한다. Fig. 2는 두 코딩 방식을 그림으로 표현한다. Fig. 2a에서 보듯이 integer encoding에서는 sandy와 clay 사이의 거리(l_{1-2})는 sandy-loam 사이의 거리(l_{1-3})와 다르며, l_{1-3} 는 clay-loam 사이의 거리(l_{2-3})와 다르다. 하지만, OHE는 soil type의 세 범주가 3차원상에서 등간격을 유지한다(Fig. 2b).

본 논문에서는 soil type (p_1), manure application method (p_8), crops type (p_{10})과 measuring technique (p_{11}) 4개 범주형 입력 변수에 대해 OHE를 적용하여 10개의 입력변수로 변형하였다. 즉, 3개의 값을 갖는 하나의 입력변수(p_1 과 p_{11})에 대하여 [0, 0], [1, 0], [0, 1]로 표현 가능함으로, p_1 과 p_{11} 은 각각 2개, p_8 과 p_{10} 은 각각 3개의 입력 변수로 확장하였다. 따라서 초기의 11개 입력변수를 17개 입력변수로 확장된 차원으로 전처리를 수행하였다.

3-1-4. Generative adversarial network (GAN)

Generative adversarial network (GAN)는 반복적이고 수치적인 학습 기법으로, Fig. 3에서 보여주듯이 생성기(G)와 판별기(D)를 포함하는 두 개의 인공지능망으로 구성된다. 원본 데이터의 통계적 특성을 계승하는 가상 데이터의 생성을 위한 신경망 구조이고[17], 다음과 같은 수식으로 표현된다[18].

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z)))] \quad (6)$$

여기서 V 는 최소-최대 최적화 문제의 목적함수, x 는 실제 데이터, z 는 무작위로 생성된 노이즈, $G(z)$ 는 노이즈로 생성된 가상 데이터, $D(x)$ 가 실제 데이터라고 분류할 확률이고, $D(G(z))$ 는 판별기가 가상 데이터라고 분류할 확률이다. 또한 $p_{data}(x)$ 는 원본 데이터의 분포이고, $p_z(z)$ 는 노이즈의 분포를 뜻한다. Fig. 3은 생성기(G)와 판별기(D) 모형도이다. G 에서 노이즈를 통해 실제와 가까운 형태의 가상 데이터가 생산되고, D 는 G 가 생산한 데이터를 검출하면서 서로 경쟁적으로 모델을 향상시키는 기법이다[18].

GAN 모델은 Python을 통해 구현되었고, Fig. 4는 GAN의 생성기와 판별기에 대한 인공지능망 구조를 보여준다. G 는 2-64-128-19의 구조를 갖는다. 이때, G 의 출력 변수 19개는 암모니아 방출량에 영향을 주는 입력변수 17개, 출력변수(K_m and N_{max}) 2개에 해당한다. D 는 19-512-256-64-1의 구조를 갖고, G 에서 생성된 가상 데이터와 실제 데이터를 훈련데이터로 사용한다. G 에서 생성된 가상 데이터 500개와 실제데이터 66개를 이용하여 GAN의 weight와 bias를 결정하였다. 이러한 weight와 bias를 사용하여 또다른 random noise 13개를 G 의 입력변수로 제공하고, 13개 가상 데이터(66개 훈련 데이

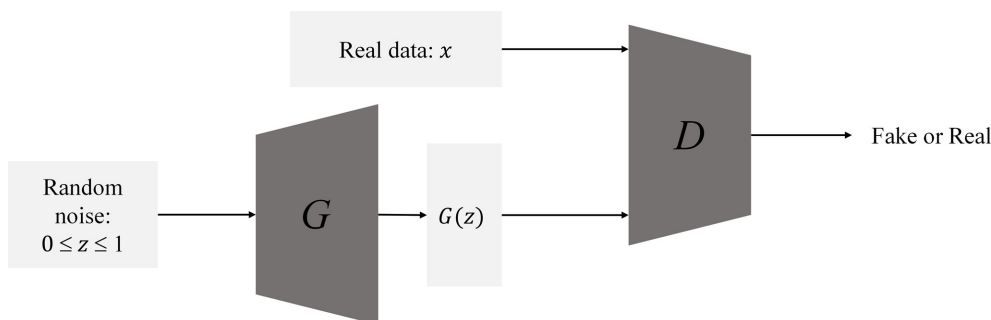


Fig. 3. Diagram of generative adversarial network (GAN).

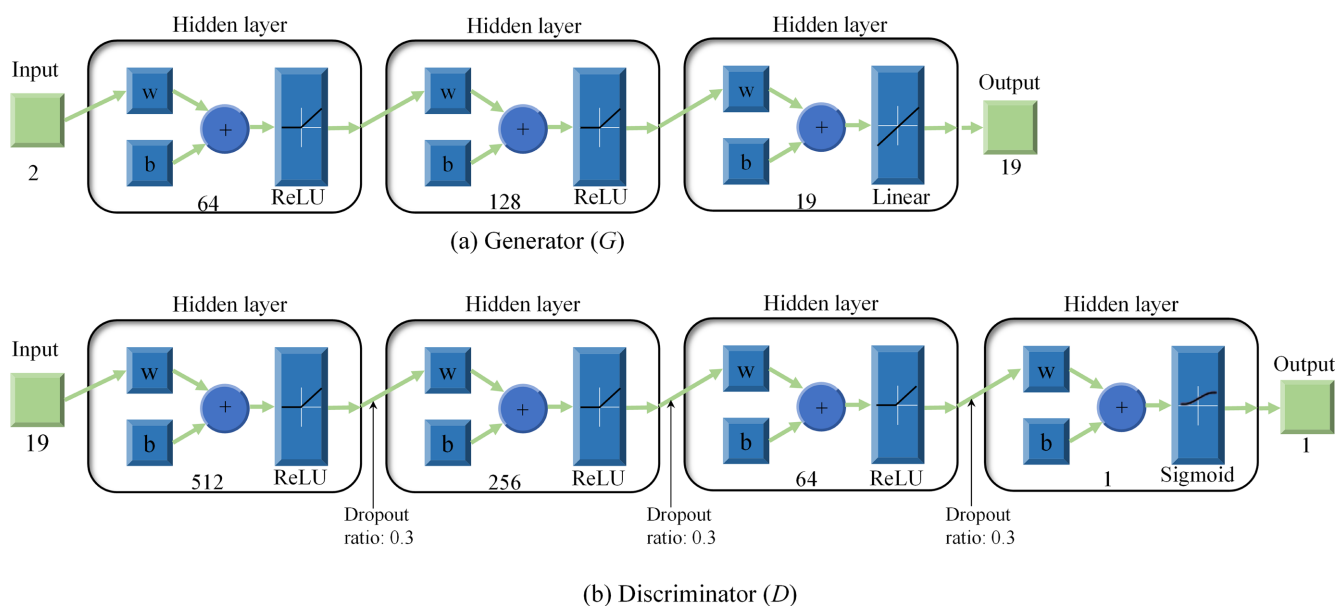


Fig. 4. Architectures of the generator (G) and discriminator (D) networks.

터의 20%)를 생성하였다.

Appendix A2는 66개 기존 훈련데이터, GAN으로 생성된 13개 가상 훈련데이터, 17개 평가데이터를 포함하는 총 96개 데이터를 보여준다(Table A2.1). Fig. A2.1는 입력변수와 출력변수에 따른 96개 데이터의 분포를 보여준다. 초록색으로 구분된 가상데이터는 기존 데이터의 분포에 순응하면서, 일부 구간에서 부족한 데이터를 보충한다.

3-2. Hyperparameter optimization

초매개변수(hyper-parameter)는 인공지능망 구조와 설정을 위한 인자들로서 사용자가 초매개변수들을 결정해야 한다. Lim et al. (2007)[1]에서는 은닉층의 수, 각 은닉층의 뉴런 수, 활성화 함수 등은 모든 영역 탐색을 통하여 최소 오차를 보이는 조건에서 정해졌다. 하지만, 이 방법은 초매개변수 수가 많고, 은닉층 수가 증가하면서 계산 시간이 많이 걸린다. 본 연구에서는 Gaussian process를 이용하여 체계적으로 초매개변수 최적화하고, 이전 방법과 성능 지표를 비교한다.

3-2-1. Gaussian process (GP) and Expected Improvement (EI)

Bayesian 최적화를 기반으로 하는 초매개변수 최적화 방법은 인

공신경망 구조의 최적화를 위하여 많이 사용되어 왔다[8]. Bayesian 최적화 기법 중의 하나인 Gaussian process(GP)는 다음과 같다[33].

$$f(x) \sim \mathcal{GP}(m(x), k(x_p, x_j)) \quad (7)$$

식 (7)의 \mathcal{GP} 는 모든 매개변수 지점, $x \in \mathcal{X}$ 에 대해 평균 $m(x)$ 과 공분산 $k(x_p, x_j)$ 에 의해 정의되고, 탐색하지 않은 각 초매개변수 구성 x 에 대한 $f(x)$ 값은 평균 및 분산에 따라 계산된다. 다양한 Acquisition function을 GP 모델과 함께 사용하여 유망한 초매개변수 구성을 추정한다. Acquisition function 중 하나인 expected improvement (EI)는 다음과 같다[33].

$$\alpha_{EI}(x) = \mathbb{E}[\max(f(x) - y^+, 0)] \quad (8)$$

$$x_i = \operatorname{argmax}_{x \in \mathcal{X}} \alpha(x) \quad (9)$$

위 식에서 y^+ 는 $\max_{i \leq j} (y_i)$ 로 정의되고, $y_i = f(x_i)$ 이며, 현재까지 탐색된 가장 높은 성능값이다. GP-EI의 매 반복마다 $f(x)$ 에 근거하여 $\alpha_{EI}(x)$ 를 도출하고, 식 (9)를 통해 초매개변수의 최적 조합을 제시한다.

Fig. 5는 GP-EI를 사용하여 초매개변수의 조합을 찾아내는 것을 2차원 그래프로 나타낸다. 실제값은 빨간 채선으로 표현되어 있고 (Fig. 5a, 5c, 5e, 그리고 5g 참조), 다음과 같이 정의되었다.

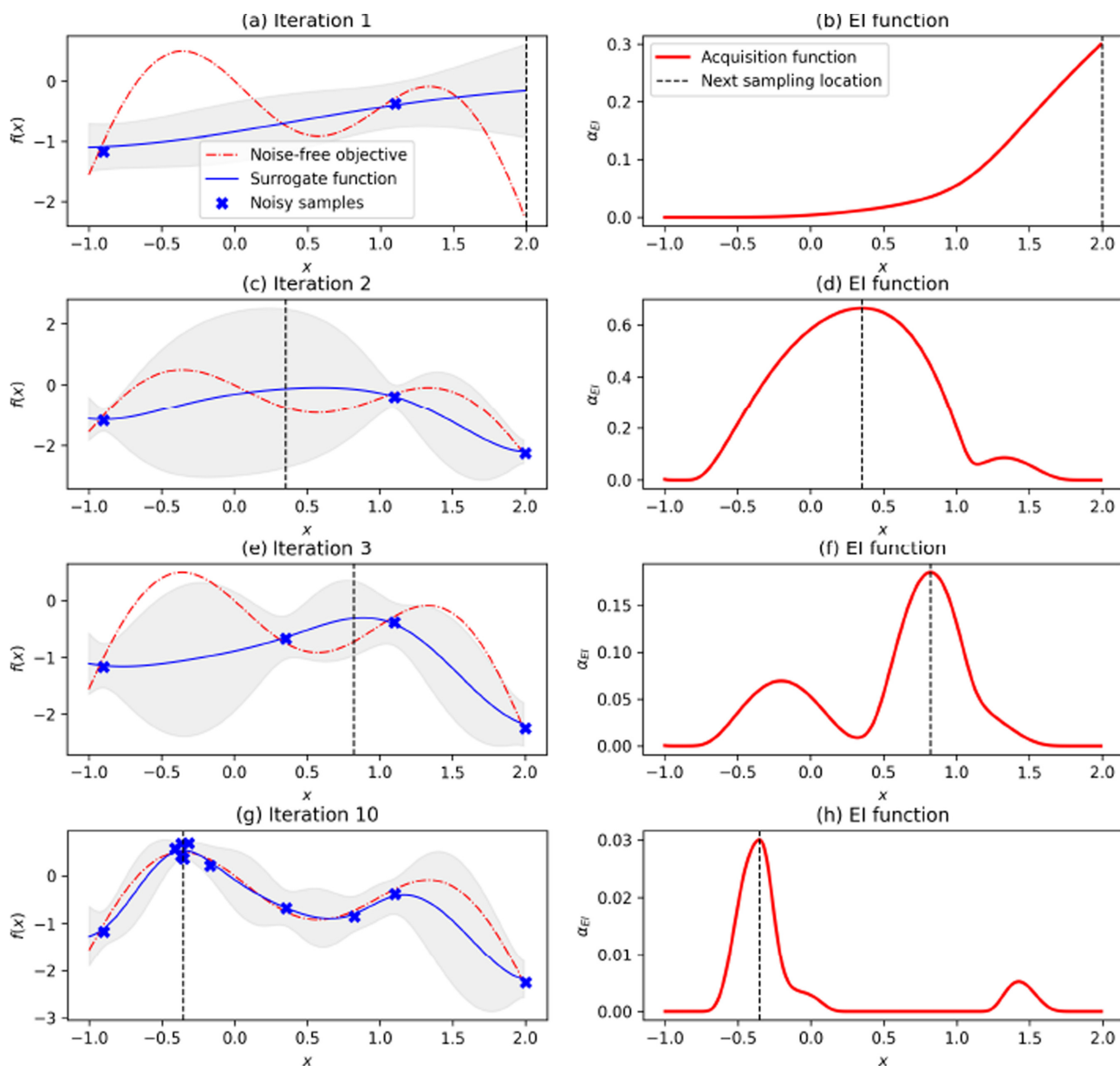


Fig. 5. Procedure of Gaussian process with expected improvement (GP-EI) for hyperparameter optimization.

$$y = -\sin 3x - x^2 + 0.7x \quad (10)$$

각 반복에서 탐색한 초매개변수(x)를 이용하여 surrogate function (파란색 실선)을 정의한다. 이때 GP는 Gaussian 분포를 이용하여, 탐색하지 못한 초매개변수에 대해 불확실성(회색 그림자 영역)을 설정한다(Fig. 5a). EI는 이러한 불확실성을 바탕으로 가장 높은 성능($f(x)$)을 보일 것이라 예상되는 지점(검은색 점선)을 도출하고(Fig. 5b), 높은 성능을 보일 것이라고 예상되는 지점을 다음 반복에서 탐색한다. 다음 반복에서 해당 지점을 탐색하고, 탐색된 지점들(noisy samples)을 바탕으로 다시 surrogate function을 재정의 한다(Fig. 5c). 이전 과정과 마찬가지로 다음 유망 지점을 도출하는 것을 반복(Fig. 5d)하며, 최적의 매개변수 조합을 탐색한다. Fig. 5g는 GP-EI가 실제값(noise-free objective)에 근사한 초매개변수 조합을 탐색한 것이다.

본 논문에서 초매개변수는 은닉층 수(1-3), 은닉층의 뉴런 수(4-32개), 활성화 함수(ReLU, tanh, 그리고 sigmoid), 최적화 해석기(SGD, Adam, 그리고 Rprop)이고, 이들 초매개변수에 대한 모든 조합의 수는 227,331 개이다. 하지만, GP-EI를 통하여 1,000개의 초매개변수 조합을 탐색하였다.

3-3. Four case studies and ANN structures

토양에 살포된 가축 분뇨로부터 암모니아 방출량 예측을 위한 4개의 인공신경망을 구축하여 각 인공신경망의 성능을 비교하였다(Table 3 참고). Case 1은 데이터 전처리로 max normalization과 PCA를 사용하였고, Lim et al. (2007)[1]에서 제시된 구조 및 초매개변수를 그대로 사용하였다. Cases 2-4는 데이터 전처리로 max normalization과 OHE를 사용하였고, Case 4에서는 추가적으로 GAN을 이용하여 13개 훈련데이터를 증식하였다. Case 2의 인공신

Table 3. Four case studies of ANN for ammonia emission rate from field-applied manure

| Cases | Data preprocessing | ANN structure | Remarks |
|--------|---|-----------------|---|
| Case 1 | Max normalization and PCA | 1 hidden layer | Lim et al. (2007) |
| Case 2 | Max normalization and OHE | 1 hidden layer | |
| Case 3 | Max normalization and OHE | 3 hidden layers | Using GP-EI for hyperparameter optimization |
| Case 4 | Max normalization, OHE, and data augmentation | 3 hidden layers | Using GP-EI for hyperparameter optimization |

Table 4. Prediction performance of four case studies for ammonia emission rate using test data

| Performance index | MAE | | R^2 | |
|----------------------------|--------|-----------|--------|--------|
| Output variables | K_m | N_{max} | K_m | |
| Case 1 (Lim et al. 2007) | 0.0668 | 0.1860 | 0.0045 | 0.0127 |
| Case 2 (OHE) | 0.0546 | 0.1605 | 0.1632 | 0.4361 |
| Case 3 (OHE + GP-EI) | 0.0431 | 0.1006 | 0.3896 | 0.8834 |
| Case 4 (OHE + GAN + GP-EI) | 0.0414 | 0.0818 | 0.4163 | 0.9329 |

경망 구조 및 초매개변수는 Lim et al. (2007)[1]과 동일하고, Cases 3-4는 3개 은닉층을 사용하였고, 초매개변수들이 GP-EI를 이용하여 최적화 되었다.

Cases 1-4 모든 경우에서 출력층의 뉴런 수는 2개이고, 입력층의 뉴런 수는 Case 1에서는 PCA로 차원이 축소되어 10개, Case 2-4에서는 OHE의 사용으로 17개 뉴런을 가진다. Cases 1-2의 은닉층은 26개의 뉴런을 갖고, 활성화 함수는 sigmoid이다. Case 2는 범주형 데이터의 대한 전처리 기법의 비교 검증을 위하여 PCA가 아닌 OHE를 이용하였다.

4. Results and Discussion

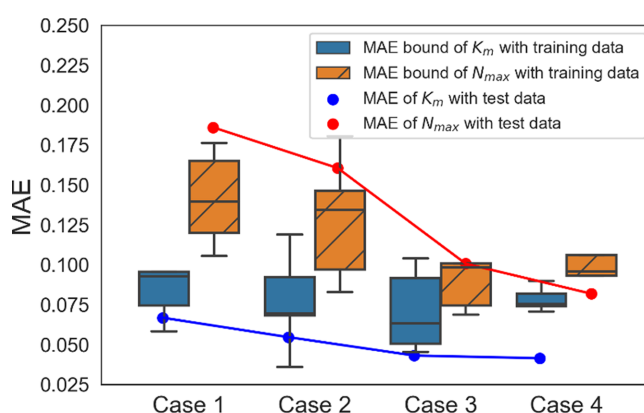
Table 3에서 보여주는 4개 사례연구에 대하여 암모니아 방출량 예측 성능 및 Cases 3-4에서 사용된 인공신경망 구조에 관하여 기술한다.

4-1. Prediction performance of four cases for ammonia emission rate

Table 4는 Cases 1-4에 대하여 암모니아 방출량을 결정하는 출력 변수 K_m 과 N_{max} 의 MAE (식 (3) 참조)와 상관계수 R^2 (식 (4) 참조)를 보여준다. 앞서 설명하였듯이 두 개의 성능지표(MAE와 R^2)는 훈련 데이터에서 사용되지 않은 평가데이터 17개를 이용하여 평가되었다.

Lim et al. (2017)에서 제시된 인공신경망을 사용할 경우(Case 1), 사용된 평가데이터에 대하여 높은 MAE와 낮은 R^2 값을 보여준다. 범주형 입력변수를 OHE를 사용하여 다차원 등간격으로 변환한 Case 2는 Case 1과 같은 인공신경망 구조를 활용했음에도 성능지표가 향상되었다. GP-EI 기법으로 초매개변수를 최적화하여 3개 은닉층을 사용한 Case 3은 1개 은닉층을 사용한 Cases 1-2 보다 상당히 향상된 예측 성능을 보여준다. GAN기법을 사용하여 13개 증식데이터를 추가하여 훈련된 Case 4는 Table 4에서 제시한 4개 사례연구 중에서 가장 적은 오차(MAE)와 가장 높은 상관계수(R^2)를 보여준다.

Fig. 6은 4개 사례연구에 대하여 K_m 과 N_{max} 의 MAE 분포를 보여준다. K_m 의 MAE는 N_{max} 의 MAE 보다 작다. 사각형은 훈련데이터 66개의 5-fold cross-validation에서 얻은 MAE의 중앙값 대비 $\pm 25\%$ 신뢰구간을 보여주며, MAE의 최대 및 최소값은 error bar로 표시되어 있고, 사각형 내 선은 평균값을 의미한다. 실선은 17개 test data

**Fig. 6. Error bounds of mean absolute error (MAE) for four cases.**

로부터 얻은 MAE 값을 보여준다. 13개 증식 데이터가 추가된 Case 4의 MAE 신뢰구간 폭이 Case 3의 신뢰구간 폭보다 작기 때문에, Case 4의 암모니아 방출량 예측값이 훈련데이터에 대하여 보다 균일하게 추정되었다. 17개 평가 데이터에 대한 MAE는 Case 4에서 최소를 보이므로 GAN을 통해 데이터를 증식한 인공신경망의 학습이 더 효과적일 수 있음을 알 수 있다.

Fig. 7은 4개 인공신경망 구조에 대하여 K_m 과 N_{max} 의 평가데이터 실제값과 예측값을 비교한다. Cases 1-2의 경우, 83개 모든 데이터를 이용하여 학습된 과적합된 ANN 구조에서 임의의 17개 평가 데이터에 대하여 얻은 상관계수값(R^2)임으로 R^2 이 낮고, 대각의 등가선에서 많이 벗어나 있다(Fig. 7a-7d 참조). Cases 3-4는 66개 혹은 79개 훈련데이터를 이용하여 초매개변수가 최적화 되었고, Cases 1-2에서 사용된 동일한 평가데이터에 대하여 상대적으로 높은 R^2 값을 보여준다(Fig. 7a-7d 참조). Case 4는 증식된 훈련데이터로 인하여 향상된 실제값과 예측값 간의 상관계수값을 보여준다.

4-2. Architecture of optimized ANN for Cases 3 and 4

Cases 3-4의 인공신경망 구조는 GP-EI를 이용하여 최적화 되었고, Fig. 8-9에서 은닉층별 뉴런 수와 활성화 함수를 보여준다. Case 3의 3개 은닉층은 모두 ReLU 활성화 함수를 사용했고, 뉴런수는 각각 32, 26, 11이다. 2개의 출력변수를 계산하는 출력층에서는 활성화 함수를 사용하지 않았다. Case 4의 3개 은닉층은 2개의 ReLU 활성화

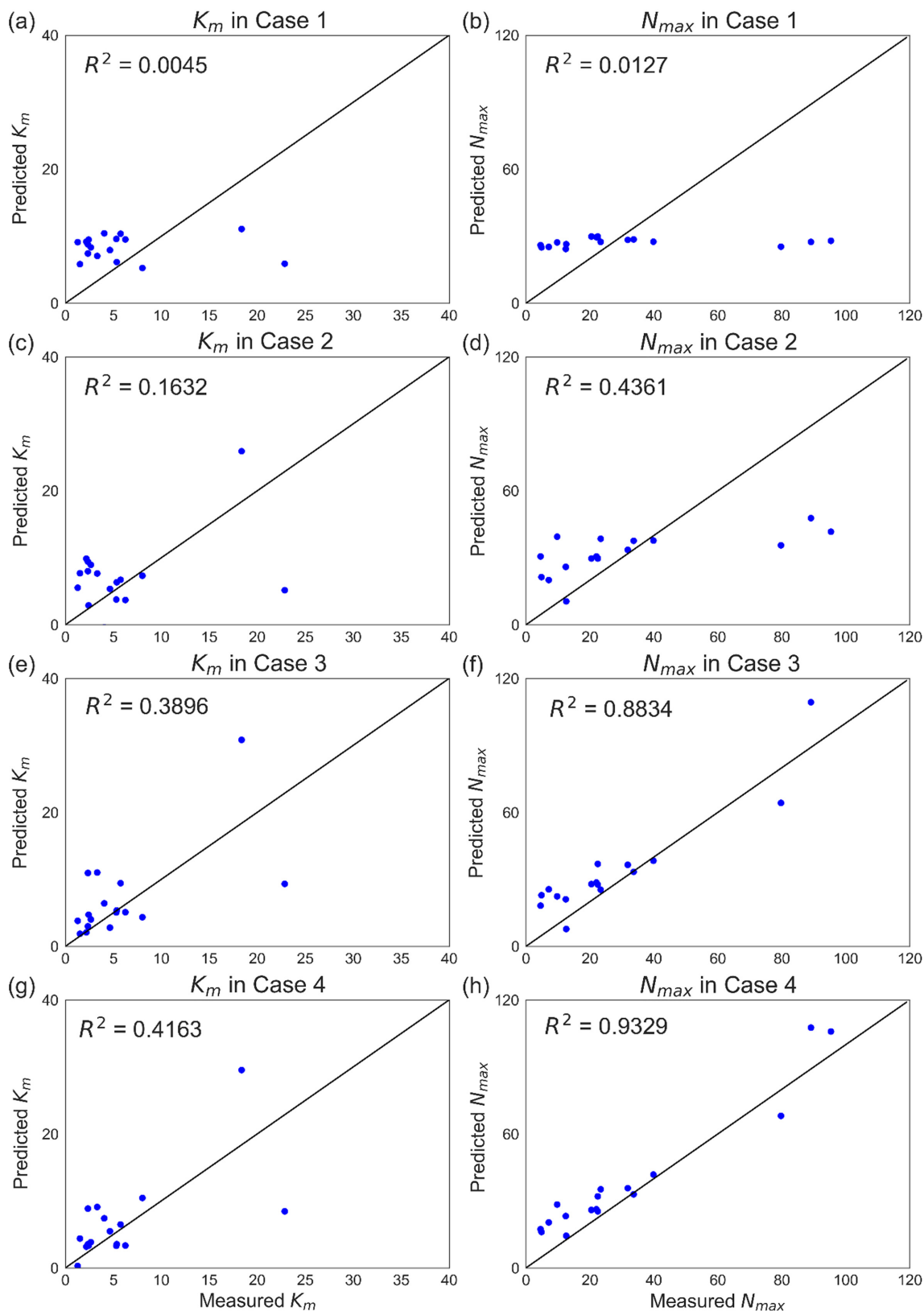


Fig. 7. Parity plot between measured and predicted parameters (K_m and N_{max}) obtained from four cases.

함수와 한개의 tanh 함수를 사용했고, 뉴런수는 각각 32, 6, 28이다. 2개의 출력변수를 계산하는 출력층에서는 활성화 함수를 사용하지 않았다. Cases 3-4 모두 최적화 해석기는 Rprop [34]이 사용되었다.

Case 3은 OHE만을 사용한 Case 2 보다 MAE가 K_m 과 N_{max} 에서 각각 21.1%, 37.3%가 감소하였다. 증식 데이터를 사용한 Case 4는 Case 3 대비 MAE가 K_m 과 N_{max} 에서 각각 3.9%, 18.7%가 감소하였다.

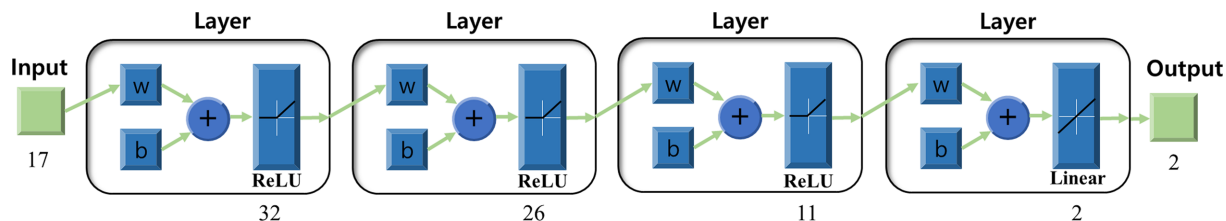


Fig. 8. Architecture of Case 3 with optimized hyper-parameters using GP-EI.

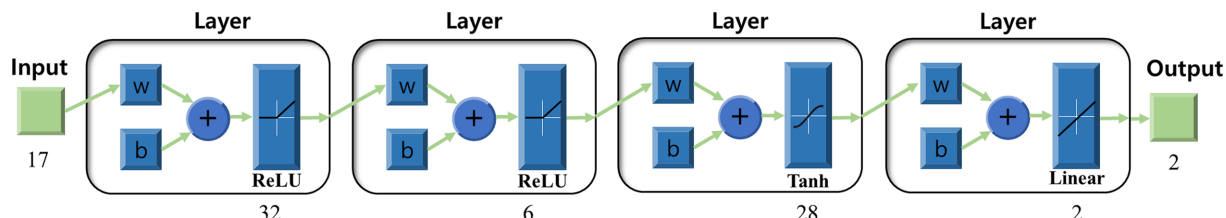


Fig. 9. Architecture of Case 4 with optimized hyper-parameters using GP-EI.

5. Conclusion

인공신경망 기법에서 데이터 수 부족, 범주형 데이터의 처리, 초매개변수 최적화 등의 문제를 해결하기 위하여, 본 연구에서는 범주형 데이터를 다차원 등간격 데이터로 변환하는 one-hot encoding (OHE), generative adversarial network (GAN)를 이용한 데이터 증식 및 Gaussian process and expected improvement (GP-EI)를 이용한 초매개변수의 최적화를 수행하였다. 암모니아 방출량을 예측하고자 하는 4개 서로 다른 artificial neural network (ANN) 구조를 비교하였고, OHE 및 GAN을 이용한 데이터 전처리와 GP-EI를 이용하여 초매개변수를 최적화한 Case 4는 평가 데이터에 대하여 우수한 성능 지표를 보여주었다.

OHE를 적용한 데이터는 0과 1로만 표기되어야 하지만, GAN을 사용하여 증식된 데이터는 0과 1이 아닌 소수값을 보여주었다. 추후 연구를 통하여 증식된 데이터도 OHE의 규칙에 따르도록 해야 할 것이다. 또한, 평가 데이터를 무작위로 추출하기 보다는 균일하게 추출하는 기법이 요구된다. 본 연구에서 제시된 전처리 방법과 초매개변수 최적화 방법은 적은 양의 데이터를 갖는 공학 문제에서 인공신경망 성능을 향상하기 위하여 활용할 수 있을 것이다.

Acknowledgement

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (Grant number: NRF-2021R1A2C1011618).

References

- Lim, Y., Moon, Y.-S. and Kim, T.-W., "Artificial Neural Network Approach for Prediction of Ammonia Emission from Field-Applied Manure and Relative Significance Assessment of Ammonia Emission Factors," *Eur. J. Agron.*, **26**(4), 425-434(2007).
- Sintermann, J., Neftel, A., Ammann, C., Häni, C., Hensen, A., Loubet, B. and Flechard, C. R., "Are Ammonia Emissions from Field-Applied Slurry Substantially over-Estimated in European Emission Inventories?," *Biogeosciences*, **9**(5), 1611-1632(2012).
- Pedersen, J., Andersson, K., Feilberg, A., Delin, S., Hafner, S. and Nyord, T., "Effect of Exposed Surface Area on Ammonia Emissions from Untreated, Separated, and Digested Cattle Manure," *Biosyst. Eng.*, **202**, 66-78(2021).
- Moon, Y. S., Lim, Y. and Kim, T. W., "Prediction of Ammonia Emission Rate from Field-Applied Animal Manure Using the Artificial Neural Network," *Korean Chem. Eng. Res.*, **45**(2), 133-142(2007).
- Jain, S., Shukla, S. and Wadhvani, R., "Dynamic Selection of Normalization Techniques Using Data Complexity Measures," *Expert Syst. Appl.*, **106**, 252-262(2018).
- Jackson, J. E., *A User's Guide to Principal Components*, 1st ed., John Wiley and Sons, New York(1991).
- Ren, J., Chen, J., Shi, D., Li, Y., Li, D., Wang, Y. and Cai, D., "Online Multi-Fault Power System Dynamic Security Assessment Driven by Hybrid Information of Anticipated Faults and Pre-Fault Power Flow," *Int. J. Electr. Power Energy Syst.*, **136**, 107651(2022).
- Hovden, I. T., "Optimizing Artificial Neural Network Hyperparameters and Architecture," (2019).
- Islam, A., Belhaouari, S. B., Rehman, A. U. and Bensmail, H., "Knnor: An Oversampling Technique for Imbalanced Datasets," *Appl. Soft Comput.*, **115**, 108288(2022).
- Ngo, S. I. and Lim, Y.-I., "Solution and Parameter Identification of a Fixed-Bed Reactor Model for Catalytic CO₂ Methanation Using Physics-Informed Neural Networks," *Catalysts*, **11**(11), 1304(2021).
- Raissi, M., Perdikaris, P. and Karniadakis, G. E., "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations," *J. Comput. Phys.*, **378**, 686-707(2019).
- Yang, Y., Zha, K., Chen, Y.-C., Wang, H. and Katabi, D., "Delving into Deep Imbalanced Regression," *arXiv preprint arXiv:2102.09554*, (2021).
- Zhuo, Y. and Brgoch, J., "Opportunities for Next-Generation Luminescent Materials through Artificial Intelligence," *J. Phys. Chem. Lett.*, **12**(2), 764-772(2021).

14. Coomans, D. and Massart, D. L., "Alternative K-Nearest Neighbour Rules in Supervised Pattern Recognition: Part 1. K-Nearest Neighbour Classification by Using Alternative Voting Rules," *Anal. Chim. Acta*, **136**, 15-27(1982).
15. Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P., "Smote: Synthetic Minority over-Sampling Technique," *J. Artif. Intell. Res.*, **16**, 321-357(2002).
16. Torgo, L., Ribeiro, R., Pfahringer, B. and Branco, P., *Smote for Regression*, ed., Springer, Berlin, Heidelberg, Correia(2013).
17. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., "Generative Adversarial Nets," *Adv. Neural Inf. Process. Syst.*, **27**, (2014).
18. Asadi, M. and McPhedran, K. N., "Greenhouse Gas Emission Estimation from Municipal Wastewater Using a Hybrid Approach of Generative Adversarial Network and Data-Driven Modelling," *Sci. Total Environ.*, **800**, 149508(2021).
19. Du, X., Xu, H. and Zhu, F., "Understanding the Effect of Hyperparameter Optimization on Machine Learning Models for Structure Design Problems," *Comput. Aided Des.*, **135**, 103013(2021).
20. Bergstra, J., Bardenet, R., Bengio, Y. and Kégl, B., "Algorithms for Hyper-Parameter Optimization," *Adv. Neural Inf. Process. Syst.*, (2011).
21. Yang, Z. and Zhang, A., "Hyperparameter Optimization Via Sequential Uniform Designs," *J. Mach. Learn. Res.*, **22**(149), 1-47(2021).
22. Yang, L. and Shami, A., "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice," *Neurocomputing*, **415**, 295-316(2020).
23. Jankovic, A., Chaudhary, G. and Goia, F., "Designing the Design of Experiments (Doe) – an Investigation on the Influence of Different Factorial Designs on the Characterization of Complex Systems," *Energy Build.*, **250**, 111298(2021).
24. Misselbrook, T. H., Nicholson, F. A. and Chambers, B. J., "Predicting Ammonia Losses Following the Application of Livestock Manure to Land," *Bioresour. Technol.*, **96**(2), 159-168(2005).
25. Rodríguez, P., Bautista, M. A., González, J. and Escalera, S., "Beyond One-Hot Encoding: Lower Dimensional Target Embedding," *Image Vis. Comput.*, **75**, 21-31(2018).
26. Demir, S., Mincev, K., Kok, K. and Paterakis, N. G., "Data Augmentation for Time Series Regression: Applying Transformations, Autoencoders and Adversarial Networks to Electricity Price Forecasting," *Appl. Energ.*, **304**, 117695(2021).
27. Ramachandran, P., Zoph, B. and Le, Q. V., "Searching for Activation Functions," *arXiv preprint arXiv:1710.05941*, (2017).
28. LeCun, Y., Bengio, Y. and Hinton, G., "Deep Learning," *Nature*, **521**(7553), 436-444(2015).
29. Kohavi, R., "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, Morgan Kaufmann Publishers Inc., Montreal, Quebec, Canada, pp. 1137-1143(1995).
30. Li, W. and Liu, Z., "A Method of Svm with Normalization in Intrusion Detection," *Procedia Environ. Sci.*, **11**, 256-262(2011).
31. Jahirul, M. I., Rasul, M. G., Brown, R. J., Senadeera, W., Hosen, M. A., Haque, R., Saha, S. C. and Mahlia, T. M. I., "Investigation of Correlation between Chemical Composition and Properties of Biodiesel Using Principal Component Analysis (Pca) and Artificial Neural Network (Ann)," *Renew. Energy*, **168**, 632-646 (2021).
32. Liu, Y., Zhao, S., Wang, Q. and Gao, Q., "Learning More Distinctive Representation by Enhanced Pca Network," *Neurocomputing*, **275**, 924-931(2018).
33. Berk, J., Nguyen, V., Gupta, S., Rana, S. and Venkatesh, S., "Exploration Enhanced Expected Improvement for Bayesian Optimization," *Machine Learning and Knowledge Discovery in Databases*, September, Cham(2018).
34. Riedmiller, M. and Braun, H., "A Direct Adaptive Method for Faster Backpropagation Learning: The Rprop Algorithm," *IEEE International Conference on Neural Networks*, March, 1993.

Authors

Pyeong-Gon Jung: Undergraduate course, Department of Chemical Engineering, Hankyong National University, Anseong 17579, Korea; zkjxc975@naver.com

Young-il Lim: Professor, Department of Chemical Engineering, Hankyong National University, Anseong 17579, Korea; limyi@hknu.ac.kr

Appendix

A1. Original data for ammonia emission rate

Table A1. Michaelis-Menten equation parameters derived from experimental data measured at 11 input variables for field-applied manure (Lim et al., 2007) [1]

| No. | K_m | N_{max} | P_1 | P_2 | P_3 | P_4 | P_5 | P_6 | P_7 | P_8 | P_9 | P_{10} | P_{11} |
|-----|-------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|
| 1 | 14.05 | 21.12 | 1 | 6.60 | 12.1 | 3.1 | 3.9 | 3.9 | 7.9 | 0 | 30 | 3 | 1 |
| 2 | 10.30 | 24.66 | 1 | 6.60 | 12.1 | 3.4 | 3.9 | 3.9 | 7.9 | 0 | 30 | 3 | 1 |
| 3 | 13.70 | 38.30 | 1 | 6.60 | 10.8 | 3 | 3.9 | 3.9 | 7.9 | 0 | 30 | 3 | 1 |
| 4 | 17.60 | 15.87 | 1 | 6.60 | 10.8 | 3.1 | 3.9 | 3.9 | 7.9 | 0 | 30 | 3 | 1 |
| 5 | 85.40 | 35.75 | 1 | 6.60 | 10.8 | 3.2 | 3.9 | 3.9 | 7.9 | 0 | 30 | 3 | 1 |
| 6 | 38.74 | 21.83 | 1 | 6.60 | 6.1 | 2.8 | 3.9 | 2 | 7.9 | 0 | 30 | 3 | 1 |
| 7 | 44.45 | 14.41 | 1 | 6.60 | 6.1 | 3.2 | 3.9 | 2 | 7.9 | 0 | 30 | 3 | 1 |
| 8 | 13.54 | 28.67 | 1 | 6.60 | 6.1 | 3 | 3.9 | 2 | 7.9 | 0 | 30 | 3 | 1 |
| 9 | 14.64 | 5.99 | 1 | 6.60 | 5.1 | 3 | 3.9 | 2 | 7.9 | 0 | 30 | 3 | 1 |
| 10 | 22.89 | 26.34 | 1 | 6.60 | 3.1 | 2.9 | 3.9 | 2 | 7.9 | 0 | 30 | 3 | 1 |
| 11 | 18.36 | 12.41 | 1 | 6.60 | 3.1 | 3.2 | 3.9 | 2 | 7.9 | 0 | 30 | 3 | 1 |
| 12 | 34.61 | 17.03 | 1 | 6.60 | 3.1 | 3 | 3.9 | 2 | 7.9 | 0 | 30 | 3 | 1 |
| 13 | 1.07 | 9.06 | 3 | 6.30 | 8.41 | 5.78 | 3.70 | 3.84 | 7.35 | 0 | 31.8 | 4 | 2 |
| 14 | 4.62 | 9.11 | 3 | 6.30 | 8.44 | 5.74 | 3.30 | 3.81 | 7.48 | 1 | 26.0 | 4 | 2 |
| 15 | 0.63 | 8.93 | 3 | 6.30 | 14 | 4.19 | 3.50 | 3.93 | 7.66 | 0 | 30.9 | 4 | 2 |
| 16 | 4.55 | 4.01 | 3 | 6.30 | 14.06 | 4.22 | 3.50 | 3.57 | 7.64 | 1 | 26.5 | 4 | 2 |
| 17 | 3.12 | 14.37 | 3 | 6.30 | 11.4 | 6.31 | 3.10 | 3.54 | 7.53 | 0 | 30.9 | 4 | 2 |
| 18 | 2.06 | 2.72 | 3 | 6.30 | 11.4 | 6.31 | 3.00 | 3.33 | 7.53 | 1 | 25.5 | 4 | 2 |
| 19 | 2.15 | 7.92 | 3 | 6.30 | 8.86 | 5.17 | 3.40 | 3.06 | 7.55 | 0 | 24.3 | 4 | 2 |
| 20 | 3.39 | 10.80 | 3 | 6.30 | 8.86 | 5.17 | 3.20 | 3.09 | 7.56 | 1 | 19.4 | 4 | 2 |
| 21 | 0.86 | 8.02 | 3 | 6.30 | 10.81 | 4.06 | 3.40 | 3.17 | 7.76 | 0 | 27.5 | 4 | 2 |
| 22 | 5.98 | 6.93 | 3 | 6.30 | 10.81 | 4.06 | 2.80 | 2.70 | 7.71 | 1 | 21.6 | 4 | 2 |
| 23 | 1.18 | 10.18 | 3 | 6.30 | 10.57 | 6.34 | 3.40 | 3.16 | 7.61 | 0 | 24.9 | 4 | 2 |
| 24 | 3.46 | 6.76 | 3 | 6.30 | 10.65 | 6.34 | 2.70 | 2.60 | 7.58 | 1 | 24.2 | 4 | 2 |
| 25 | 1.40 | 6.87 | 3 | 6.30 | 12.85 | 4.31 | 5.20 | 1.49 | 6.70 | 0 | 38.3 | 4 | 2 |
| 26 | 4.93 | 5.39 | 3 | 6.30 | 12.75 | 4.3 | 5.20 | 1.49 | 6.70 | 1 | 31.3 | 4 | 2 |
| 27 | 2.49 | 10.52 | 3 | 6.30 | 5.87 | 5.52 | 6.90 | 3.65 | 7.01 | 0 | 34.4 | 4 | 2 |
| 28 | 4.65 | 12.54 | 3 | 6.30 | 5.87 | 5.52 | 6.90 | 3.65 | 7.01 | 1 | 22.9 | 4 | 2 |
| 29 | 2.97 | 17.98 | 3 | 6.30 | 10.93 | 4.91 | 4.20 | 3.44 | 6.97 | 0 | 31.9 | 4 | 2 |
| 30 | 3.77 | 9.03 | 3 | 6.30 | 10.93 | 4.91 | 4.20 | 3.44 | 6.97 | 1 | 31.3 | 4 | 2 |
| 31 | 34.67 | 15.75 | 2 | 7.80 | 0.85 | 3.15 | 1.93 | 3.00 | 7.52 | 1 | 28.7 | 4 | 2 |
| 32 | 0.64 | 0.65 | 2 | 7.65 | 5.5 | 4.37 | 2.79 | 2.80 | 7.66 | 1 | 29.9 | 3 | 2 |
| 33 | 9.23 | 13.95 | 2 | 7.49 | 9.62 | 3.58 | 3.56 | 3.2 | 7.63 | 1 | 30.2 | 4 | 2 |
| 34 | 16.84 | 41.90 | 2 | 7.46 | 11.65 | 3.85 | 3.54 | 2.8 | 7.66 | 1 | 30.2 | 4 | 2 |
| 35 | 12.79 | 26.42 | 2 | 6.91 | 15.05 | 2.82 | 3.76 | 2.6 | 7.42 | 1 | 30.9 | 4 | 2 |
| 36 | 2.33 | 7.12 | 2 | 7.47 | 15.12 | 2.85 | 4.04 | 2.69 | 7.51 | 1 | 31.2 | 4 | 2 |
| 37 | 3.31 | 4.81 | 2 | 7.69 | 17.03 | 2.08 | 3.7 | 2.4 | 7.52 | 1 | 31.1 | 4 | 2 |
| 38 | 6.05 | 5.02 | 2 | 7.74 | 17.15 | 1.88 | 3.44 | 2.45 | 7.68 | 1 | 31.3 | 4 | 2 |
| 39 | 7.64 | 9.85 | 2 | 8.00 | 15.7 | 1.8 | 11.30 | 6.31 | 8.00 | 2 | 7.9 | 1 | 2 |
| 40 | 3.92 | 26.46 | 2 | 8.00 | 15.89 | 1.75 | 11.30 | 6.31 | 8.00 | 2 | 14.9 | 1 | 2 |
| 41 | 1.29 | 79.66 | 2 | 8.00 | 12.93 | 1.8 | 11.30 | 6.31 | 8.00 | 0 | 17.5 | 1 | 2 |
| 42 | 5.60 | 13.17 | 2 | 8.00 | 15.71 | 1.8 | 11.30 | 6.31 | 8.00 | 3 | 17.3 | 1 | 2 |
| 43 | 0.74 | 21.00 | 2 | 7.80 | 24 | 1.88 | 5.13 | 3.51 | 7.8 | 0 | 8.4 | 1 | 2 |
| 44 | 5.34 | 9.87 | 2 | 7.30 | 7.29 | 5.11 | 8.66 | 5.00 | 7.30 | 2 | 12.0 | 1 | 2 |
| 45 | 4.77 | 84.11 | 2 | 7.30 | 7.29 | 4.98 | 8.66 | 5.00 | 7.30 | 0 | 16.3 | 1 | 2 |

Table A1. Michaelis-Menten equation parameters derived from experimental data measured at 11 input variables for field-applied manure (Lim et al., 2007) [1]

| No. | K_m | N_{max} | P_1 | P_2 | P_3 | P_4 | P_5 | P_6 | P_7 | P_8 | P_9 | P_{10} | P_{11} |
|-------|-------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|----------|
| 46 | 3.07 | 88.30 | 2 | 7.30 | 7.3 | 5.14 | 8.66 | 5.03 | 7.3 | 0 | 15.2 | 1 | 2 |
| 47 | 4.05 | 4.54 | 2 | 7.30 | 7.34 | 5.19 | 8.66 | 5.03 | 7.3 | 2 | 10.6 | 1 | 2 |
| 48 | 22.31 | 32.65 | 2 | 5.14 | 8.31 | 0.93 | 3.07 | 4.02 | 7.50 | 0 | 40.0 | 1 | 1 |
| 49 | 5.75 | 31.79 | 2 | 5.14 | 8.36 | 1.15 | 3.07 | 4.02 | 7.50 | 0 | 40.0 | 1 | 1 |
| 50 | 5.36 | 41.71 | 2 | 5.14 | 8.36 | 1.04 | 3.07 | 4.02 | 7.50 | 0 | 40.0 | 1 | 1 |
| 51 | 2.66 | 32.04 | 2 | 5.88 | 8.59 | 1.05 | 4.01 | 4.14 | 7.44 | 0 | 40.0 | 1 | 1 |
| 52 | 8.13 | 30.23 | 2 | 5.88 | 8.59 | 1.15 | 4.01 | 4.14 | 7.44 | 0 | 40.0 | 1 | 1 |
| 53 | 2.46 | 45.66 | 2 | 5.88 | 8.59 | 1.06 | 4.01 | 4.14 | 7.44 | 0 | 40.0 | 1 | 1 |
| 54 | 1.86 | 27.37 | 3 | 5.71 | 18.3 | 0.95 | 1.52 | 2.00 | 7.47 | 0 | 40.0 | 1 | 1 |
| 55 | 1.99 | 34.97 | 3 | 5.71 | 18.3 | 0.87 | 1.52 | 2.00 | 7.47 | 0 | 40.0 | 1 | 1 |
| 56 | 1.22 | 29.02 | 3 | 5.71 | 18.3 | 1.13 | 1.52 | 2.00 | 7.47 | 0 | 40.0 | 1 | 1 |
| 57 | 3.89 | 17.96 | 3 | 5.75 | 5.05 | 1.06 | 1.72 | 3.71 | 7.38 | 0 | 40.0 | 1 | 1 |
| 58 | 6.25 | 22.39 | 3 | 5.75 | 5.05 | 1.05 | 1.72 | 3.71 | 7.38 | 0 | 40.0 | 1 | 1 |
| 59 | 5.30 | 20.45 | 3 | 5.75 | 5.05 | 1.17 | 1.72 | 3.71 | 7.38 | 0 | 40.0 | 1 | 1 |
| 60 | 3.27 | 24.68 | 3 | 5.46 | 4.9 | 1.04 | 2.06 | 3.05 | 7.45 | 0 | 40.0 | 1 | 1 |
| 61 | 2.41 | 21.96 | 3 | 5.46 | 4.9 | 1.24 | 2.06 | 3.05 | 7.45 | 0 | 40.0 | 1 | 1 |
| 62 | 6.34 | 28.44 | 3 | 5.46 | 4.9 | 1.14 | 2.06 | 3.05 | 7.45 | 0 | 40.0 | 1 | 1 |
| 63 | 3.94 | 9.09 | 2 | 7.90 | 21.58 | 0.50 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 64 | 1.71 | 11.85 | 2 | 7.90 | 21.58 | 1.90 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 65 | 2.13 | 37.45 | 2 | 7.90 | 21.58 | 2.92 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 66 | 2.65 | 9.70 | 2 | 7.90 | 21.58 | 0.89 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 67 | 2.33 | 23.32 | 2 | 7.90 | 21.58 | 1.95 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 68 | 2.17 | 39.82 | 2 | 7.90 | 21.58 | 2.93 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 69 | 3.15 | 10.00 | 2 | 7.90 | 21.58 | 0.49 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 70 | 3.41 | 50.16 | 2 | 7.90 | 21.58 | 4.07 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 71 | 3.21 | 22.70 | 2 | 7.90 | 21.58 | 0.95 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 72 | 2.78 | 75.87 | 2 | 7.90 | 21.58 | 4.01 | 4.74 | 3.19 | 7.17 | 0 | 40 | 2 | 1 |
| 73 | 5.82 | 5.77 | 3 | 7.50 | 17.02 | 0.82 | 10.50 | 3.75 | 7.77 | 3 | 40.0 | 2 | 1 |
| 74 | 5.35 | 33.62 | 3 | 7.50 | 17.02 | 0.73 | 10.50 | 3.75 | 7.77 | 0 | 40.0 | 2 | 1 |
| 75 | 5.03 | 22.59 | 3 | 7.50 | 17.02 | 0.69 | 10.50 | 3.75 | 7.77 | 0 | 40.0 | 2 | 1 |
| 76 | 1.92 | 108.42 | 3 | 7.50 | 27.31 | 1.38 | 4.36 | 2.58 | 7.85 | 0 | 58.0 | 1 | 3 |
| 77 | 1.51 | 89.03 | 3 | 7.50 | 27.31 | 1.38 | 4.03 | 2.56 | 7.99 | 0 | 58.0 | 1 | 3 |
| 78 | 8.02 | 95.34 | 3 | 7.50 | 27.31 | 1.38 | 3.96 | 2.50 | 7.82 | 1 | 58.0 | 1 | 3 |
| 79 | 13.55 | 47.31 | 3 | 7.60 | 17.81 | 0.76 | 11.00 | 3.70 | 7.77 | 1 | 35.0 | 2 | 1 |
| 80 | 12.49 | 47.78 | 3 | 7.60 | 17.81 | 0.85 | 11.00 | 3.70 | 7.77 | 1 | 35.0 | 2 | 1 |
| 81 | 3.79 | 35.79 | 3 | 7.60 | 17.81 | 0.82 | 11.00 | 3.70 | 7.77 | 1 | 35.0 | 2 | 1 |
| 82 | 1.90 | 33.74 | 3 | 7.60 | 22.39 | 0.77 | 6.17 | 2.49 | 7.25 | 0 | 45.0 | 2 | 1 |
| 83 | 22.85 | 22.42 | 3 | 7.60 | 22.39 | 0.66 | 6.17 | 2.49 | 7.25 | 1 | 45.0 | 2 | 1 |
| Min. | 0.63 | 0.65 | 1 | 5.14 | 0.85 | 0.49 | 1.52 | 1.49 | 6.7 | 0 | 7.9 | 1 | 1 |
| Max. | 85.40 | 108.42 | 3 | 8 | 27.31 | 6.34 | 11.3 | 6.31 | 8 | 3 | 58 | 4 | 3 |
| Aver. | 8.35 | 25.89 | 2.31 | 6.84 | 12.83 | 2.78 | 4.87 | 3.36 | 7.53 | 0.43 | 32.73 | 2.43 | 1.49 |
| s.d. | 12.26 | 22.88 | 0.71 | 0.85 | 6.53 | 1.76 | 2.73 | 1.03 | 0.31 | 0.70 | 9.97 | 1.23 | 0.57 |

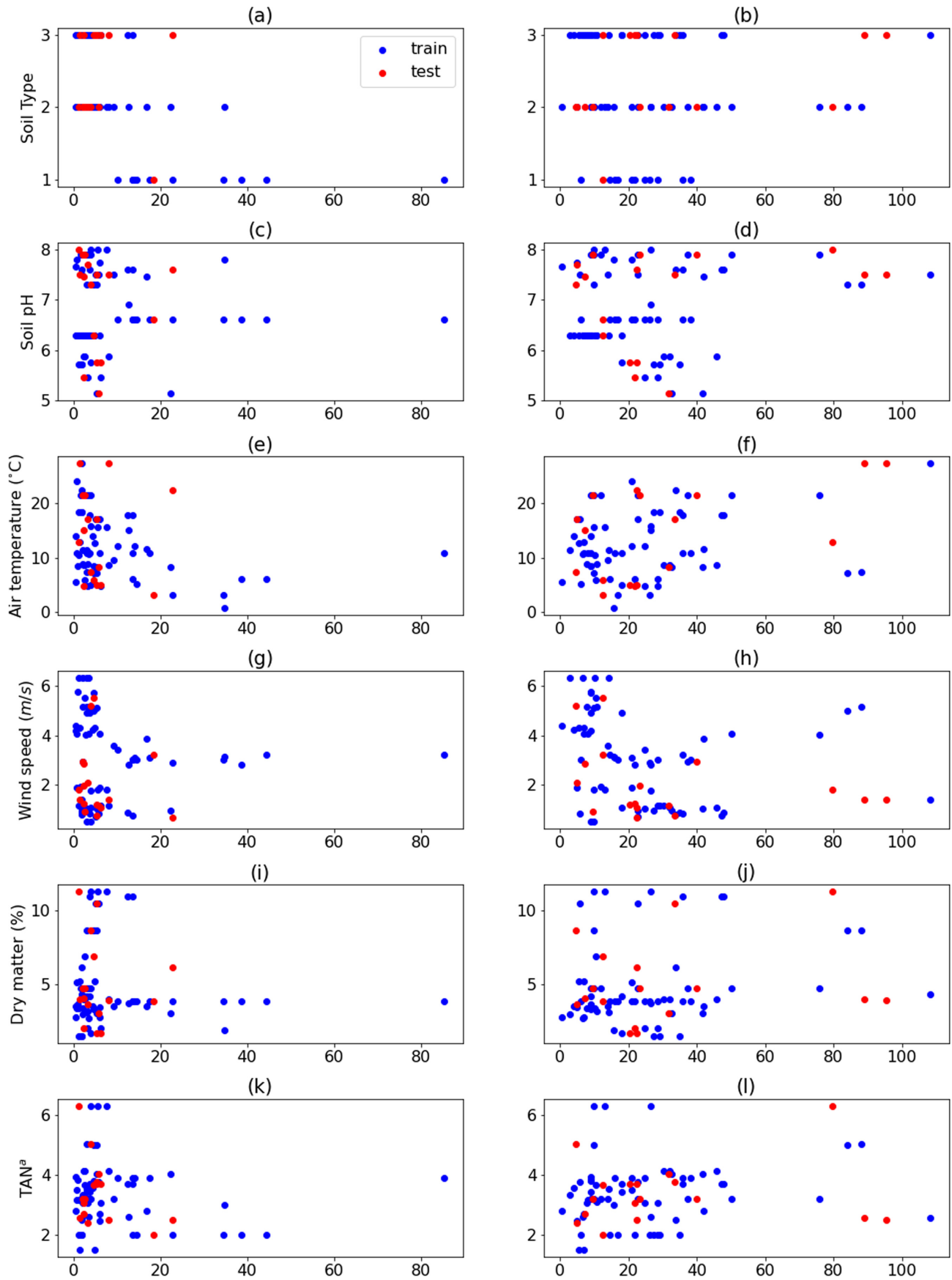


Fig. A1. Distribution of original data (blue: training data; red: test data).

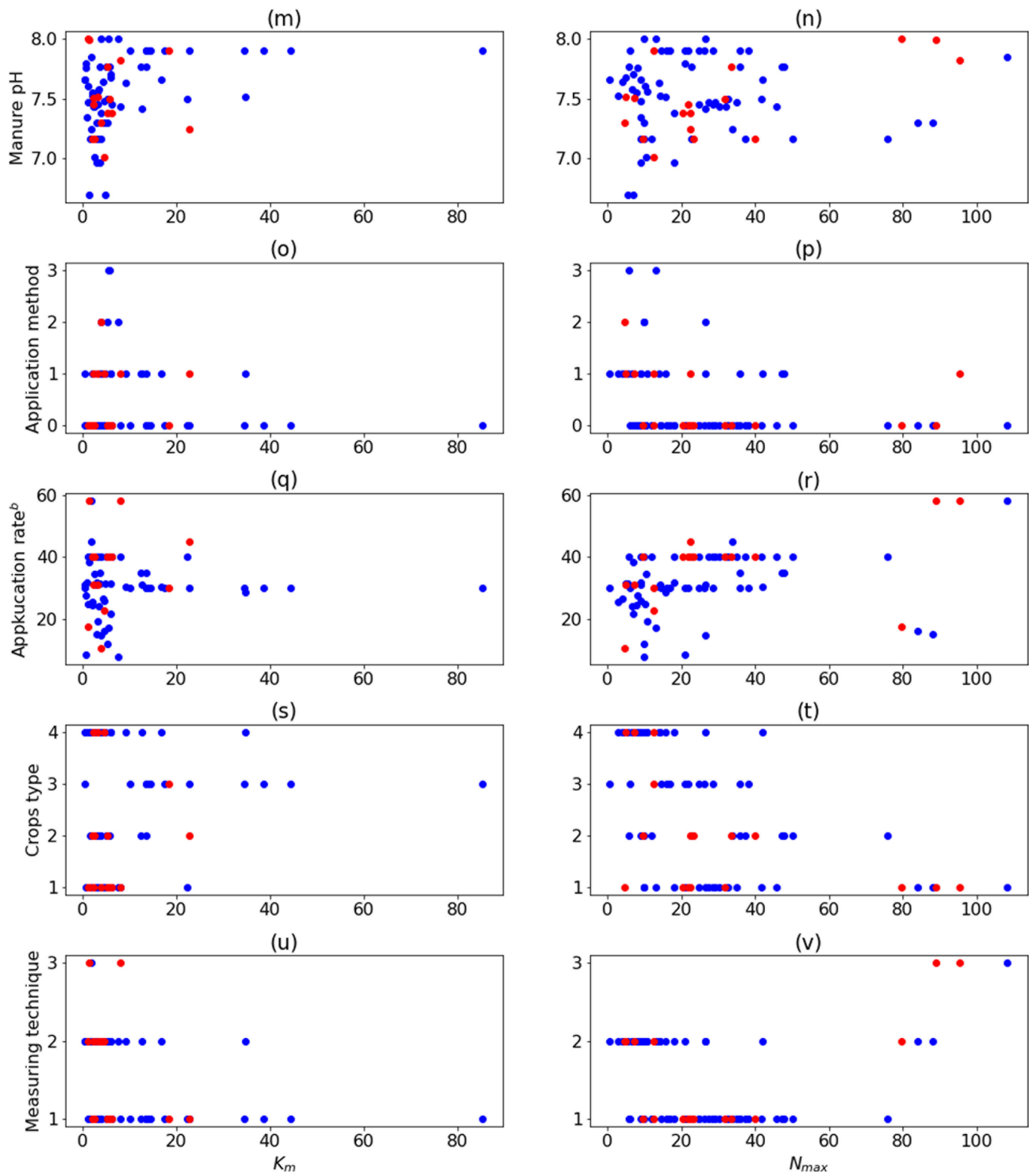


Fig. A1. (Continued)

A2. Augmented data generated by GAN

Table A2. Augmented data sets (96 training data = 66 original training data + 13 augmented training data + 17 test data)

| No. | K_m | N_{max} | p_2 | p_3 | p_4 | p_5 | p_6 | p_7 | p_9 | OHE p_1 | | OHE p_8 | | | OHE p_{10} | | | OHE p_{11} | |
|-----|-------|-----------|-------|-------|-------|-------|-------|-------|-------|-----------|----------|-----------|----------|----------|--------------|----------|----------|--------------|----------|
| | | | | | | | | | | p_1 | p_{12} | p_8 | p_{13} | p_{14} | p_{10} | p_{15} | p_{16} | p_{11} | p_{17} |
| 1 | 14.05 | 21.12 | 6.60 | 12.10 | 3.10 | 3.90 | 3.90 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 2 | 10.30 | 24.66 | 6.60 | 12.10 | 3.40 | 3.90 | 3.90 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 3 | 13.70 | 38.30 | 6.60 | 10.80 | 3.00 | 3.90 | 3.90 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 4 | 17.60 | 15.87 | 6.60 | 10.80 | 3.10 | 3.90 | 3.90 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 5 | 85.40 | 35.75 | 6.60 | 10.80 | 3.20 | 3.90 | 3.90 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 6 | 38.74 | 21.83 | 6.60 | 6.10 | 2.80 | 3.90 | 2.00 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 7 | 44.45 | 14.41 | 6.60 | 6.10 | 3.20 | 3.90 | 2.00 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 8 | 13.54 | 28.67 | 6.60 | 6.10 | 3.00 | 3.90 | 2.00 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 9 | 14.64 | 5.99 | 6.60 | 5.10 | 3.00 | 3.90 | 2.00 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 10 | 22.89 | 26.34 | 6.60 | 3.10 | 2.90 | 3.90 | 2.00 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 11 | 18.36 | 12.41 | 6.60 | 3.10 | 3.20 | 3.90 | 2.00 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 12 | 34.61 | 17.03 | 6.60 | 3.10 | 3.00 | 3.90 | 2.00 | 7.90 | 30.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| 13 | 1.07 | 9.06 | 6.30 | 8.41 | 5.78 | 3.70 | 3.84 | 7.35 | 31.80 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 14 | 4.62 | 9.11 | 6.30 | 8.44 | 5.74 | 3.30 | 3.81 | 7.48 | 26.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 15 | 0.63 | 8.93 | 6.30 | 14.00 | 4.19 | 3.50 | 3.93 | 7.66 | 30.90 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 16 | 4.55 | 4.01 | 6.30 | 14.06 | 4.22 | 3.50 | 3.57 | 7.64 | 26.50 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 17 | 3.12 | 14.37 | 6.30 | 11.40 | 6.31 | 3.10 | 3.54 | 7.53 | 30.90 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 18 | 2.06 | 2.72 | 6.30 | 11.40 | 6.31 | 3.00 | 3.33 | 7.53 | 25.50 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 19 | 2.15 | 7.92 | 6.30 | 8.86 | 5.17 | 3.40 | 3.06 | 7.55 | 24.30 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 20 | 3.39 | 10.80 | 6.30 | 8.86 | 5.17 | 3.20 | 3.09 | 7.56 | 19.40 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 21 | 0.86 | 8.02 | 6.30 | 10.81 | 4.06 | 3.40 | 3.17 | 7.76 | 27.50 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 22 | 5.98 | 6.93 | 6.30 | 10.81 | 4.06 | 2.80 | 2.70 | 7.71 | 21.60 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 23 | 1.18 | 10.18 | 6.30 | 10.57 | 6.34 | 3.40 | 3.16 | 7.61 | 24.90 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 24 | 3.46 | 6.76 | 6.30 | 10.65 | 6.34 | 2.70 | 2.60 | 7.58 | 24.20 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 25 | 1.40 | 6.87 | 6.30 | 12.85 | 4.31 | 5.20 | 1.49 | 6.70 | 38.30 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 26 | 4.93 | 5.39 | 6.30 | 12.75 | 4.30 | 5.20 | 1.49 | 6.70 | 31.30 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 27 | 2.49 | 10.52 | 6.30 | 5.87 | 5.52 | 6.90 | 3.65 | 7.01 | 34.40 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 28 | 4.65 | 12.54 | 6.30 | 5.87 | 5.52 | 6.90 | 3.65 | 7.01 | 22.90 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 29 | 2.97 | 17.98 | 6.30 | 10.93 | 4.91 | 4.20 | 3.44 | 6.97 | 31.90 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 30 | 3.77 | 9.03 | 6.30 | 10.93 | 4.91 | 4.20 | 3.44 | 6.97 | 31.30 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 31 | 34.67 | 15.75 | 7.80 | 0.85 | 3.15 | 1.93 | 3.00 | 7.52 | 28.70 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 32 | 0.64 | 0.65 | 7.65 | 5.50 | 4.37 | 2.79 | 2.80 | 7.66 | 29.90 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 1.00 | 0.00 |
| 33 | 9.23 | 13.95 | 7.49 | 9.62 | 3.58 | 3.56 | 3.20 | 7.63 | 30.20 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 34 | 16.84 | 41.90 | 7.46 | 11.65 | 3.85 | 3.54 | 2.80 | 7.66 | 30.20 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 35 | 12.79 | 26.42 | 6.91 | 15.05 | 2.82 | 3.76 | 2.60 | 7.42 | 30.90 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 36 | 2.33 | 7.12 | 7.47 | 15.12 | 2.85 | 4.04 | 2.69 | 7.51 | 31.20 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 37 | 3.31 | 4.81 | 7.69 | 17.03 | 2.08 | 3.70 | 2.40 | 7.52 | 31.10 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 38 | 6.05 | 5.02 | 7.74 | 17.15 | 1.88 | 3.44 | 2.45 | 7.68 | 31.30 | 1.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 |
| 39 | 7.64 | 9.85 | 8.00 | 15.70 | 1.80 | 11.30 | 6.31 | 8.00 | 7.90 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 40 | 3.92 | 26.46 | 8.00 | 15.89 | 1.75 | 11.30 | 6.31 | 8.00 | 14.90 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 41 | 1.29 | 79.66 | 8.00 | 12.93 | 1.80 | 11.30 | 6.31 | 8.00 | 17.50 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 42 | 5.60 | 13.17 | 8.00 | 15.71 | 1.80 | 11.30 | 6.31 | 8.00 | 17.30 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 43 | 0.74 | 21.00 | 7.80 | 24.00 | 1.88 | 5.13 | 3.51 | 7.80 | 8.40 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 44 | 5.34 | 9.87 | 7.30 | 7.29 | 5.11 | 8.66 | 5.00 | 7.30 | 12.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 45 | 4.77 | 84.11 | 7.30 | 7.29 | 4.98 | 8.66 | 5.00 | 7.30 | 16.30 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 46 | 3.07 | 88.30 | 7.30 | 7.30 | 5.14 | 8.66 | 5.03 | 7.30 | 15.20 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 47 | 4.05 | 4.54 | 7.30 | 7.34 | 5.19 | 8.66 | 5.03 | 7.30 | 10.60 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 |
| 48 | 22.31 | 32.65 | 5.14 | 8.31 | 0.93 | 3.07 | 4.02 | 7.50 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 49 | 5.75 | 31.79 | 5.14 | 8.36 | 1.15 | 3.07 | 4.02 | 7.50 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Table A2. (Continued)

| No. | K_m | N_{max} | p_2 | p_3 | p_4 | p_5 | p_6 | p_7 | p_9 | OHE p_1 | | OHE p_8 | | | OHE p_{10} | | | OHE p_{11} | |
|-------|-------|-----------|-------|-------|-------|-------|-------|-------|-------|-----------|----------|-----------|----------|----------|--------------|----------|----------|--------------|----------|
| | | | | | | | | | | p_1 | p_{12} | p_8 | p_{13} | p_{14} | p_{10} | p_{15} | p_{16} | p_{11} | p_{17} |
| 50 | 5.36 | 41.71 | 5.14 | 8.36 | 1.04 | 3.07 | 4.02 | 7.50 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 51 | 2.66 | 32.04 | 5.88 | 8.59 | 1.05 | 4.01 | 4.14 | 7.44 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 52 | 8.13 | 30.23 | 5.88 | 8.59 | 1.15 | 4.01 | 4.14 | 7.44 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 53 | 2.46 | 45.66 | 5.88 | 8.59 | 1.06 | 4.01 | 4.14 | 7.44 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 54 | 1.86 | 27.37 | 5.71 | 18.30 | 0.95 | 1.52 | 2.00 | 7.47 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 55 | 1.99 | 34.97 | 5.71 | 18.30 | 0.87 | 1.52 | 2.00 | 7.47 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 56 | 1.21 | 29.02 | 5.71 | 18.30 | 1.13 | 1.52 | 2.00 | 7.47 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 57 | 3.89 | 17.96 | 5.75 | 5.05 | 1.06 | 1.72 | 3.71 | 7.38 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 58 | 6.25 | 22.39 | 5.75 | 5.05 | 1.05 | 1.72 | 3.71 | 7.38 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 59 | 5.30 | 20.45 | 5.75 | 5.05 | 1.17 | 1.72 | 3.71 | 7.38 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 60 | 3.27 | 24.68 | 5.46 | 4.90 | 1.04 | 2.06 | 3.05 | 7.45 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 61 | 2.41 | 21.96 | 5.46 | 4.90 | 1.24 | 2.06 | 3.05 | 7.45 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 62 | 6.34 | 28.44 | 5.46 | 4.90 | 1.14 | 2.06 | 3.05 | 7.45 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 63 | 3.94 | 9.09 | 7.90 | 21.58 | 0.50 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 64 | 1.71 | 11.85 | 7.90 | 21.58 | 1.90 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 65 | 2.13 | 37.45 | 7.90 | 21.58 | 2.92 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 66 | 2.65 | 9.70 | 7.90 | 21.58 | 0.89 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 67 | 2.33 | 23.32 | 7.90 | 21.58 | 1.95 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 68 | 2.17 | 39.82 | 7.90 | 21.58 | 2.93 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 69 | 3.15 | 10.00 | 7.90 | 21.58 | 0.49 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 70 | 3.41 | 50.16 | 7.90 | 21.58 | 4.07 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 71 | 3.21 | 22.70 | 7.90 | 21.58 | 0.95 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 72 | 2.78 | 75.87 | 7.90 | 21.58 | 4.01 | 4.74 | 3.19 | 7.17 | 40.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 73 | 5.82 | 5.77 | 7.50 | 17.02 | 0.82 | 10.50 | 3.75 | 7.77 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 74 | 5.35 | 33.62 | 7.50 | 17.02 | 0.73 | 10.50 | 3.75 | 7.77 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 75 | 5.03 | 22.59 | 7.50 | 17.02 | 0.69 | 10.50 | 3.75 | 7.77 | 40.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 76 | 1.92 | 108.42 | 7.50 | 27.31 | 1.38 | 4.36 | 2.58 | 7.85 | 58.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 77 | 1.51 | 89.03 | 7.50 | 27.31 | 1.38 | 4.03 | 2.56 | 7.99 | 58.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 78 | 8.02 | 95.34 | 7.50 | 27.31 | 1.38 | 3.96 | 2.50 | 7.82 | 58.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |
| 79 | 13.55 | 47.31 | 7.60 | 17.81 | 0.76 | 11.00 | 3.70 | 7.77 | 35.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 80 | 12.49 | 47.78 | 7.60 | 17.81 | 0.85 | 11.00 | 3.70 | 7.77 | 35.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 81 | 3.79 | 35.79 | 7.60 | 17.81 | 0.82 | 11.00 | 3.70 | 7.77 | 35.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 82 | 1.90 | 33.73 | 7.60 | 22.39 | 0.77 | 6.17 | 2.49 | 7.25 | 45.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 83 | 22.85 | 22.42 | 7.60 | 22.39 | 0.66 | 6.17 | 2.49 | 7.25 | 45.00 | 0.00 | 1.00 | 1.00 | 0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 84 | 7.28 | 20.03 | 6.83 | 12.38 | 3.21 | 3.02 | 1.92 | 7.57 | 27.62 | 0.96 | 0.11 | 1.09 | -0.06 | 0.04 | 0.05 | 0.04 | 1.01 | 1.06 | -0.03 |
| 85 | 5.53 | 8.81 | 6.60 | 8.47 | 5.33 | 3.18 | 3.12 | 7.43 | 26.54 | 0.02 | 1.09 | 1.07 | -0.09 | -0.03 | 0.05 | 0.06 | 1.01 | 1.07 | -0.08 |
| 86 | 7.89 | 11.42 | 6.59 | 9.60 | 5.03 | 3.83 | 3.58 | 8.54 | 28.12 | 0.03 | 1.03 | 0.06 | -0.05 | -0.06 | 0.04 | 0.09 | 1.06 | 1.07 | -0.04 |
| 87 | 20.35 | 16.07 | 6.70 | 2.25 | 3.14 | 1.22 | 2.11 | 6.81 | 29.23 | 0.86 | 0.17 | 1.06 | -0.05 | 0.07 | 0.07 | 0.03 | 1.03 | 1.02 | -0.02 |
| 88 | 5.51 | 8.76 | 6.47 | 8.29 | 5.24 | 3.13 | 3.04 | 7.33 | 26.08 | 0.02 | 1.07 | 1.10 | -0.09 | -0.02 | 0.05 | 0.06 | 1.00 | 1.06 | -0.08 |
| 89 | 5.39 | 21.61 | 7.10 | 14.31 | 3.41 | 3.33 | 1.95 | 7.90 | 28.34 | 0.96 | 0.12 | 1.12 | -0.07 | 0.02 | 0.05 | 0.05 | 1.04 | 1.10 | -0.03 |
| 90 | 8.80 | 19.57 | 6.82 | 11.16 | 3.22 | 2.80 | 1.96 | 7.50 | 27.98 | 0.95 | 0.11 | 1.10 | -0.06 | 0.04 | 0.06 | 0.04 | 1.01 | 1.06 | -0.03 |
| 91 | 27.33 | 12.80 | 7.48 | -1.87 | 3.33 | 0.54 | 2.81 | 7.45 | 27.45 | 0.88 | 0.23 | 1.04 | -0.09 | 0.13 | 0.11 | 0.05 | 1.02 | 1.04 | -0.05 |
| 92 | 7.68 | 19.94 | 6.84 | 12.08 | 3.22 | 2.97 | 1.93 | 7.56 | 27.73 | 0.95 | 0.11 | 1.10 | -0.06 | 0.04 | 0.05 | 0.04 | 1.01 | 1.06 | -0.03 |
| 93 | 4.68 | 7.18 | 5.38 | 7.20 | 4.58 | 2.94 | 2.56 | 6.36 | 22.51 | -0.03 | 0.94 | 1.06 | -0.07 | -0.02 | 0.04 | 0.06 | 0.89 | 0.95 | -0.06 |
| 94 | 5.46 | 8.93 | 6.52 | 8.73 | 5.18 | 3.16 | 3.12 | 7.51 | 27.13 | 0.00 | 1.09 | 0.80 | -0.09 | -0.04 | 0.04 | 0.07 | 1.00 | 1.05 | -0.08 |
| 95 | 4.58 | 6.99 | 5.19 | 6.96 | 4.41 | 2.85 | 2.48 | 6.18 | 21.84 | -0.02 | 0.90 | 1.04 | -0.07 | -0.02 | 0.04 | 0.05 | 0.87 | 0.92 | -0.06 |
| 96 | 16.97 | 18.59 | 6.86 | 5.74 | 3.28 | 1.75 | 2.06 | 7.13 | 30.44 | 0.94 | 0.15 | 1.08 | -0.05 | 0.05 | 0.07 | 0.03 | 1.05 | 1.06 | -0.02 |
| Min. | 0.63 | 0.65 | 5.14 | -1.87 | 0.49 | 0.54 | 1.49 | 6.18 | 7.90 | -0.03 | 0.00 | 0.00 | -0.09 | -0.06 | 0.00 | 0.00 | 0.00 | 0.00 | -0.08 |
| Max. | 85.40 | 108.42 | 8.00 | 27.31 | 6.34 | 11.30 | 6.31 | 8.54 | 58.00 | 1.00 | 1.09 | 1.12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.06 | 1.10 | 1.00 |
| Aver. | 8.54 | 24.27 | 6.80 | 12.19 | 2.95 | 4.57 | 3.24 | 7.50 | 31.95 | 0.41 | 0.47 | 0.36 | 0.03 | 0.02 | 0.19 | 0.14 | 0.40 | 0.51 | 0.03 |
| s. d. | 11.68 | 21.74 | 0.82 | 6.47 | 1.72 | 2.66 | 1.02 | 0.37 | 9.51 | 0.49 | 0.49 | 0.49 | 0.20 | 0.14 | 0.39 | 0.34 | 0.49 | 0.51 | 0.18 |

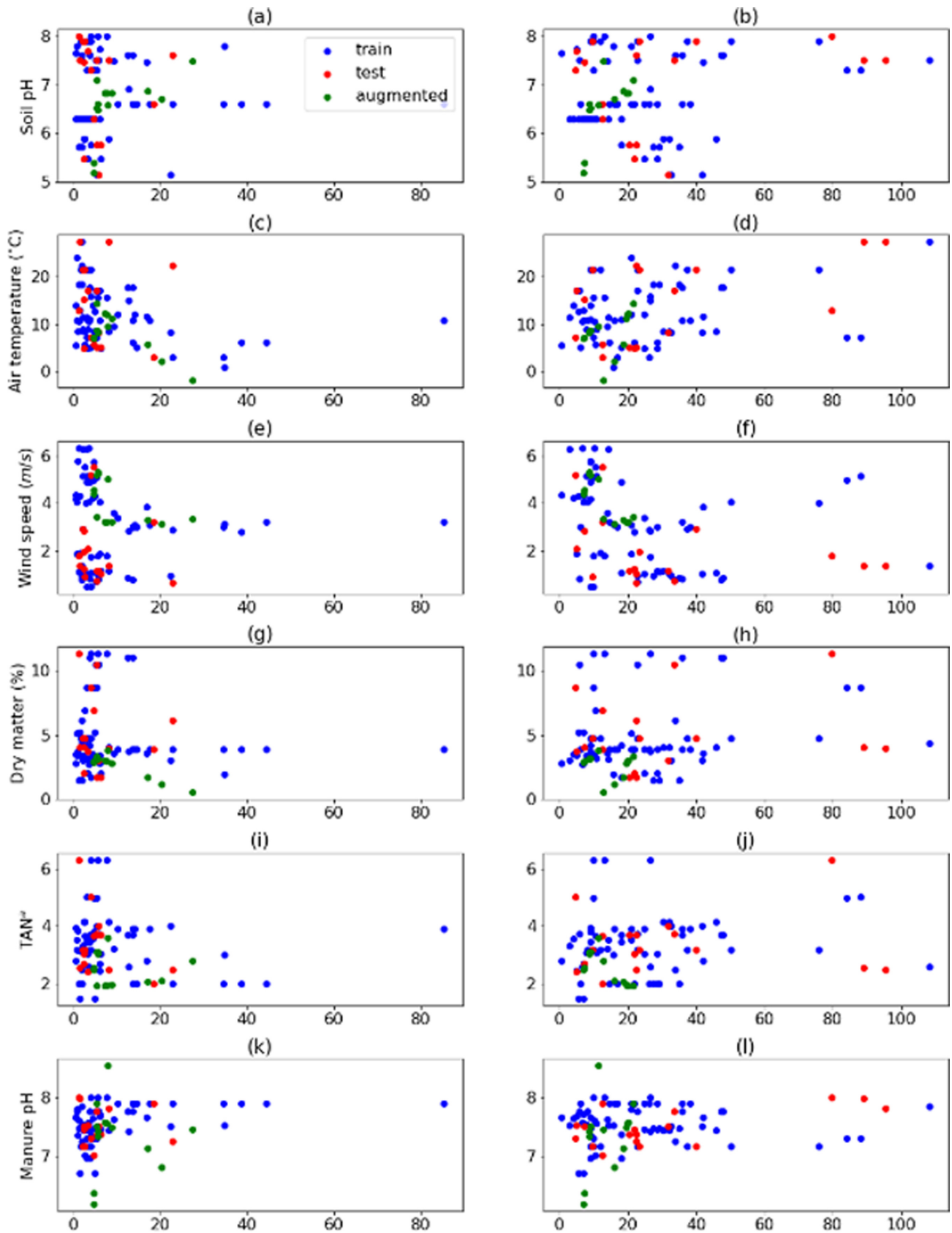


Fig. A2. Distribution of augmented data (blue: original training data; green: augmented training data; red: test data).

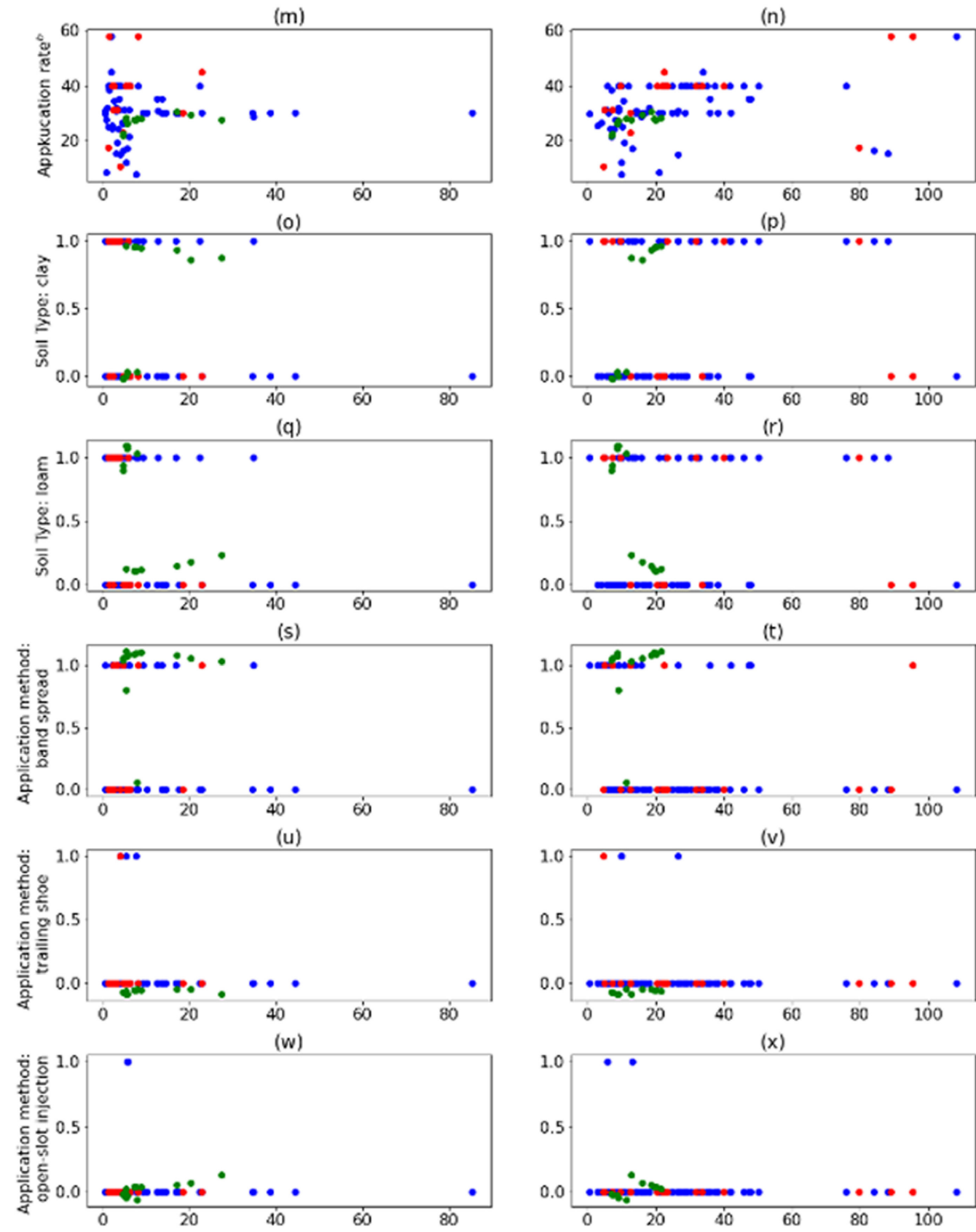


Fig. A2. (Continued)

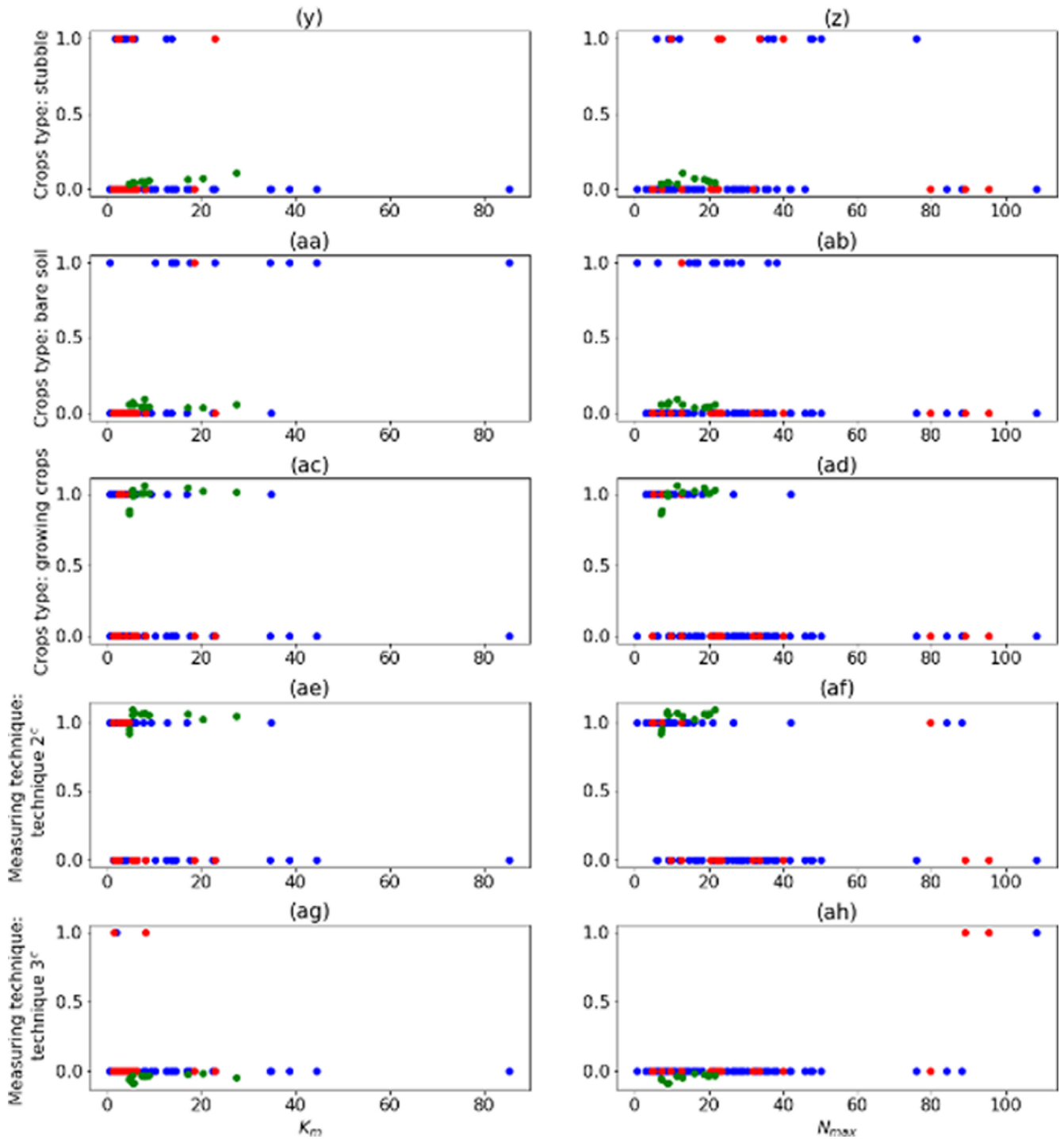


Fig. A2. (Continued)