

논문

대규모 합성 데이터를 활용한 딥러닝 기반 라이다 에어로졸 소산 프로파일링 역산

Deep Learning-Based Lidar Inversion for Aerosol Extinction Profile Using Large-Scale Synthetic Data

이권호^{1),2)}, 노영민^{3),*}, 김관철⁴⁾

¹⁾강원대학교 대기환경과학과, ²⁾강원대학교 복사-위성 연구소,
³⁾국립부경대학교 환경공학과, ⁴⁾서울대학교 차세대융합기술연구원 첨단환경감시센터

Kwon-Ho Lee^{1),2)}, Youngmin Noh^{3),*}, Kwanchul Kim⁴⁾

¹⁾Department of Atmospheric & Environmental Sciences, Kangwon National University, Gangneung, Republic of Korea

²⁾Research Institute for Radiation-Satellite, Kangwon National University, Gangneung, Republic of Korea

³⁾Department of Environmental Engineering, Pukyong National University, Busan, Republic of Korea

⁴⁾Advanced Environmental Monitoring Center, Advanced Institute of Convergence Technology, Seoul National University, Suwon, Republic of Korea

접수일 2026년 2월 13일
수정일 2026년 3월 24일
채택일 2026년 4월 12일

Received 13 February 2026
Revised 24 March 2026
Accepted 12 April 2026

*Corresponding author
Tel : +82-(0)51-629-5626
E-mail : nym@pknu.ac.kr

Abstract Deriving aerosol extinction coefficient profiles from the Light Detection and Ranging (Lidar) observations has been constrained in accuracy and generalization performance due to structural limitations in the initial assumptions of the lidar signal equation. This study aimed to overcome these limitations by developing a deep learning-based inversion methodology combining a large-scale synthetic dataset with a 1D-Convolutional Neural Networks (1D-CNN). A dataset of 10,000 synthetic observations, encompassing diverse boundary layer and upper-level aerosol structures, was constructed for training and testing the deep learning model. The deep learning model completed training for 200 epochs after optimizing the required hyperparameters through grid search. The optimized model achieved excellent performance across all altitudes, with train RMSE = $1.54 \times 10^{-3} \text{ km}^{-1}$ and validation RMSE = $4.59 \times 10^{-3} \text{ km}^{-1}$. Analysis of altitude-specific means and standard deviations showed the predicted profiles from the deep learning model closely overlapped with true values, with bias approaching zero. Furthermore, comparing the proposed deep learning model's performance against traditional Fernald's and Klett's inversion methods revealed that the deep learning model consistently outperformed in both layer structure reproducibility and quantitative accuracy. Therefore, this study demonstrates that the synthetic data-based deep learning inversion method can derive aerosol extinction profiles from Lidar observations with higher accuracy than existing techniques.

Key words: LiDAR, Inversion, Aerosol extinction coefficient, Convolutional neural network, Deep learning

1. 서론

에어로졸의 수직 분포와 광학적 특성은 지표에서 관측된 중량 농도 또는 수 농도만으로는 파악하기 어렵기 때문에, 라이다(Lidar)와 같은 원격 측정 장비를 이용한 방법은 오랫동안 중요한 연구 주제로 다루어

져 왔다(Hong *et al.*, 2004; Schmid *et al.*, 2003; Welton *et al.*, 2000). 라이다 관측 연구는 대기 에어로졸의 수직 구조와 광학 특성을 정량화하여 기후 영향과 미세먼지 문제를 이해하는 데 중요한 역할을 하고 있다(Lee, 2012; Ansmann *et al.*, 1992). 라이다 관측은 대기 중으로 조사된 레이저 빔이 되돌아오는 신호를 정밀하

게 측정하는 과정에서 에어로졸 후방산란계수(back-scattering coefficient)와 소산계수(extinction coefficient)를 분리해야 하므로, 전통적으로는 Klett 또는 Fernald 알고리즘과 같은 수치 해석적 역산법이 널리 사용되어 왔다(Fernald, 1984; Klett, 1981; Fernald *et al.*, 1972). 그러나 수치 해석적 역산법은 라이다 비(lidar ratio), 기준고도에서의 초기 광학계수 등 여러 물리량을 사전에 가정해야 하고, 실제 대기의 시공간 변동성을 충분히 반영하지 못할 경우 시스템적인 오차가 발생한다는 한계가 지적되어 왔다(Wandinger, 2005; Ansmann *et al.*, 1992).

라이다 관측 신호는 다양한 잡음(noise)과 기기 특성, 대기 조건에 따라 변동성이 크므로, 에어로졸 소산계수를 산출하는 과정은 많은 가정과 경험적 판단에 의존하고 있다(Kaestner, 1986). 라이다의 관측 신호 해석을 위한 초기 가정들이 실제 대기의 상태를 충분히 반영하지 못할 경우에는 산출된 소산계수는 계통오차를 포함하게 되며, 특히 대류권 상단이나 에어로졸 층이 중첩된 상황에서는 오차가 더욱 확대되는 경향을 보인다(Kovalev, 1993; Russell, *et al.*, 1979). 이로 인해 동일한 관측 자료에 대해서도 설정 조건에 따라 상이한 소산계수 프로파일이 도출되고, 장기 관측 자료의 비교나 모델 검증에 활용하기에 충분히 일관된 자료를 확보하기 어렵다는 문제가 제기되어 왔다(Böckmann *et al.*, 2004; Klett, 1981). 또한 라이다 역산 과정은 신호 전처리, 기기 보정, 초기값 설정, 품질 관리 단계가 복합적으로 얹혀 있어, 현장에서 실시간으로 안정적인 소산계수 산출 결과를 제공하기가 쉽지 않으므로, 대부분의 라이다 관측에서는 운영자의 판단에 따라 파라미터를 조정하거나, 자료의 품질을 개별 사례별로 검토해야 하는 경우가 발생한다(Thorsen and Fu, 2015). 이와 같은 복잡성은 관측 자료의 활용성을 제한할 뿐 아니라, 동일 자료를 재분석할 때 재현성을 확보하는 데도 장애 요인으로 작용한다. 따라서, 초기 가정값에 대한 민감도가 낮고, 다양한 관측 조건에서 일관된 성능을 보이며, 자동화된 새로운 에어로졸 소산계수 산출 기법에 대한 요구가 꾸준히 제기되고 있다(Liang *et al.*, 2022; Granados-Muñoz *et al.*, 2019).

최근에는 이러한 한계를 보완하기 위하여 합성 자료나 보조 관측과 결합한 새로운 역산 기법, 그리고 물리 모델과 통계적 기계학습을 결합한 하이브리드 접근법이 제안되고 있다(Lopatin *et al.*, 2013; Böckmann *et al.*, 2004). 특히 딥러닝을 포함한 인공지능 기법은 라이다 신호와 에어로졸 광학 특성 사이의 비선형 관계를 직접 학습할 수 있는 장점이 있으며, 인공위성 및 지상 라이다 자료를 이용한 에어로졸 특성 산출에 적용된 사례들이 보고되고 있다(Liang *et al.*, 2022; Granados-Muñoz *et al.*, 2019). 이러한 선형 연구사례들은 물리 방정식을 명시적으로 풀지 않고 라이다 관측 신호와 목표의 물리량 간의 관계를 데이터 기반으로 학습할 수 있다는 장점이 있지만, 인공지능 모델의 학습을 위한 고품질 데이터 확보가 어려운 경우는 여전히 제약이 따른다.

본 연구는 대규모 합성 데이터를 활용하여 딥러닝 기반의 라이다 에어로졸 소산계수 역산 기법을 제안함으로써, 기존의 물리 기반 수치 해석법이 가진 초기 조건 민감성 및 재현성 한계를 극복하고자 하였다. 합성 데이터 기반 인공지능 모델의 학습 전략은 실제 관측에서 얻기 어려운 다양한 시나리오를 설계할 수 있다는 점에서 관측 자료의 부족 문제를 완화하고 모델의 일반화 능력을 강화하는 데 기여할 수 있다. 이와 같은 방법론은 국내외적으로도 최초의 시도이며, 다양한 대기 조건을 포괄하는 합성 데이터 기반 학습을 통해 보다 안정적이고 자동화된 소산계수 산출 가능성을 확인하였다. 실제 라이다 관측 자료에 모델을 적용하여 전통적 방법과 비교 평가함으로써, 본 연구의 방법론이 실질적 관측 환경에 유의미한 개선을 제공할 수 있음을 입증하였다. 이러한 연구는 대기 에어로졸 관측과 기후 및 환경 연구 분야에서 라이다 관측 자료의 활용도를 한층 높이고, 향후 자동화된 실시간 대기 모니터링 시스템 개발에 기여할 것으로 기대된다.

2. 연구 방법

본 연구의 방법론은 합성 데이터를 기반으로 한 AI

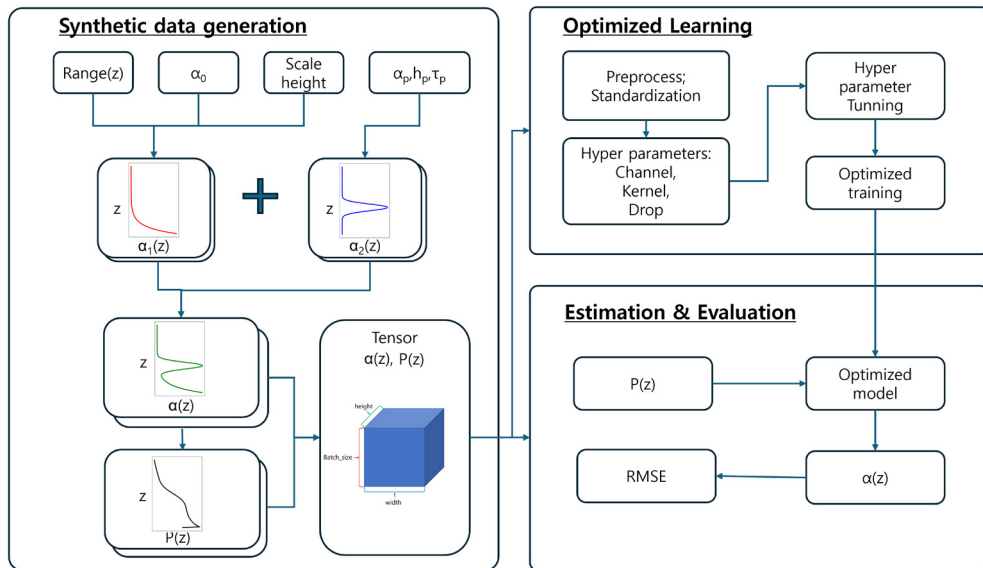


Fig. 1. Block diagram of the deep learning model designed for estimating the aerosol extinction coefficient profile from Lidar signal.

모델 학습과 그 검증 절차로 구성된다(그림 1). 합성 데이터 생성단계에서는 다양한 에어로졸 광학 두께(τ) 및 거리(z)별 분포 특성을 반영한 에어로졸 소산계수 프로파일($\alpha(z)$)을 생성한다. 에어로졸 소산계수 프로파일은 지표를 기준으로 지수적으로 감쇠하는 기본형 프로파일($\alpha_1(z)$)과 특정 고도에서 피크값을 가지는 증가형 프로파일($\alpha_2(z)$)을 통합한 결과로 가정한다. 이로부터 라이다 시스템의 특성값(거리, 해상도, 시스템 상수값 등)이 고려된 이론적인 라이다 관측 신호($P(z)$)를 모의하여 합성 데이터의 현실성을 높였다. 이렇게 생성된 데이터를 입력(라이다 신호)과 출력(소산계수) 쌍으로 사용하는 데이터셋을 거리와 시간 축의 텐서(tensor)로 변환하였다. 다음으로, 이 자료를 학습하는 1차원 합성곱 신경망(1 Dimension Convolution Neural Network (1D-CNN)) 기반 딥러닝 모델을 설계하여 신경망 구조와 손실함수, 물리적 제약 조건을 함께 정의하였다. 합성 데이터셋은 각각 학습, 검증, 시험 구간으로 분할한 후, 1D-CNN 모델의 하이퍼파라미터(Hyper-parameter)를 조정하며 모델 학습을 수행하였다. 하이퍼파라미터의 종류는 채널(channel),

커널(kernel), 드롭아웃(drop out) 값이다.

모델의 예측 결과인 에어로졸 소산계수는 합성 데이터의 에어로졸 소산계수를 참값으로 간주하여 통계적 오차와 상관성을 이용해 성능을 평가하였다. 성능 평가를 위한 지수값은 모델 예측치와 참값 간의 평균 제곱근오차(Mean Squared Error; MSE)와 상관계수(Correlation coefficient)를 주요 지표로 사용하였다. RMSE는 오차의 절대 크기를 직관적으로 나타내며, 상관계수는 예측 결과와 실제값 사이의 선형적 일치 정도를 파악하는 데 적합하다. 이러한 평가 지표들은 Granados-Muñoz *et al.* (2019)과 Lopatin *et al.* (2013) 등의 선행 연구에서도 널리 활용되어 본 연구에서도 검증의 신뢰도를 확보하는 근거로 삼았다. 모델 학습 후 최적 하이퍼파라미터 조정을 거쳐 검증 및 시험 데이터에서 평가를 수행하였다. 이때 가장 우수한 성능을 나타내는 모델을 재학습하여 기존의 수치 해석적 역산 결과와의 차이를 분석함으로써 제안 기법의 유효성을 검증하였다. 각 단계별 데이터 처리와 모델 학습 절차, 성능평가 방법은 다음 장에서 상세히 설명하였다.

2.1 합성 데이터 생성

이 장에서는 라이다 신호 해석을 위한 딥러닝 모델의 학습을 위한 합성 데이터 생성 과정을 상세히 설명한다. 합성 데이터 생성은 에어로졸 광학 특성값과 라이다 방정식을 결합하여, 라이다 관측 신호와 이에 상응하는 에어로졸 소산계수 프로파일을 동시에 생성하는 방식을 적용하였다. 우선, 대류권 경계층 내 혼합층, 상층 연무층, 장거리 수송 황사층 등 대기환경에서 자주 관측되는 상황을 반영할 수 있도록 에어로졸 연직 분포 형태와 광학 두께 범위를 설정하고, 각 거리별(또는 고도별) 에어로졸 소산계수($\alpha(z)$)는 식 (1)과 같이 계산하였다.

$$\alpha(z) = \alpha_0 \cdot \exp\left[-\frac{z}{H_s}\right] + \alpha_p \cdot \exp\left[-\frac{(z-z_p)^2}{(2 \cdot t_p)^2}\right] \quad (1)$$

식 (1)에서, α_0 와 $\alpha(z)$ 는 각각 지표와 거리 z 에서 에어로졸 소산계수, H_s 는 소산계수가 일정 비율로 낮아지는 비율인 척도고도(scale height), α_p 는 거리 z_p 에서 최대값을 가지는 에어로졸 소산계수, t_p 는 증가된 에어로졸 층의 두께를 의미한다. 에어로졸 층의 광범위한 조건을 반영할 수 있도록 각 물리적 변수값의 범위($\alpha_0 = 0.1 \sim 0.5 \text{ km}^{-1}$, $H_s = 500 \sim 3,000 \text{ m}$, $\alpha_p = 0.1 \sim 1.0 \text{ km}^{-1}$, $z_p = 500 \sim 5,000 \text{ m}$, $t_p = 50 \sim 500 \text{ m}$)를 설정하여 무작위로 반복 추출하여 1만 개의 데이터를 생성하였다. 이는 딥러닝 모델의 초기 모델 검증에 충분한 통계적 대표성을 제공하면서도, 최적의 하이퍼파라미터 탐색을 위한 전수조사에 적합한 계산 효율성을 확보하기 위한 규모로 선정되었다.

식 (1)에 의하여 정의된 에어로졸 소산계수 프로파일은 가정된 시스템 파라미터(파장 = 532 nm, 기기상수 = 1 가정)를 고려하여 이론적인 라이다 수신신호를 산출하였다(식 (2)).

$$P(z) = P_0 \cdot C \cdot \frac{O(z)}{z^2} \cdot \beta(z) \cdot \exp(-2 \int \alpha(z') dz') \quad (2)$$

식 (2)에서, $P(z)$ 은 거리 z 에서 수신되는 라이다 신호 강도, P_0 는 대기 중으로 방출되는 레이저 펄스의 초기 송신 강도, C 는 시스템 상수, $O(z)$ 는 송신되는 레이저 빔과 수신부 망원경의 시야(Field of View)가 겹치는

정도를 나타내는 중첩함수, $\beta(z)$ 은 체적 후방산란계수($\text{km}^{-1} \text{ sr}^{-1}$), $\alpha(z)$ 은 체적 소산계수(km^{-1})이다. 이 연구에서는 이론적인 합성 데이터 생성과 딥러닝 학습의 목적으로 사용하기 위하여 C 와 $O(z)$ 의 영향은 없는 것으로 가정하였으나, 향후 특정 라이다 관측 시스템을 대상으로 하는 경우엔 이를 반영한 값이 사용 가능하도록 설계되었다. 이렇게 생성된 라이다 신호는 실제 관측에서 사용하는 수직 해상도와 시간 평균 간격에 맞추어 격자화된 데이터셋으로 변환하여, 최종적으로 소산계수 프로파일에 대한 라이다 신호 데이터의 쌍으로 이루어진 텐서 형태로 재구성하였다. 추후 실제 라이다 관측 자료에 모델을 적용하기 위해, 동일한 전처리 체계를 관측 자료에도 적용할 수 있도록 거리 보정, 신호 정규화 절차를 합성 데이터 생성 단계와 일관되게 정의하였다. 이러한 데이터 생성 과정은 학습용 합성 데이터와 실제 관측 데이터 사이의 형식적 차이를 최소화하여, AI 모델이 합성 데이터로 학습된 후에도 실제 라이다 신호에 안정적으로 적용될 수 있도록 하기 위한 것이다.

2.2 딥러닝 모델 아키텍처

여기에서는 라이다 관측 신호($P(z)$)로부터 에어로졸 소산계수 프로파일($\alpha(z)$)을 산출하기 위하여 설계한 딥러닝 모델의 구조와 구성 요소를 설명한다. 이전 단계에서 생성된 합성 데이터를 이용하여 학습된 딥러닝 모델이 실제 관측 신호에도 안정적으로 적용될 수 있도록 입출력 과정에 필요한 데이터는 일반적인 라이다 관측 자료의 특성과 최대한 일치시켰다. 그리고 딥러닝의 내부 신경망 구조는 수직 프로파일 정보 보존과 잡음 완화를 동시에 고려한 형태로 구성하였다(그림 2). 입력 자료는 2.1장에서 생성한 합성 라이다 신호를 고도에 대한 시퀀스 배열로 변환한 2차원 필드로 정의하고, 이를 채널 차원으로 쌓아 다채널 입력 텐서를 구성하였다. 각 시점의 신호값은 거리를 제곱한 보정값(range corrected signal)과 표준화 과정을 거쳤으며, 고도에 따른 신호 감소와 관측 조건 간의 규모 차이가 모델 학습에 과도한 영향을 주지 않도록 사전 처리하였다. 목표 출력은 동일한 고도-시간 격자에

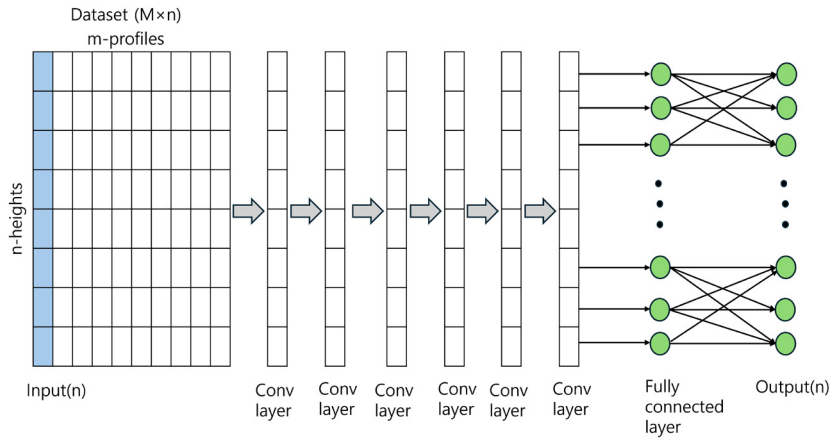


Fig. 2. Architecture of the deep learning model with 6 convolutional layers for aerosol extinction profile estimation from the Lidar receiving signals. The size of the input layer is n-heights units out of m-profiles.

서의 에어로졸 소산계수 프로파일로 정의하였으며, 물리적으로 음수가 될 수 없는 양이라는 점을 반영하기 위해 후단의 활성화 함수(activation function)는 선택과 손실함수 설계에서 이를 명시적으로 고려하였다.

그림 2에서 딥러닝 모델의 구조는 에어로졸 층 구조와 경계층 상단과 같은 수직 특징을 효과적으로 포착하기 위해, 인코더(encoder)-디코더(decoder) 형태의 1D-CNN 구조를 기본 틀로 삼았다. 인코더 부분에서는 여러 단계의 1차원 합성곱과 다운샘플링을 통해 라이다 신호의 국소 패턴과 더불어 광범위한 구조 정보를 추출한다. 여기서 사용되는 1차원 합성곱은 입력 신호(x)와 필터(커널)(w)가 있을 때, 출력(y)의 i 번째 요소는 다음과 같이 계산한다.

$$y[i] = (x \times w)[i] = \sum_j x[i+j] \cdot w[j] \quad (3)$$

식 (3)에서 x 는 입력데이터(lidar signal), w 는 필터 또는 커널(학습되는 가중치), 그리고 i 와 j 는 각각 출력 인덱스와 커널 내에서의 인덱스값이다. 디코더 부분에서는 업샘플링과 합성곱을 반복하여 원래의 고도-시간 해상도로 복원하면서 소산계수를 예측하도록 설계하였다. 이 과정에서 인코더와 디코더 사이에 잔차 연결 또는 U-Net 형태의 Skip Connection을 모사하였다. U-Net은 인코더에서 데이터를 압축하며 특징을

추출하고, 디코더에서는 다시 원래 해상도로 복원하게 되는데, 이때, 인코더의 각 층에서 나온 특징 맵(Feature Map)을 복사하여 대응되는 디코더 층의 결과물과 합쳐주는 것이 Skip Connection으로서, 정보 손실 방지와 학습의 효율화를 위하여 사용된다(Mao *et al.*, 2016; Ronneberger *et al.*, 2015). 따라서, 라이다 신호의 학습과정에서 미세한 층 구조나 경계층의 급격한 변화가 다운샘플링 과정에서 손실되지 않도록 하였다.

이렇게 구성된 딥러닝 모델의 손실함수는 기본적으로 예측된 소산계수와 참값 사이의 평균제곱오차를 최소화하는 것을 목표로 하며, 고도 방향으로의 과도한 진동을 억제하고 물리적으로 불합리한 급격한 변화를 줄이기 위하여 1차 미분항을 추가하였다. 또한 출력층에는 Rectified Linear Unit (ReLU) 활성화 함수를 사용하여 음의 소산계수가 발생하지 않도록 제약을 두었다. ReLU는 단순한 임계값 연산을 통해 비선형성을 부여함으로써, 기존의 Sigmoid나 Tanh 함수에 비해 연산 비용이 낮고 수렴 속도가 훨씬 빠르다는 장점이 있다(Krizhevsky *et al.*, 2012). 그리고 학습 과정에서 발생가능한 과적합을 방지하기 위해 합성곱 계층에 대한 정규화와 일부 은닉층에 대한 드롭아웃을 병행하였다. 이러한 설계를 통해 딥러닝 모델이 학습한 라이다 관측 신호와 에어로졸 소산계수와의 관계

를 일반화하면서도, 다양한 관측 조건에서 안정적으로 작동하는 에어로졸 소산계수 산출 기법이 되도록 하는 것을 목표로 하였다.

2.3 AI 모델 학습 및 검증

라이다 관측 신호 해석을 위한 딥러닝 모델을 학습 시키고, 그 성능을 정량적으로 평가하는 절차는 다음과 같다. 이전의 합성 데이터 생성 파트에서 생성한 이론적인 라이다 관측 신호와 에어로졸 소산계수에 관한 데이터셋을 학습, 검증, 시험 세 부분으로 무작위 분할하였다. 학습 자료는 전체의 70%를 할당하여 모델 파라미터의 최적화를 위해 사용하였고, 검증 자료는 전체의 15%를 배정하여 학습 과정에서의 하이퍼파라미터 조정과 조기 종료 조건 판단에 활용하였다. 나머지 15%는 학습과 검증 과정에 사용되지 않은 독립 시험 자료로 사용하였으며, 딥러닝 모델의 일반화 성능을 평가하였다.

모델의 학습은 배치 경사 하강법을 기반으로 한 Adam 최적화 기법을 사용하여 수행하였다(Ruder, 2016; Kingma and Ba, 2014). 초기 학습률은 몇 가지 후보값에 대해 사전 실험을 통해 결정하고, 학습이 진행됨에 따라 검증 손실의 감소 폭이 일정 기준 이하로 줄어든다면 학습률을 단계적으로 감소시키는 방식으로 안정적인 수렴을 유도하였다. 각 에폭(epoch)마다 학습 손실과 검증 손실을 동시에 모니터링하였으며, 검증 손실이 연속된 여러 에폭 동안 개선되지 않을 경우 학습을 종료하는 조기 종료(early stopping) 전략을 도입하여 과적합을 방지하였다. 하이퍼파라미터(학습률, 배치 크기, 네트워크 깊이, 필터 수, 정규화 계수 등)는 사전 정의한 후보 조합에 대해 반복 실험을 수행한 후, 검증 자료에서의 성능 지표가 가장 우수한 구성을 최종 모델로 선정하는 격자 탐색(Grid search) 기법을 적용하였다(Bergstra and Bengio, 2012). 격자 탐색 과정에서 탐색을 위한 비용이 발생하지만 주요 하이퍼파라미터 영역 내에서 전역적인 성능 경향성을 파악하고 최적의 조합을 확정하기 위해 이 방법을 채택하였다.

딥러닝 모델의 정확도 평가는 예측된 에어로졸 소

산계수와 참값 사이의 평균제곱오차(Mean Squared Error; MSE), 평균절대오차(Mean Absolute Error; MAE), 결정계수(r^2) 등을 계산하여 전체 오차 수준을 분석하였다. 또한 소산계수를 적분하여 얻은 에어로졸 광학 두께와 참값 사이의 차이를 산정함으로써, 프로파일 전체에 걸친 누적 오차를 평가하였다. 또한, 동일한 라이다 관측 신호 사례에 대해 전통적인 역산 기법(Klett 또는 Fernald 알고리즘)을 적용하여 얻은 소산계수 프로파일과 딥러닝 모델의 산출 결과를 비교하였다. 이때 고도별 차이뿐만 아니라, 대류권 경계층 상단 위치, 주요 에어로졸 층의 두께와 최대 소산계수 등 구조적 특징의 재현성을 함께 검토하여, 제안된 기법이 다양한 관측 조건에서 기존 방법과 비교해 어느 정도의 일관성과 안정성을 제공하는지를 평가하였다.

역산 알고리즘 간의 수치적 거동 및 오차 특성을 객관적으로 비교하기 위해, 경계 조건 설정 방식을 단일화하였다. 구체적으로, 합성 데이터의 최고 고도인 10 km 지점에서 소산계수 값을 매우 적은 수(0.00001 km^{-1})로 고정하여 모든 역산 알고리즘의 공통 초기값으로 사용하였다. 이러한 통제된 경계 조건 하에서의 상대 비교는, 각 알고리즘이 상층부의 희박한 에어로졸 구간에서 지표 부근의 농도 구간으로 진입할 때 발생하는 물리적 해법의 차이를 극명하게 보여준다. 특히 지표 부근에서 발생하는 오차의 증폭 현상은 경계 조건의 차이가 아닌, 모델 간의 수치적 수렴 특성 차이임을 명확히 규명하고자 하였다.

Fernald 역산 방법은 라이다 방정식을 기반으로, 분자 및 에어로졸 산란 성분을 분리하여 에어로졸 소산계수 프로파일을 연속적으로 복원하는 고전적 알고리즘이다(Fernald *et al.*, 1972). 이 방법의 장점은 분자 산란이 포함된 일반적인 라이다 방정식에 직접 적용 가능하며, 기준고도를 임의의 고도에 설정할 수 있어 관측 조건에 유연하게 대응할 수 있다는 점이다. 반면, 기준고도에서의 에어로졸 소산계수 가정값과 라이다 비 선택에 민감하며, 수치적 적분 과정에서 잡음 증폭이 발생할 수 있어 적절한 평활화와 품질 관리가 요구된다. 본 연구에서는 에어로졸 소산계수와 라이다 신

호와의 비선형성을 평가하기 위한 목적으로 분자 산란에 관한 항을 무시한 단순화된 형태를 사용하였다. 기준고도 z_{ref} 에서의 에어로졸 소산계수 $\alpha_{fer}(z_{ref})$ 를 알고 있다고 가정하면(또는, 에어로졸의 영향이 거의 없는 조건의 매우 적은 값), 에어로졸 소산계수는 식 (4)와 같이 표현할 수 있다.

$$\alpha_{fer}(z) = \frac{\alpha_{fer}(z_{ref}) \frac{P(z)}{P(z_{ref})} \exp[-2 \int_z^{z_{ref}} \alpha_{fer}(u) du]}{1 + 2S_a \frac{P(z)}{P(z_{ref})} \int_z^{z_{ref}} \exp[-2 \int_z^{z_{ref}} \alpha_{fer}(v) dv] du} \quad (4)$$

위 식은 α_{fer} 가 적분항에도 포함되어 있어 직접 해를 구할 수 없으므로, 초기 산출값 $\alpha_{fer}^0(z)$ 으로부터 적분항을 $\alpha_{fer}^n(z)$ 으로 계산하여 $\alpha_{fer}^{n+1}(z)$ 를 업데이트한다. 그리고 $\alpha_{fer}^{n+1}(z)$ 와 $\alpha_{fer}^n(z)$ 의 차이가 수렴 기준 이하가 될 때까지 반복한다. 즉, Fernald 역산법은 기준고도 조건을 이용하여 수치 해석적으로 전후진 적분(Iterative forward-backward integration)으로 에어로졸 소산계수를 복원한다.

Klett 역산법은 라이다 방정식에서 기준고도 z_{ref} 의 에어로졸 소산계수나 후방산란계수를 알고 있다고 가정하면, 임의 고도 z 에서의 에어로졸 소산계수($\alpha_{klet}(z)$)는 다음의 식 (5)와 같이 계산한다(Klett, 1981).

$$\alpha_{klet}(z) = \frac{\frac{P(z)}{P(z_{ref})}}{\frac{1}{\beta_a(z_{ref})} + 2S_a \int_z^{z_{ref}} \frac{P(u)}{P(z_{ref})} du} \cdot S_a \quad (5)$$

식 (5)는 에어로졸의 영향이 거의 없는 상층의 기준고도 z_{ref} 에서 시작해 지표 방향으로 후진적분(Backward integration)을 수행한다. 이때 기준고도에서 $\beta_a(z_{ref})$ 를 설정하고, 인접 고도 간격 Δz 에 대해, 적분을 누적 합 형태로 근사하여 $\alpha_{fer}(z)$ 를 역순으로 계산한다. 그러나 여기에서는 라이다 신호에서 분자 기여도를 고려하지 않았다. 이후 Klett (1985)는 분자와 에어로졸 성분을 분리한 형태로 확장하였으며, 분자 산란을 고려하여 에어로졸 후방 산란계수를 추출하는 역산식(식 (5))이 제시되었다. 이 역산 알고리즘은 고도에 따른 후방 산란 대 소멸 비율의 변화에 대한 추가적인 개선 사항을 포함한다.

$$\beta(z) = B_p(z)\alpha_p(z) + B_m(z)\alpha_m(z)$$

$$= \frac{\exp(S' - S'_m)}{\frac{1}{\beta(z_{ref})} + 2S_a \int_z^{z_{ref}} \frac{\exp(S' - S'_m)}{B_p(dr')} dr'} \quad (6)$$

위 식에서, 아래 첨자 p 와 m 은 각각 입자와 분자 성분을 의미하며, B_p 는 입자에 의한 라이다 비, 분자에 의한 라이다 비 $B_m = 3/(8\pi) = 0.119 \text{ sr}^{-1}$ 이다.

Fernald 방법이 전진과 후진의 반복적 적분 구조를 가지는 반면, Klett 방법은 한 번의 후진 적분으로 해를 구할 수 있어 계산 효율성이 높으나 기준고도 선택과 라이다 비 가정에 따른 체계적 편향에 더 민감하다(Granados-Muñoz *et al.*, 2014). 즉, Klett 방법은 계산이 간단하고 수렴 특성이 우수하여 운영 라이다 자료 처리에 널리 사용되지만, 초기값 가정이 부정확할 경우 전체 프로파일에 체계적 편향을 야기할 수 있다는 한계를 가진다. 따라서 본 연구에서는 Fernald와 Klett 두 방법을 모두 구현하고, 합성 데이터셋에서의 성능 비교 및 AI 기반 역산 결과와의 상호 검증을 수행함으로써 전통적 역산 기법의 한계를 정량적으로 평가하였다.

3. 결과 및 고찰

3.1 합성 데이터 특성

2.1장에서 설명한 에어로졸 소산계수 프로파일 1만 개와 이에 대응하는 라이다 관측 신호 1만 개의 합성 데이터셋을 생성하였다. 에어로졸 광학 특성은 단일 라이다 비(Lidar ratio = 60 sr) 조건으로 고정하였고, 이는 국내의 라이다 관측에서 관측되는 값의 범위에 해당하며(Noh, 2020; Shin *et al.*, 2017; Kim *et al.*, 2011), 라이다 비 자체의 변동성보다는 소산계수 연직 구조와 신호 잡음이 역산 정확도에 미치는 순수한 영향을 분리 및 평가하기 위함이다. 소산계수 프로파일은 연직 해상도(15 m)와 최대 탐지 고도(10 km)에 맞춘 격자 단위의 값을 가지는 것으로 가정하였다. 단일 라이다 비 조건임에도 불구하고, 소산계수의 층 두께, 최대 값 위치, 감쇠 패턴 등의 다양성으로 인해 딥러닝 모델

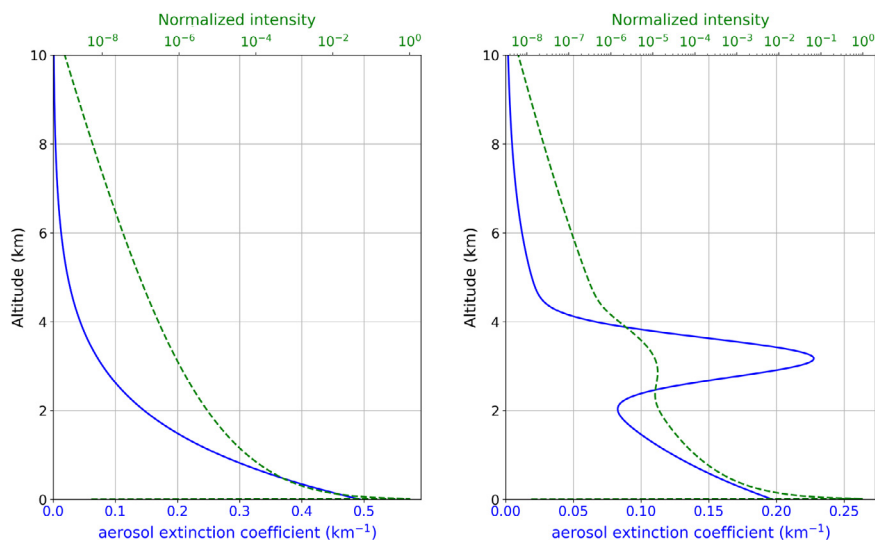


Fig. 3. Two representative aerosol extinction profiles (blue) and corresponding normalized lidar signals (green) randomly selected from the synthetic dataset.

이 라이다 신호 패턴으로부터 소산계수를 학습할 수 있는 충분한 변이를 고려하였다.

그림 3은 생성된 합성 데이터셋에서 무작위로 추출한 두 가지 에어로졸 소산계수 프로파일과 이에 대응하는 라이다 관측 신호를 비교한 사례를 나타낸다. 왼쪽 패널에서 소산계수 프로파일(청색 실선)은 지표에서 0.5 km⁻¹ 수준의 고농도 에어로졸과 이후 지수적으로 감소하는 특성을 보인다. 이에 대응하는 라이다 신호(녹색 점선)도 고도 증가에 따라 일정하게 감소하는 기울기를 나타낸다. 이는 도시에서 전형적인 지표 근처의 고농도 오염층 구조를 잘 모의한 사례이다. 오른쪽 패널에서 소산계수는 지표에서 낮은 값(0.2 km⁻¹)을 보이다가 고도 3 km에서 증가된 최대값을 가지는 2중 에어로졸 레이어의 구조를 나타낸다. 이러한 구조는 장거리 수송되거나 연직방향으로 상승된 에어로졸 층을 반영한다. 이 두 사례는 소산계수 연직 구조의 차이에 따라 라이다 신호의 감쇠 특성이 뚜렷하게 구분됨을 보여준다. 특히 고도에 따른 감소 기울기와 최대값 피크 위치가 에어로졸 층의 위치와 강도를 직접적으로 반영하는 패턴은 딥러닝 모델이 학습해야 할 핵심 신호 특징이다. 실제 라이다 관측에서도 이러한 패

턴이 빈번히 관측되므로, 본 합성 데이터셋이 모델의 현실적 일반화에 기여할 수 있음을 입증한다.

그림 4는 합성 데이터셋의 평균과 $\pm 1\sigma$ 범위에 대한 고도별 에어로졸 소산계수와 이론적 라이다 신호를 나타낸다. 무작위로 생성된 에어로졸 합성 프로파일 데이터는 지표 부근에 강한 에어로졸 층이 형성되고 고도가 증가할수록 점진적으로 사라지는 추세를 나타내며, 라이다 신호도 이에 대응해 지수적으로 약해지는 전형적인 대류권 에어로졸 구조를 잘 재현하고 있다. 보다 상세하게는 에어로졸 소산계수 분포는 지표에서 최대값을 보인 뒤 고도가 증가함에 따라 감소하며, 고도 약 4~5 km에서 완만한 형태를 유지한다. $\pm 1\sigma$ 음영은 고도 0~3 km 구간에서 특히 넓게 나타나고 있어 경계층 내 에어로졸 농도와 층 구조가 다양한 시나리오로 구성되어 있음을 보여주며, 고도 5 km 이상에서는 표준편차 폭이 급격히 줄어 배경 상태가 상대적으로 균질하게 설정되었음을 의미한다. 이에 대응하는 평균 라이다 신호는 로그 스케일에서 거의 직선에 가까운 감소를 보인다. 이는 소산계수 평균 프로파일에 의해 결정되는 전형적인 지수적 감소 형태의 분포 특성을 반영한다. 라이다 신호의 $\pm 1\sigma$ 범위는 고

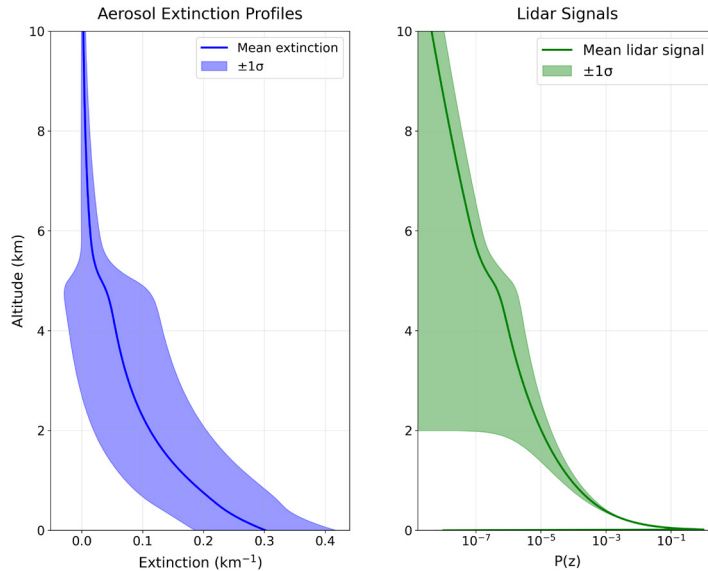


Fig. 4. Vertical profiles of mean aerosol extinction coefficient (left) and normalized lidar backscattered signal (right) from synthetic profiles, with $\pm 1\sigma$ shading indicating statistical variability across diverse boundary layer and free tropospheric aerosol structures.

도 2~5 km 구간에서 크게 벌어져 있는데, 이는 합성 데이터에서 다양한 에어로졸 층 고도와 두께가 고려되어 해당 고도대의 신호 감쇠가 시나리오에 따라 크게 달라질 수 있음을 의미하며, 상층(8~10 km)에서는 신호가 거의 배경 수준으로 수렴하면서 변동성도 크게 줄어든다.

그림 4에서 왼쪽 패널의 평균 소산계수와 오른쪽 패널의 평균 라이다 신호는 서로 일관된 연직분포 형태로 대응하며, 이는 전향 모의 과정이 라이다 방정식을 물리적으로 잘 구현했음을 뒷받침한다. 동시에 두 변수 모두에서 상당한 수의 수직 구조 다양성이 확보되어 있으므로 역산을 위한 딥러닝 모델에서는 경계층 내 강한 오염 이벤트부터 상층의 증가된 에어로졸 층까지 폭넓은 상황을 학습할 수 있는 풍부한 학습 통계 집단을 제공한다는 점에서 합성 데이터셋의 활용성이 높다고 할 수 있다. 이러한 대규모 합성 데이터는 실제 라이다 자료로는 확보하기 어려운 연직 구조의 통계적 다양성을 제공하여, 딥러닝 기반 소산계수 산출 모델의 견고한 학습과 일반화 성능 평가의 기초 자료로 활용되었다.

3.2 AI 모델의 성능평가

생성된 합성 데이터셋을 활용하여 딥러닝 기반 에어로졸 소산계수 역산 모델의 최적화를 위한 하이퍼파라미터 최적화와 성능 평가를 수행하였다. 이 과정에서는 전통적인 격자 탐색 기법을 통해 체계적으로 탐색한 최적 모델을 예측용으로 사용하였다. 즉, 딥러닝 모델의 성능 최적화를 위해 여러 가지 조합의 하이퍼파라미터 조건에서 모델의 성능을 비교하여 가장 성능이 좋은 모델의 하이퍼파라미터 조합을 선택하게 된다. 본 연구에서는 다음과 같이 64개의 하이퍼파라미터 조합(Base channels: [2, 4, 12, 48], Kernel size: [2, 3, 5, 7], Dropout rate: [0.1, 0.2, 0.3, 0.4])에서 격자 탐색을 수행하였다. 각 조합은 100 에폭 동안 합성 데이터셋의 15%에 해당하는 검증 데이터셋에서 오차 수준을 평가하였으며, 이때 계산된 RMSE를 주요 기준으로 최적의 하이퍼파라미터 조합을 선정하였다.

최적화된 하이퍼파라미터는 최종 산출된 소산계수 프로파일의 물리적 타당성에 결정적인 역할을 한다. 일반적으로, 커널 크기는 입력 신호의 특징 추출 범위를 결정하며, 이는 물리적으로 공간 해상도(Spatial

resolution)와 신호 대 잡음비(Signal-to-Noise Ratio) 간의 트레이드오프를 조절하는 과정과 유사하다. 구체적으로, 최적화된 필터 가중치는 고도별 신호의 급격한 변화율을 인식하여 에어로졸 층의 경계를 명확히 식별하는 동시에, 무작위 노이즈에 의한 수치적 진동을 효과적으로 억제한다. 또한, 적절한 학습 파라미터의 설정은 모델이 라이다 방정식의 결정론적 해법을 학습하는 과정에서 지표 부근의 수치적 불안정성을 최소화하고, 물리적으로 유효한 양(+)의 소산계수 범위를 유지하며 해에 수렴하도록 보장한다. 결론적으로, 하이퍼파라미터 튜닝은 단순한 오차 최소화를 넘어, 물리 모델이 가진 파라미터 민감도를 데이터 기반의 가중치 조절을 통해 완화함으로써 최종 해의 강건성(Robustness)과 물리적 재현성을 확보하는 핵심적인 수치 해석 과정이라 할 수 있다.

그림 5는 64개 조합별 모델의 학습 및 검증 결과에 대한 RMSE 값의 크기 순으로 정렬한 막대그래프이다. 하이퍼파라미터 중 Base channel은 1D-CNN 모델의 첫 번째 합성곱 층에서 생성하는 피쳐 맵(Feature Map)의 개수를 결정하는 핵심 하이퍼파라미터이다. 채널 수가 많을수록 라이다 신호에 포함된 다양한 물리적 특성(에어로졸 층의 경계, 신호의 급격한 감쇠, 미세한 노이즈 패턴 등)을 서로 다른 필터가 독립적으로 포착할 수 있다. 채널 수가 너무 많은 경우에는 일반화 성능이 저하되고 연산량이 증가하게 되며, 반대로 채널 수가 적은 경우에는 입력 데이터의 복잡성을 반영하지 못하는 정보 손실이 발생하여 정확도가 떨어질 수 있다(Krizhevsky *et al.*, 2012). 라이다 신호에서 특정 고도의 데이터는 인접한 고도의 상태와 강한 물리적 연관성을 가지게 되는데, Kernel size는 이러한 국소적 패턴(Local Pattern)을 읽어내어 에어로졸 층의 경계나 경사도를 인식한다. 작은 Kernel size는 세밀한 신호 변화를 포착하고, 큰 커널은 신호의 전반적인 흐름이나 거시적인 대기 구조를 파악하는 데 유리한 것으로 알려져 있다(Bai *et al.*, 2018; Luo *et al.*, 2016). 반대로, 작은 커널은 잡음에 민감하게 반응하여 소산계수 프로파일에 지터링(Jittering)이 발생할 수 있

며, 큰 커널은 에어로졸 층의 급격한 경계면이 무시될 수 있으며, 연산 비용이 증가한다. Dropout rate는 모델의 학습 과정 중 무작위로 일부 뉴런의 활성화를 차단하여 모델의 과적합(Overfitting)을 방지하고 일반화 성능을 높이는 핵심적인 규제(Regularization) 기법이다. 일반적으로, Dropout 값이 적을수록 규제 효과가 미미하여 모델이 라이다 신호의 무작위 노이즈까지 학습하게 되어 오차가 커지는 과적합 현상이 발생할 수 있다. 반대로 커질수록 과도한 정규화로 인하여 학습 과정이 불안정해지고 수렴 속도가 느려지며, 데이터의 중요한 패턴을 제대로 학습하지 못한 상태인 과소적합(Underfitting)이 발생한다(Srivastava *et al.*, 2014). 라이다 신호와 같이 물리적 연속성이 중요한 데이터에서는 적절한 Dropout rate를 설정함으로써 딥러닝 모델이 노이즈 패턴에 편향되는 것을 효과적으로 억제할 수 있다.

하이퍼파라미터 튜닝은 딥러닝 모델 성능 개선에 핵심적인 역할을 하므로 학습 데이터와 검증 데이터에서 각각의 평가 지표를 기반으로 최적 조합을 선정한다. 그러나 그림 5의 결과에서는 실제 실험 결과에서 학습 RMSE와 검증 RMSE가 가장 낮은 조합이 일치하지 않는 경우와 두 지표 간 차이를 나타내는 과적합 지표(Overfit (Δ RMSE)) 또한 조합별로 차별적으로 나타난다. 학습 RMSE는 새로운 데이터를 적용하는 경우 성능이 저하될 수 있는 과적합 위험이 있으므로, 최적 조합의 결정을 위해서 검증 RMSE가 최소화되는 조합은 일반화 성능이 높음을 의미하여 우선 후보군으로 선정하였다. 과적합 지표인 'Overfit (Δ RMSE)'은 학습 RMSE와 검증 RMSE의 차이로, 이 값이 클수록 학습 데이터에만 적합되어 일반화 성능이 떨어질 수 있다. 따라서, 검증 RMSE의 최소값의 허용 오차(tolerance) 10% 범위에서 과적합 지표가 가장 낮은 조합을 우선적으로 선택하여 모델의 안정성과 신뢰성을 확보하였다. 이러한 기준을 적용하면, 그림 5의 격자 탐색 결과에서는 Base channel=48, Kernel size=3, Dropout=0.3인 조합에서 최소값 Train RMSE = $3.262 \times 10^{-2} \text{ km}^{-1}$, Validation RMSE = $3.285 \times 10^{-2} \text{ km}^{-1}$, Overfit Δ RMSE =

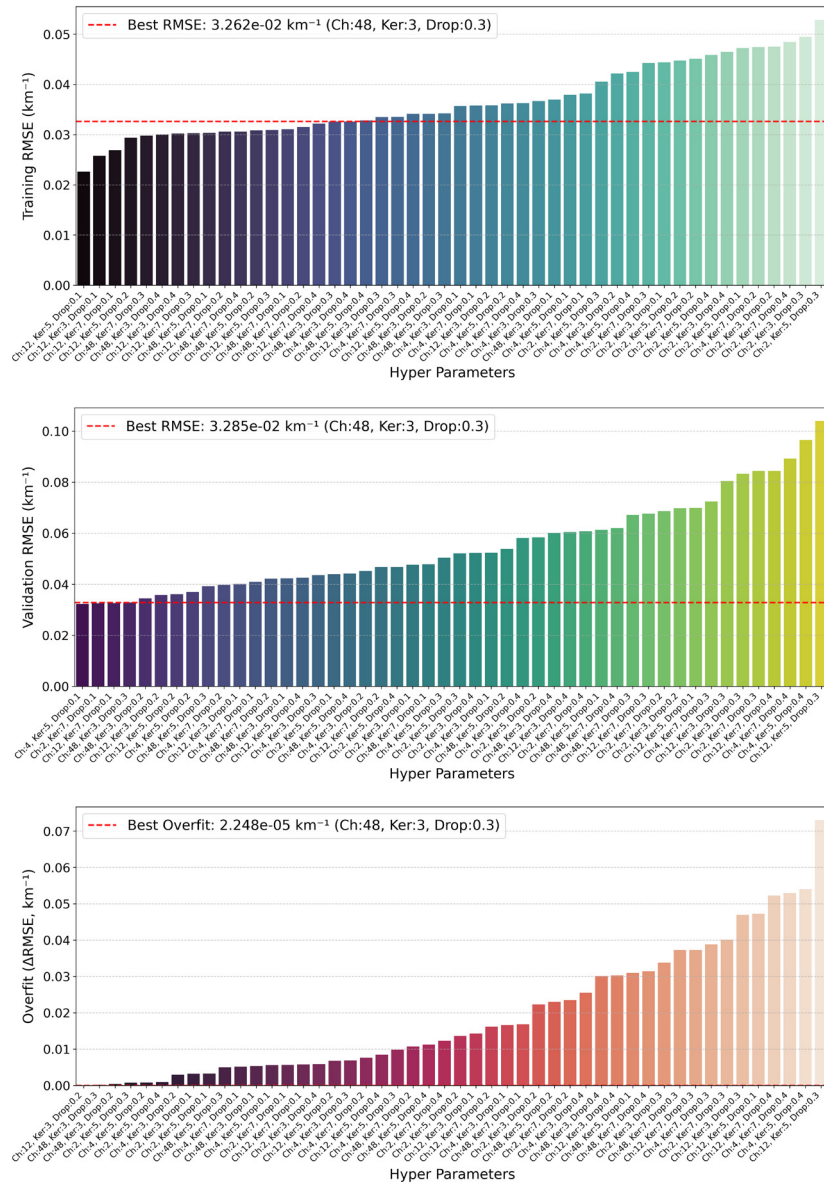


Fig. 5. Comparison of RMSE values for different hyperparameter combinations obtained through a grid search process with training and validation datasets. Training RMSE (top), validation RMSE (middle), and overfit Δ RMSE (bottom) are displayed. The red dashed lines and corresponding text indicate the configuration that yielded the best performance metrics for each dataset: base channel (Ch), kernel size (Ker), and dropout rate (Drop).

$2.248 \times 10^{-5} \text{ km}^{-1}$ 를 나타냈으며, 이 조건에서 프로파일의 통계적 다양성을 효과적으로 활용하면서 과적합이 억제된 것을 확인하였다. 따라서, 이 조합을 라이다 신호의 역산을 위한 딥러닝 모델에서 필요한 최적의

하이퍼파라미터 조합으로 판단할 수 있으며, 이 조합을 이용하여 재학습된 모델을 이용하여 예측과 검증을 시도하였다.

격자 탐색 기법을 통하여 최적 하이퍼파라미터 조

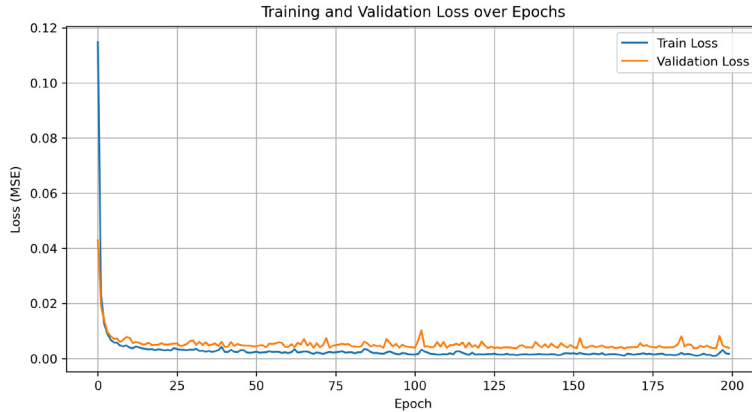


Fig. 6. Training and validation loss curves of the optimized 1D-CNN model.

합을 사용한 1D-CNN 모델에 대해 200 에폭 동안 재 학습을 수행하고, 이때의 에폭별 손실함수 변화는 그림 6과 같다. 학습 손실과 검증 손실 모두 초기 수십 에폭 동안 급격히 감소한 뒤, 약 20 에폭 이후에는 변화 폭이 미미한 평탄 구간에 도달하여 모델이 수렴 상태에 도달했음을 보여준다. 최종 에폭에서의 손실값은 학습 손실 RMSE $1.54 \times 10^{-3} \text{ km}^{-1}$, 검증 손실 RMSE $4.59 \times 10^{-3} \text{ km}^{-1}$ 수준으로 수렴하였으며, 두 곡선 사이의 간격이 에폭 전 구간에 걸쳐 크게 벌어지지 않는 점을 고려할 때, 심각한 과적합은 발생하지 않은 것으로 판단된다. 또한 사용된 손실함수는 소산계수 프로파일의 평균제곱오차에 기반하므로, 최종 손실값은 고도별 에어로졸 소산계수 오차가 약 10^{-3} km^{-1} 수준임을 의미한다. 이러한 결과는 앞에서 제시한 하이퍼파라미터 최적화 단계에서 획득한 검증수치인 Train RMSE = $3.262 \times 10^{-2} \text{ km}^{-1}$, Validation RMSE = $3.285 \times 10^{-2} \text{ km}^{-1}$ 보다 낮은 수준이다. 이렇게 손실함수의 수렴 특성과 시험 세트 평가 결과가 상호 부합하고 있는 결과는 본 연구에서 수행한 격자 탐색 기반 하이퍼파라미터 최적화가 모델의 일반화 성능 확보에 유효하게 기여하였음을 의미한다. 그리고 손실곡선이 특정 에폭 이후 안정적으로 유지되는 패턴은 실제 운영 환경에서 계산 효율을 고려한 조기 종료 기준을 설정할 수 있는 근거를 제시한다.

재학습된 1D-CNN 모델의 예측 성능을 평가하기 위하여 시험 세트 중 임의로 선택한 네 사례에 대해 예측된 소산계수 프로파일과 참값을 비교한 결과는 그림 7과 같다. 각 패널에서 청색 실선은 합성 데이터에 포함된 소산계수(참값), 적색 점선은 라이다 신호를 입력값으로 1D-CNN 모델이 예측한 결과이다. 그림 7a~c 사례에서는 지표 부근의 에어로졸 층과 상층의 증가된 에어로졸 층이 동시에 존재하는 구조임에도 불구하고, 모델이 전체 고도 범위에서 참값의 곡선을 매우 잘 해석하는 양상이 나타난다. 참값과 모델 결과의 비교에 따른 $\text{Bias} > -1.4 \times 10^{-2} \text{ km}^{-1}$ 와 $\text{RMSE} < 5.1 \times 10^{-2} \text{ km}^{-1}$ 로 낮은 수치를 보였다. 특히, 경계층 상단의 급격한 소산계수 감소 구간과 중층의 에어로졸 피크값의 위치가 예측 프로파일에서 정확히 재현되어, 모델이 라이다 신호에 내재된 수직 구조 정보를 효과적으로 학습하였음을 보여준다. 그림 7d~f 사례는 지표 근처에서 주로 존재하는 에어로졸 층과 지수적으로 감소하는 조건을 보여준다. 이 경우에도 모델이 예측한 에어로졸 프로파일은 참값에 비해 다소 거칠게 표현되는 경향이 있으나, 강도와 고도별 감소 경향을 잘 따르고 있다. 참값과 모델 결과의 비교에 따른 $\text{Bias} > -1.6 \times 10^{-2} \text{ km}^{-1}$ 와 $\text{RMSE} < 2.7 \times 10^{-2} \text{ km}^{-1}$ 로 다중 레이어 사례보다 낮은 수치를 보였다. 이러한 결과는 모델이 잡음을 평균화하는 특성 때문에 세부 피

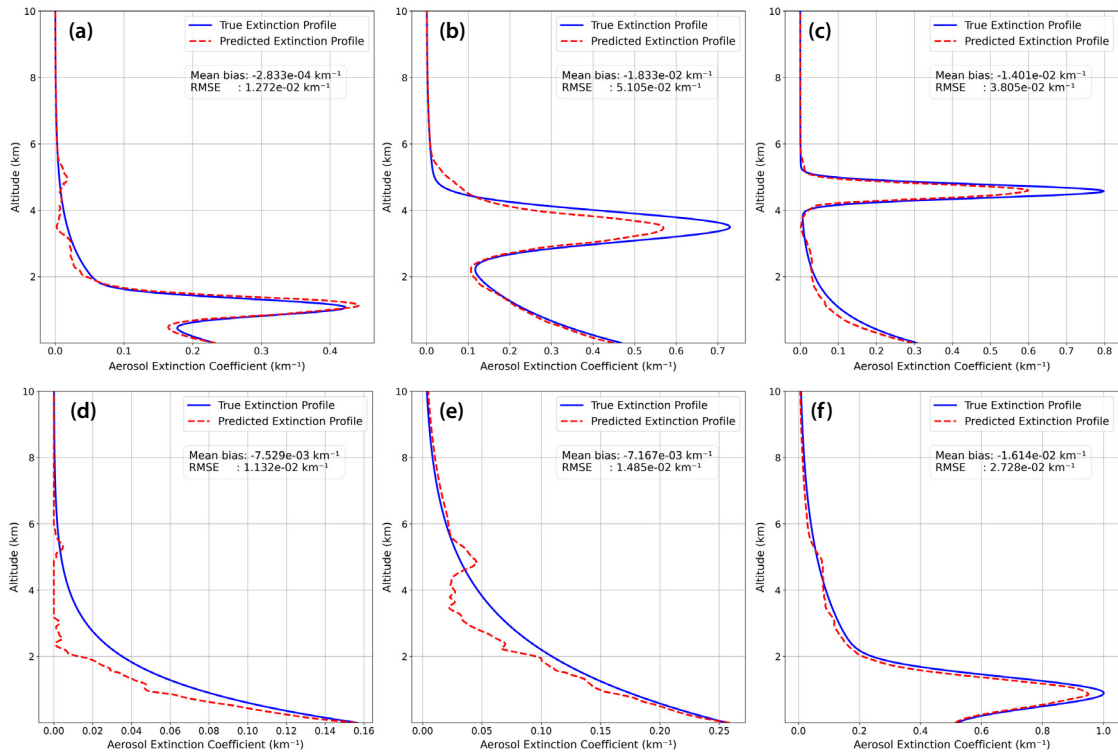


Fig. 7. Test cases comparing 1D-CNN retrieved aerosol extinction profiles (red dashed lines) with the corresponding synthetic ground-truth profiles (blue solid lines).

크의 진폭은 약간 저감시키지만, 대기 에어로졸의 주요 구조적 특징(층 두께, 중심 고도, 수직 기울기)을 신뢰도 높게 복원한다는 점을 의미한다. 종합하면, 그림 7에 제시된 개별 사례 비교는 이전 절에서 제시한 통계적 성능 지표의 결과와 더불어 본 연구에서 개발한 딥러닝 모델이 다양한 형태의 에어로졸 수직 분포에 대해 높은 재현성을 갖는다는 것을 시각적으로 뒷받침한다. 특히, 경계층 내 에어로졸 층과 상승된 에어로졸 층 모두에서 참값과 예측값 간 차이가 낮음을 확인할 수 있어 실제 관측 라이다 자료에 적용할 경우에도 유의미한 역산 성능을 기대할 수 있을 것이다.

그림 8은 최적화된 딥러닝 모델의 일반화 성능을 평가하기 위하여, 합성 데이터셋 중 전체 시험 데이터셋에 대해 예측된 에어로졸 소산계수와 참값의 통계적 특성을 고도별로 비교한 결과이다. 그림 8a는 참값

과 딥러닝 모델 예측값에 대한 고도별 평균 프로파일(실선과 점선)과 ± 1 표준편차 범위(음영)를 함께 나타내며, 전 고도 범위에서 딥러닝 모델 예측값(붉은 점선)은 참값(파란 실선)과 매우 유사한 수직 분포를 보이고 있음을 확인할 수 있다. 특히 고도 0~2 km 범위 내 경계층에서의 소산계수 감소와 3~5 km 부근의 증가된 에어로졸 층 구조가 잘 재현되고 있다. 참값과 딥러닝 모델 예측값의 범위는 $(7.14 \times 10^{-2}) \pm 0.126 \text{ km}^{-1}$, $(6.92 \times 10^{-2}) \pm 0.128 \text{ km}^{-1}$ 로 유사하며, 표준편차 음영이 대부분의 고도에서 상당 부분 중첩되는 것은 모델이 개별 프로파일의 변동성까지 일정 수준 잘 반영하고 있음을 의미한다. 다만 예측 표준편차가 약간 좁게 나타나는 구간은 딥러닝 모델이 일부 극단적인 고농도 또는 저농도 사례를 다소 보수적으로 예측하는 경향이 있음을 의미한다.

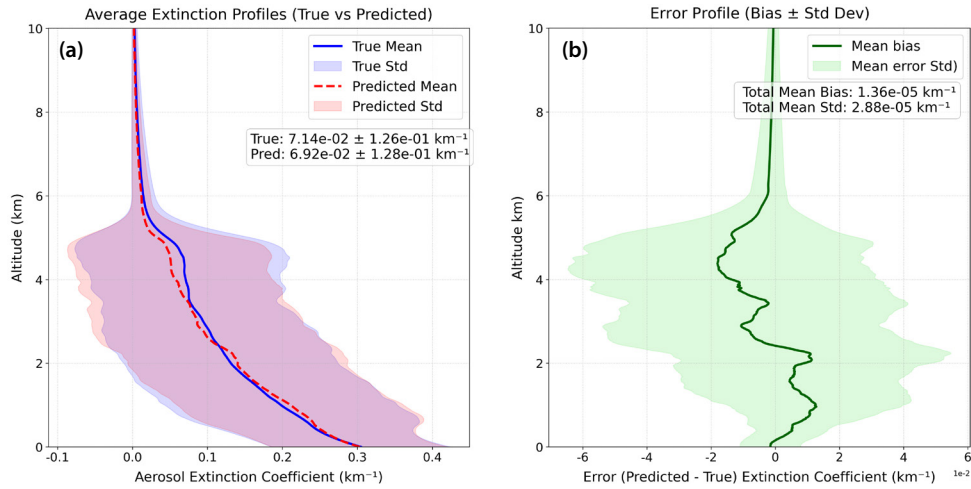


Fig. 8. Height-resolved comparison between true and 1D-CNN predicted aerosol extinction over the entire test dataset: (a) mean extinction profiles with ± 1 standard deviation for true (blue) and predicted (red) values, and (b) corresponding vertical profile of bias (predicted minus true) showing mean bias (solid line) and ± 1 standard deviation (shade).

그림 8b는 딥러닝 모델 예측값에서 참값을 뺀 오차의 고도별 평균 편향과 표준편차를 시각화한 것이다. 평균 편향은 $(1.36 \times 10^{-5}) \text{ km}^{-1}$ 로서 전 고도에 걸쳐 거의 0에 근접하며, $\pm 1\sigma$ 범위는 $(2.88 \times 10^{-5}) \text{ km}^{-1}$ 의 좁은 구간에 수렴한다. 이는 최적 모델이 특정 고도대에서 체계적으로 과대·과소 산출하는 경향이 거의 없고, 오차가 주로 무작위 성분에 의해 지배되고 있음을 보여준다. 경계층 근처 고도 2 km와 중층 고도 4 km에서 오차 표준편차 범위가 다소 증가하는데, 이 구간은 소산계수 수직구배와 층 구조 변화가 가장 큰 영역으로, 역산 난이도가 높은 조건에서도 모델이 전반적으로 편향 없이 변동성만 약간 크게 보이는 수준에 머무른다는 점을 확인할 수 있다. 이와 같은 통계적 분석 결과는, 개별 사례 비교 및 전역 손실함수 지표 결과와 일관성을 가지며, 본 연구에서 개발한 1D-CNN 기반의 딥러닝 모델이 다양한 수직 에어로졸 구조에 대해 높은 정확성과 낮은 편향을 동시에 달성하고 있음을 정량적으로 뒷받침한다. 특히 고도별 평균 편향이 거의 0에 수렴하고, 표준편차가 물리적으로 허용 가능한 범위에 제한된다는 점은, 실제 라이다 관측 자료에 적용할 때도 전체 칼럼 내 층 구조를 신뢰성 있

게 복원할 수 있다는 중요한 근거가 된다.

3.3 역산 기법과의 비교

그림 9a~f는 시험 데이터셋에서 무작위로 추출한 6가지 사례에 대해, 참값 에어로졸 소산계수 프로파일 (검은 실선)과 딥러닝 모델 예측(붉은 점선), Fernald 역산법(파란 점선), Klett 역산법(주황 점선)의 결과를 비교한 것이다. 개별 역산 방법별로 산출된 에어로졸 소산계수는 참값과의 비교를 통해 RMSE를 계산하였으며, 정량적 오차와 수직 구조 재현성을 동시에 평가하였다. 모든 사례에서 딥러닝 모델이 참값에 가장 근접한 형태를 보이며, 전 고도 범위에서 딥러닝 모델 기반의 에어로졸 소산계수에 대한 RMSE 값은 Fernald 및 Klett 방법에 의한 결과보다 약 2~20배 정도 낮게 나타난다. 특히, 경계층 내 에어로졸 층과 상층의 급격한 수직 구배가 존재하는 구간에서는 딥러닝 모델의 예측 프로파일은 참값의 곡선을 잘 따르는 반면, Fernald 역산 결과는 층의 최대값과 경계층 상단 고도를 과소 산출하는 경향을 보인다. 이는 Fernald 역산법이 고정 라이다 비와 기준고도 가정에 의해 결정되는 해가 제약되는 반면, 딥러닝 모델은 라이다 신호와

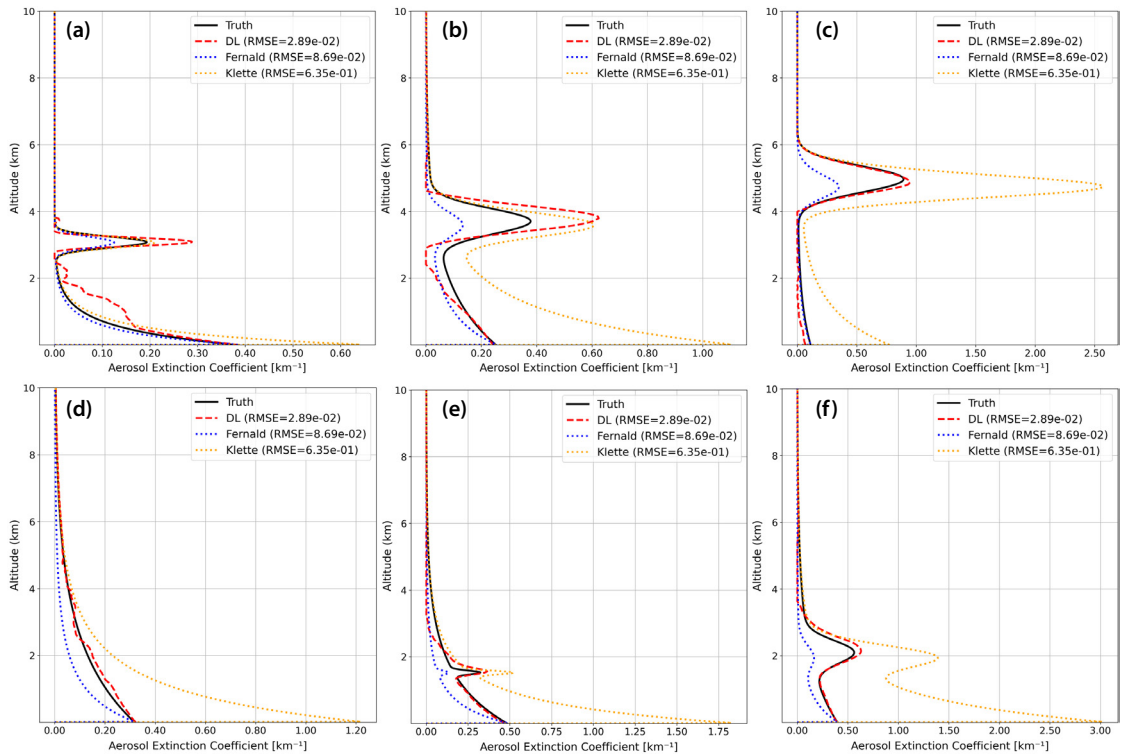


Fig. 9. Case studies comparing extinction profiles from the DL model (red dashed), Fernald inversion (blue dotted), and Klett inversion (orange dotted) against the truth profiles (black), with RMSE.

소산계수 사이의 비선형 관계를 직접 학습함으로써 다양한 층 구조를 더 유연하게 재현할 수 있기 때문이다. Klett 역산법은 이론적으로는 수렴 특성이 우수하지만, 본 연구에서 합성 자료로 재구성한 자료를 이용한 실험 결과에서는 지표 부근(약 0.5 km 이하)에서 오차가 다소 증가하는 경향을 보였다. 이러한 현상은 단일 기준고도에서 소산계수를 가정하는 Klett 알고리즘의 구조적 한계와 관련이 있다. Klett 역산법은 단일 고도의 초기값을 전체 프로파일에 적용하는 과정에서 상층부에서 발생한 초기값 오차가 하층까지 일정 부분 영향을 미침으로 인하여 하층부 오차가 증가하는 것으로 판단된다. 그리고 고도별로 변화하는 에어로졸 특성을 고려하지 않고 고정된 라이다 비를 가정함으로써, 수치적 수렴 속도가 저하되고 이로 인해 지표 부근에서 불확실성이 커지는 것이 원인이 될 수

있다. 반면 딥러닝 모델은 동일한 라이다 신호를 사용함에도 불구하고, 합성 데이터에서 다양한 에어로졸 연직분포 조건을 학습한 덕분에 기준고도 설정 오류에 덜 민감한 결과를 산출하는 것으로 판단된다.

결국, 전통적인 수치 해석 기반의 역산법은 기준고도 근처에서는 참값에 대한 편향이 상대적으로 작지만, 고도가 변할수록 역산을 위한 오차가 누적되고 농도 층의 세부 구조를 충분히 포착하지 못한다는 점이 확인된다. 반대로 딥러닝 모델은 잡음 또는 계통오차를 평균화하면서 에어로졸 층의 위치, 두께, 최대값을 정확히 재현하여 라이다 관측에서 중요한 경계층 높이 및 상층 에어로졸 층 판별에 더 유리한 특성을 보인다. 이러한 사례 비교 결과는 제안된 딥러닝 기반 모델이 고전적인 역산 기법보다 수직 구조 재현성과 정량적 정확도 측면에서 우수한 성능을 제공함

을 뚜렷이 보여준다. 특히 기준고도와 라이다 비 가정에 따른 계통 오차가 큰 상황에서도 딥러닝 모델이 보다 안정적인 해를 제공한다는 점은, 실제 운용 라이다 관측에 딥러닝 기반 역산을 도입할 경우 전통 기법의 한계를 보완할 수 있는 중요한 장점으로 평가된다.

4. 결 론

본 연구에서는 라이다 관측 신호로부터 에어로졸 소산계수 프로파일을 역산하기 위한 딥러닝 기반 방법론을 제안하고, 대규모 합성 데이터셋을 통해 그 성능을 체계적으로 검증하였다. 물리적인 요소를 반영한 합성 데이터셋 구축을 위하여, 다양한 경계층 구조와 상층의 에어로졸 층을 포함하는 에어로졸 소산계수 프로파일과 라이다 신호 쌍을 1만 개 규모로 생성하였으며, 생성된 데이터의 통계적 특성과 수직 구조가 실제 관측에서 기대되는 패턴을 잘 재현함을 확인하였다. 합성 데이터셋을 기반으로 1D-CNN 아키텍처를 설계하고, 기본 채널 수, 커널 크기, 드롭아웃 비율에 대한 격자 탐색 기법을 통한 전수조사를 수행하여 최적의 하이퍼파라미터 조합을 도출하였다. 그 결과, 중간 수준의 모델 용량과 적절한 정규화 강도를 갖는 구성이 검증 손실을 최소화하면서 학습-검증 손실 간의 오차 범위를 억제하며 일반화 성능이 가장 우수한 조합을 확보하였다. 최적의 하이퍼파라미터 조합을 이용하여 1D-CNN 모델을 200 에폭 동안 재학습한 결과는 각각 학습 손실 $RMSE\ 1.54 \times 10^{-3}\ km^{-1}$, 검증 손실 $RMSE\ 4.59 \times 10^{-3}\ km^{-1}$ 수준으로 수렴하였으며, 이는 에어로졸 연직 구조와 총 부하량을 동시에 신뢰성 있게 복원할 수 있음을 의미한다.

합성 데이터셋 중 시험 데이터셋에 대한 정량적 평가와 개별 프로파일 사례 분석을 통하여, 제안된 딥러닝 모델이 경계층 내부와 자유대류권의 에어로졸 층을 포함한 다양한 수직 구조에서 에어로졸 소산계수 프로파일을 높은 정확도로 재현함을 확인하였다. 고도별 평균 및 표준편차 분석 결과는 딥러닝 모델 예측 값이 참값에 거의 중첩되어 편향이 전체 고도에서 0에

근접하는 형태를 보여, 과대 또는 과소 산출 경향이 미미함을 입증하였다. 또한 수직 구조가 복잡한 개별 사례에서도 층 위치, 두께, 최대값이 잘 재현되어, 라이다 관측에서 중요한 경계층 높이 및 상층 에어로졸 층 판별에 활용 가능성이 높음을 증명하였다.

전통적인 라이다 신호 역산 방법인 Fernald 및 Klett 역산법과의 비교에서는 두 방법 모두 에어로졸 층의 최대값 위치와 강도, 고도별 분포에서 유의한 편차를 보인 반면, 딥러닝 모델 기반 역산 결과는 동일한 입력 조건에서도 적은 RMSE 값과 수직 구조 재현성 면에서 일관되게 우수한 성능을 나타냈다. 특히 Klett 방법에서 관찰된 기준고도 오차의 전 고도 전파 및 비물리적으로 평탄한 프로파일 문제는 딥러닝 모델에서는 크게 완화되었으며, Fernald 방법이 갖는 층 경계의 과도한 평활화 문제 역시 딥러닝 모델이 상당 부분 보정하는 것으로 나타났다. 이는 학습 기반 접근법이 전통 역산 알고리즘의 구조적 한계를 보완할 수 있는 유효한 대안임을 보여준다.

결국, 본 연구는 물리 기반 전향모델로 생성한 대규모 합성 데이터셋과 1D-CNN 아키텍처 최적화를 결합함으로써, 라이다 관측 신호로부터 에어로졸 소산계수를 높은 정확도로 산출할 수 있음을 입증하였다. 향후 연구에서는 (1) 다채널 정보 및 동적 라이다비를 고려, (2) 실제 장비별 잡음 특성과 보정 절차를 반영한 합성 데이터 고도화, (3) 실측 라이다 관측과의 일관성 검증 및 전이학습 적용을 통해, 제안된 딥러닝 기반 역산 기법의 실제 운영 적용 가능성을 보다 정밀하게 평가할 필요가 있다. 이러한 후속 연구를 통해, 라이다 관측을 활용한 대기 에어로졸 감시와 환경 분야 응용에서 인공지능 기반 역산 알고리즘의 활용 범위가 크게 확대될 것으로 기대된다.

감사의 글

이 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2026-25486593).

References

- Ansmann, A., Riebesell, M., Wandinger, U., Weitkamp, C., Voss, E., Lahmann, W., Michaelis, W. (1992) Combined Raman elastic-backscatter lidar for vertical profiling of moisture, aerosol extinction, backscatter, and lidar ratio, *Applied Physics B*, 55(1), 18-28. <https://doi.org/10.1007/BF00348608>
- Bai, S., Kolter, J.Z., Koltun, V. (2018) An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, *ArXiv*. <https://doi.org/10.48550/arXiv.1803.01271>
- Bergstra, J., Bengio, Y. (2012) Random Search for Hyper-Parameter Optimization, *The Journal of Machine Learning Research*, 13, 281-305.
- Böckmann, C., Wandinger, U., Ansmann, A., Bösenberg, J., Amiridis, V., Boselli, A., Delaval, A., de Tomasi, F., Frioud, M., Grigorov, I.V., Hågård, A., Horvat, M., Iarlori, M., Komguem, L., Kreipl, S., Larchevêque, G., Matthias, V., Papayannis, A., Pappalardo, G., Rocadenbosch, F., Rodrigues, J.A., Schneider, J., Shcherbakov, V., Wiegner, M. (2004) Aerosol lidar intercomparison in the framework of the EARLINET project. 2. Aerosol backscatter algorithms, *Applied Optics*, 43(4), 977-989. <https://doi.org/10.1364/AO.43.000977>
- Fernald, F.G. (1984) Analysis of atmospheric lidar observations: some comments, *Applied Optics*, 23(5), 652-653. <https://doi.org/10.1364/AO.23.000652>
- Fernald, F.G., Herman, B.M., Reagan, J.A. (1972) Determination of aerosol height distributions by lidar, *Journal of Applied Meteorology and Climatology*, 11(3), 482-489. [https://doi.org/10.1175/1520-0450\(1972\)011<0482:DOAHDB>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<0482:DOAHDB>2.0.CO;2)
- Granados-Muñoz, M.J., Guerrero-Rascado, J.L., Bravo-Aranda, J.A., Navas-Guzmán, F., Valenzuela, A., Lyamani, H., Chaikovsky, A., Wandinger, U., Ansmann, A., Dubovik, O., Grudo, J.O., Alados-Arboledas, L. (2014) Retrieving aerosol microphysical properties by Lidar-Radiometer Inversion Code (LIRIC) for different aerosol types, *Journal of Geophysical Research: Atmospheres*, 119, 4836-4858. <https://doi.org/10.1002/2013JD021116>
- Granados-Muñoz, M.J., Sicard, M., Rocadenbosch, F., Papayannis, A., Nicolae, D. (2019) Characterization of aerosol layers by synergy of lidar and sun-sky photometer measurements and neural network inversion, *Atmospheric Measurement Techniques*, 12(10), 5039-5056. <https://doi.org/10.5194/amt-12-5039-2019>
- Hong, C., Lee, K., Kim, Y.J., Iwasaka, Y. (2004) LIDAR Measurements of the Vertical Aerosol Profile and Optical Depth during the ACE-Asia 2001 IOP, at Gosan, Jeju Island, Korea, *Environmental Monitoring and Assessment*, 92, 43-57. <https://doi.org/10.1023/b:emas.0000014508.76989.a5>
- Kaestner, M. (1986) Lidar inversion with variable backscatter/extinction ratios: comment, *Applied Optics*, 25(6), 833-835. <https://doi.org/10.1364/AO.25.000833>
- Kim, M., Kim, S., Yoon, S., Sugimoto, N., Sohn, B. (2011) Characteristics of the Lidar Ratio Determined from Lidar and Sky Radiometer Measurements in Seoul, *Atmosphere*, 21(2), 57-67, (in Korean with English abstract).
- Kingma, D.P., Ba, J. (2014) Adam: A Method for Stochastic Optimization, *ArXiv*. <https://doi.org/10.48550/arXiv.1412.6980>
- Klett, J.D. (1981) Stable analytical inversion solution for processing lidar returns, *Applied Optics*, 20(2), 211-220. <https://doi.org/10.1364/AO.20.000211>
- Klett, J.D. (1985) Lidar inversion with variable backscatter/extinction ratios, *Applied Optics*, 24(11), 1638-1643. <https://doi.org/10.1364/AO.24.001638>
- Kovalev, V.A. (1993) Lidar measurement of the vertical aerosol extinction profiles with range-dependent backscatter-to-extinction ratios, *Applied Optics*, 32(30), 6053-6065. <https://doi.org/10.1364/AO.32.006053>
- Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012) ImageNet classification with deep convolutional neural networks, *Communications of the ACM*, 60(6), 84-90. <https://doi.org/10.1145/3065386>
- Lee, K.-H. (2012) Aerosol Direct Radiative Forcing by Three Dimensional Observations from Passive- and Active-Satellite Sensors, *Journal of Korean Society for Atmospheric Environment*, 28(2), 159-171, (in Korean with English abstract).
- Liang, Y., Zhang, Y., Wang, L., Li, Z., Chen, H. (2022) Deep learning-based retrieval of aerosol extinction profiles from ground-based lidar measurements, *Atmospheric Environment*, 268, 118783. <https://doi.org/10.1016/j.atmosenv.2021.118783>
- Lopatin, A., Dubovik, O., Chaikovsky, A., Goloub, P., Lapyonok, T., Tanré, D., Bojkov, B. (2013) Enhancement of aerosol characterization using synergy of lidar and sun-photometer data, *Atmospheric Measurement Techniques*, 6(8), 2065-2088. <https://doi.org/10.5194/amt-6-2065-2013>
- Luo, W., Li, Y., Urtasun, R., Zemel, R.S. (2016) Understanding the Effective Receptive Field in Deep Convolutional Neural Networks, *ArXiv*. <https://doi.org/10.48550/arXiv.1701.04128>

- Mao, X., Shen, C., Yang, Y. (2016) Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections, *ArXiv*. <https://doi.org/10.48550/arXiv.1603.09056>
- Noh, Y. (2020) A Study on the Variation of Aerosol Lidar Ratio according to Aerosol Types on Korea Based on AERONET Sun/Sky Radiometer Data. *Journal of Korean Society for Atmospheric Environment*, 36(1), 84-92, (in Korean with English abstract). <https://doi.org/10.5572/KOSAE.2020.36.1.084>
- Ronneberger, O., Fischer, P., Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation, *ArXiv*. <https://doi.org/10.48550/arXiv.1505.04597>
- Ruder, S. (2016) An overview of gradient descent optimization algorithms, *ArXiv*. <https://doi.org/10.48550/arXiv.1609.04747>
- Russell, P.B., Swissler, T.J., McCormick, P. (1979) Methodology for error analysis and simulation of lidar aerosol measurements, *Applied Optics*, 18(22), 3783-3797. <https://doi.org/10.1364/AO.18.003783>
- Schmid, B., Hegg, D.A., Wang, J., Bates, D., Redemann, J., Russell, P.B., Livingston, J., Jonsson, H.H., Welton, E.J., Seinfeld, J.H., Flagan, R.C., Covert, D.S., Dubovik, O., Jefferson, A.J. (2003) Column closure studies of lower tropospheric aerosol and water vapor during ACE-Asia using airborne Sun photometer and airborne in situ and ship-based lidar measurements, *Journal of Geophysical Research*, 108(D23). <https://doi.org/10.1029/2002JD003361>
- Shin, S., Lee, K., Lee, K. (2017) Retrieval of LIDAR Aerosol Parameter Using Sun/Sky Radiometer at Gangneung, Korea, *Current Optics and Photonics*, 1(3), 175-185. <https://doi.org/10.3807/COPP.2017.1.3.175>
- Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014) Dropout: a simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, 15, 1929-1958.
- Thorsen, T.J., Fu, Q. (2015) Automated Retrieval of Cloud and Aerosol Properties from the ARM Raman Lidar. Part II: Extinction, *Journal of Atmospheric and Oceanic Technology*, 32, 1999-2023. <https://doi.org/10.1175/JTECH-D-14-00178.1>
- Wandinger, U. (2005) Introduction to Lidar, in *Lidar*. Springer Series in Optical Sciences, vol 102. Edited by C. Weitkamp, Springer, New York, NY. https://doi.org/10.1007/0-387-25101-4_1
- Welton, E.J., Voss, K.J., Gordon, H.R., Maring, H., Smirnov, A., Holben, B.N., Schmid, B., Livingston, J., Russell, P.B., Durkee, P.A., Formenti, P., Andreae, M.O. (2000) Ground-Based Lidar Measurements of Aerosols During ACE-2 Instrument Description, Results, and Comparisons with Other Ground-Based and Airborne Measurements, *Tellus B*, 52, 636-651. <https://doi.org/10.1034/j.1600-0889.2000.00025>

Authors Information

이권호 (강원대학교 대기환경과학과 교수, 강원대학교 복사-위성 연구소 소장) (kwonho.lee@gmail.com)

노영민 (국립부경대학교 환경공학과 부교수) (nym@pknu.ac.kr)

김관철 (서울대학교 차세대융합기술연구원 첨단환경감시센터 센터장) (fehouse@snu.ac.kr)