

Comparative Effects of Different Activation Functions and Node Numbers of Hidden Layer in Artificial Neural Network Models on Real-Time Urban Air Quality Forecasting

Soo-Min Choi^{1,2}, Hyo Choi²

¹Dept. of Computer Engineering, Konkuk University, Korea

²Atmospheric & Oceanic Disaster Research Institute, Korea

The effects of different activation functions (sigmoid and hyperbolic tangent), and node numbers in a single hidden layer of artificial neural network (ANN) models on urban air quality forecasting were compared in Gangneung city (Korea). The ANN models of multilayer perceptron (MLP) with a back-propagation training algorithm for error calculation in cases of 13, 15, and 17 nodes in each hidden layer were performed using 15 input independent variables (PM, gas, and meteorological data of a Korean city), affected by PM and gas of an upwind Chinese city. Root mean square error (RMSE) and the coefficient of determination (R^2 ; Pearson R) were evaluated to assess the two models' forecasting abilities between the predicted and measured values. The values of R by ANN-sig (ANN-tanh) with 13, 15, and 17 hidden neuron numbers were 0.930 (0.950), 0.920 (0.947), 0.926 (0.953) on PM₁₀, 0.953 (0.956), 0.927 (0.938), 0.949 (0.960) on PM_{2.5}, 0.880 (0.959), and 0.917 (0.886), 0.882 (0.939) on NO₂. Regardless of node numbers and activation functions, the prediction abilities of the two models were excellent, showing the highest values of R in the ANN-tanh model with more neuron numbers in the hidden layer. Different from the previous literature insisting that if the number of neurons in the hidden layer than the input layer is set too small (large), the prediction performance will underfit (overfit), this study revealed better the prediction performance with the bigger or smaller neuron number in the hidden layer than the same neuron number as one in the input layer, except for 15 nodes on NO₂ in case of sigmoid function, through their temporal distributions and scatter plots between the predicted and measured values.

Keywords: ANN-sigmoid model, ANN-tanh model, Root mean square error, Coefficient of determination, Pearson R

환경 및 응용기상 분과 / 환경 및 응용기상 2-2

CMAQ 수준 지상 NO₂ 농도 예측을 위한 Sentinel-5P 위성 기반 인공지능 모델 개발

김현서¹, 김은혜², 김순태³, 손락훈¹

¹국립부경대학교 지구환경시스템과학부 환경대기과학전공

²국립군산대학교 환경공학과

³아주대학교 환경안전공학과

정교한 화학 메커니즘을 반영한 대기질 수치모델은 주요 오염물질의 생성 및 소멸을 정밀하게 모의할 수 있는 장점을 지니며, 이러한 특성을 바탕으로 지상 농도 추정을 비롯한 다양한 대기오염 연구에 활용된다. 그러나 높은 계산 비용과 장시간의 연산 문제로 인해 자료의 신속한 산출에는 본질적인 제약이 따른다. 본 연구에서는 Sentinel-5P TROPOMI (TROPOspheric Monitoring Instrument) 위성 자료를 활용하여 CMAQ (Community Multiscale Air Quality) 모델 수준의 지상 NO₂ 농도를 예측할 수 있는 인공지능 모델을 개발하였다. 위성 자료는 공간을 광범위하게 관측할 수 있다는 장점이 있으나, 제한된 관측 시간과 구름 및 에어로졸에 의한 간섭으로 인해 결측값이 발생하거나 자료 품질이 저하되는 한계가 존재한다. 이를 보완하기 위해, 본 연구에서는 Image inpainting 기법을 적용하여 결측값 보정 과정을 거치지 않고 시공간적으로 연속된 지상 농도 예측 자료를 산출하였다. 또한, 예측 정확도 향상을 위해 생성 확산 모델 (Diffusion model)을 추가로 적용하였으며, 그 결과 일부 관측 지점에서는 CMAQ 대비 향상된 예측 성능을 달성하였고, 지역별 오염물질 확산 패턴 또한 더욱 정밀하게 모의되었다. 본 연구는 인공지능 기법을 적용하여 고품질의 지상 대기오염물질 농도 분포를 신속하게 산출할 수 있음을 보였으며, 이는 향후 실시간 대기질 모니터링 및 정책 지원 도구로의 활용 가능성을 시사한다.

Keywords: TROPOMI, CMAQ, machine learning, NO₂

※ 이 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(RS-2024-00343921), 기상청 한국형 도심항공교통(K-UAM) 안전운용체계 핵심기술 개발(RS-2024-00404042)의 지원을 받아 수행되었습니다.

환경 및 응용기상 분과 / 환경 및 응용기상 2-3

효과적인 도시 탄소 배출량 추정을 위한 AI 기반 LPDM 에뮬레이터 구축 및 활용가능성 평가

김주은¹, 정수종²

¹서울대학교 농업생명과학대학 협동과정 농림기상학

²서울대학교 환경대학원 환경관리학과

대기 관측 및 대기수송모델링 기반 하향식(top-down) 탄소 배출량 추정은 국가 온실가스 감축목표(NDC) 이행관리의 과학적 근거를 제공한다. 도시는 국가 탄소 배출량의 큰 부분을 차지하므로, 정확하고 신속한 도시 배출량 추정은 효과적인 NDC 이행을 위해 필수적이다. 도시 규모 탄소 배출량 추정 연구에 주로 활용되는 라그랑지안 입자 확산 모델(Lagrangian Particle Dispersion Model)은 관측 지점(receptor)과 배출원(source) 간의 민감도(풋프린트)를 정량화하며, 관측 지점별로 계산하는 구조를 가진다. 최근 위성·지상 관측 자료와 고해상도 모델링 수요가 증가함에 따라 관측 지점별 풋프린트 산출이 누적되어 대기 수송 모델의 계산 시간과 저장 비용이 증가한다. 본 연구는 이러한 한계점을 보완하기 위해 WRF (Weather Research and Forecasting) 기상장의 주요 변수들과 STILT (Stochastic Time-Inverted Lagrangian Transport) 모델의 풋프린트를 학습한 AI 기반 대기수송 에뮬레이터를 구축하였다. 생성한 에뮬레이터는 기존 대기 수송 모델 대비 수백~수천 배 향상된 연산 속도를 보여주었다. 또한, 에뮬레이터로 산출한 풋프린트를 STILT 기반 풋프린트와 비교하여 공간 패턴 유사도, 시계열 상관성, 평균절대오차(MAE) 등의 지표로 적용가능성을 평가하였다. 본 연구를 통해 고해상도 도시 배출량 추정의 계산 부담을 절감하고, 정확하고 신속한 도시 탄소 배출량 추정에 기여하는 실용적인 대안을 제시하였다.

Keywords: 에뮬레이터, 풋프린트, STILT, 도시 탄소 배출량 추정

※ 본 결과물은 환경부의 재원으로 한국환경산업기술원의 관측기반 온실가스 공간정보지도 구축 기술개발사업의 지원을 받아 연구되었습니다 (RS-2023-00232066).

환경 및 응용기상 분과 / 환경 및 응용기상 2-4

위성자료와 기계학습을 활용한 도시숲 통합대기질지수 개발

우재은¹, 정수종¹, 박찬열², 서흥덕²

¹서울대학교 환경대학원

²국립산림과학원 생활권도시숲연구센터

대기오염물질 저감 기능을 위해 산림·도시숲을 대상으로 구축된 미세먼지 관측망 (전국 45개소, 135개 지점)은 미세먼지에 한정되어 있어 가스상 오염물질에 대한 자료가 부족한 실정이다. 또한, 기존에 사용되는 통합대기질지수는 대기질이 나쁜 지역을 식별하고 대기오염 노출의 심각성을 대중에게 알리기 위한 목적으로 설계되어, 도시숲의 대기질 개선 효과를 평가하는 데에는 한계가 있다. 따라서 본 연구는 인공위성 자료와 인공지능 기법으로 도시숲 내 이산화질소 (NO₂)와 이산화황 (SO₂) 농도를 추정하고, 미세먼지 측정 자료를 함께 활용하여 도시숲의 기능을 정량화하는 통합대기질지수를 제시한다. 이를 위해 Sentinel-5p의 TROPospheric Monitoring Instrument (TROPOMI)에서 산출된 대기오염물질 연직 농도로부터 일별 대기오염물질 지상 농도를 추정하였으며, 토지 이용, 지형, 식생지수 등을 변수로 함께 활용하였다. 연구 결과, Random Forest 모델이 가장 우수한 성능을 보였으며 ($R^2=0.83$, RMSE=4.726ppb), 창경궁 도시숲에 설치된 국립산림과학원 관측 장비 (AQT560)의 이산화질소 농도 자료로 검증하였을 때 높은 상관성을 보였다 ($R=0.81$). 도시숲 통합대기질지수는 도심과 도시숲 내부 대기오염물질 농도 차이를 반영하여 -100에서 100% 사이의 값으로 설정하였다. 남산 도시숲의 경우, 도시숲 통합대기질지수는 인근 용산구 도시대기측정망보다 평균 약 12% 낮게 낮으며 여름에 가장 높고, 가을에 가장 낮게 나타났다. 본 연구에서 제시한 도시숲 통합대기질지수는 주요 대기오염물질을 종합적으로 반영하여, 도시숲의 규모나 유형에 따른 대기질 영향을 정량적으로 평가하는 새로운 지표로 활용될 수 있다.

Keywords: 도시숲, 대기오염물질, 대기질지수, 기계학습, 인공위성

※ 본 결과물은 환경부의 재원으로 한국환경산업기술원의 관측기반 온실가스 공간정보지도 구축 기술개발사업의 지원을 받아 연구되었습니다 (RS-2023-00232066)

※ 본 연구는 국립산림과학원 '도시숲 지역의 대기질 관측 및 예측 기술 개발' 과제(FE0100-2024-03-2025)의 지원을 받아 진행되었습니다.

환경 및 응용기상 분과 / 환경 및 응용기상 2-5

랜덤 포레스트 기반 준실시간 잎 면적 지수(LAI) 산출 방법 개발

정지민¹, 김민중², 이재범³, 홍성철³, 이용희³

¹명지대학교 환경에너지공학과

²명지대학교 환경시스템공학과

³국립환경과학원 대기질통합예보센터

생물기원 휘발성 유기화합물(Biogenic Volatile Organic Compounds, BVOCs)은 오존(O₃) 및 이차 유기 에어로졸(Secondary Organic Aerosols, SOAs)의 주요 전구물질로, 대기질 예측에서 중요한 인자로 작용한다. 이러한 BVOC 배출량은 식생 상태를 정량적으로 나타내는 잎 면적 지수(Leaf Area Index, LAI)에 크게 의존하므로, 신뢰성 있는 LAI 입력자료의 확보는 대기화학 모의와 배출량 추정의 정확도를 높이는 데 필수적이다. 기존에는 높은 정확도를 지닌 GLASS LAI 제품이 널리 활용되었으나, 해당 자료는 계산 비용이 크고 업데이트 주기가 불규칙하여 최신 연도 자료 제공까지 장기간의 공백이 발생한다. 실제로, 2021년 이후 약 3년간 GLASS LAI가 제공되지 않아 연구자들은 과거 자료에 의존할 수밖에 없었고, 이로 인해 실제 식생 변화를 반영하지 못해 대기질 모의 과정에서 상당한 불확실성이 발생하였다. 본 연구에서는 이러한 한계를 보완하기 위해, 위성관측 자료(MODIS, GK-2A 등)와 재분석 기상자료(ERA5)를 결합한 머신러닝 기반 LAI 예측 프레임워크를 개발하였다. 이는 LAI가 계절적·기상학적 요인에 의해 변동한다는 점을 반영하기 위한 것이다. 본 연구에서 구축한 모델은 2018-2023년 자료를 기반으로 학습되었으며, 검증은 2024년(업데이트된 최신 연도) GLASS LAI를 활용하여 수행하였다. 한반도와 동아시아 영역을 대상으로 예측을 수행하여 모형의 일반화 가능성을 평가한 결과, 두 영역 모두에서 신뢰할 수 있는 예측 성능을 보였다. 다만, 여름철 남중국 등 식생 밀도가 높은 지역에서 예측 오차가 다소 증가하는 경향이 나타났는데, 이는 급격한 계절적 변화와 공간적 이질성에 기인한 것으로 판단된다. 이상의 결과는 RF 기반 LAI 예측이 GLASS 자료의 불규칙한 업데이트로 인한 장기간의 시계열 공백을 효과적으로 보완할 수 있으며, 최신 연도의 기상 변화를 반영한 보다 현실성 있는 입력자료를 제공할 수 있음을 보여준다. 본 연구에서 제안한 준실시간 LAI 산출 방법은 향후 대기질 모의 및 기후 연구에서 BVOC 배출량 산정의 불확실성을 줄이고, 최신 자료 기반의 신뢰성 높은 대기 모델링 연구를 가능하게 하는 유용한 접근법으로 활용될 수 있다.

Keywords: 생물기원 휘발성 유기화합물(BVOC), 잎 면적 지수(LAI), GLASS LAI, 랜덤 포레스트(Random Forest)

※ 이 연구는 환경부의 재원으로 국립환경과학원의 지원을 받아 수행하였습니다(NIER-2025-04-02-033).

Global Low-Level Turbulence Prediction Using Tree-Based Machine Learning Methods

Ye-Seul Lee, Hye-Yeong Chun

Department of Atmospheric Sciences, Yonsei University

Low-level turbulence (LLT), often caused by terrain-induced and convective processes, presents persistent challenges for both aviation safety and accurate prediction. This study introduces a machine learning-based framework for global LLT forecasting below 10,000 ft, utilizing Random Forest, Extreme Gradient Boosting, and LightGBM models. The models were trained on approximately 3 million matched pairs of turbulence diagnostics and in situ eddy dissipation rate (EDR) observations from commercial aircraft. All machine learning models achieved comparable performance, which remained stable even when using a reduced set of input diagnostics. Notably, they consistently outperformed the operational Graphical Turbulence Guidance (GTG) system in LLT prediction accuracy. Despite these gains, the machine learning models displayed similar seasonal, diurnal, and altitude-dependent performance patterns to GTG, suggesting that shared limitations stem from the turbulence diagnostic inputs themselves. To further refine the models, threshold sensitivity analysis was conducted to identify optimal classification boundaries, and the Synthetic Minority Over-sampling Technique (SMOTE) was applied to mitigate class imbalance. These enhancements led to improved detection rates for turbulent events, particularly under underrepresented conditions.

Keywords: Low-level turbulence (LLT), Machine-learning, in-situ flight EDR, GTG system, forecast