

Article

FFT-Based Multiscale Frequency Decomposition for Atmospheric Lidar Attenuated Backscatter Profile Forecasting

Hao Chen ¹ , Zhanpeng Zhang ¹, Jingjing Liu ², Fei Gao ^{2,*}  and Zhimin Rao ³ 
¹ School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China; chen hao@xaut.edu.cn (H.C.)

² School of Optoelectronic Science and Intelligent Instrumentation, Xi'an University of Technology, Xi'an 710048, China

³ School of Electrical and Information Engineering, North Minzu University, Yinchuan 750021, China

* Correspondence: gaofei@xaut.edu.cn

Abstract

Atmospheric light detection and ranging (lidar) measurements of the attenuated backscatter coefficient (ABSC) form high-dimensional vertical profiles that vary across multiple temporal scales. Directly processing the original time-domain sequence with a single prediction structure may entangle information associated with these different scales. To address this issue, we propose FFT-FDNet, a frequency-domain decomposition network based on fast Fourier transform (FFT). FFT-FDNet explicitly decomposes the input sequence along the temporal dimension into low-, mid-, and high-frequency components. Branch-specific predictors model these components separately, and a learnable fusion layer combines their outputs to forecast future ABSC profiles. Experiments were conducted using continuous single-site lidar observations from the Tokyo station at a temporal resolution of 15 min. FFT-FDNet was compared with eight representative time-series forecasting models at forecast horizons of 2, 4, and 6 h. Across the three horizons, FFT-FDNet achieved the lowest mean MAE_{std} and MSE_{std} and the highest mean R^2 among the evaluated methods. At the 2 h horizon, these metrics were 0.1221, 0.4116, and 0.7505, respectively. The ablation results showed consistent performance degradation after removing the FFT decomposition, low-frequency branch, or mid-frequency branch, whereas the high-frequency branch provided modest improvements at some forecast horizons. The frequency band sensitivity analysis supported the use of $(\nu_1, \nu_2) = (0.10, 0.45)$ among the evaluated cutoff combinations. These results suggest that FFT-based three-band decomposition is useful for the present Tokyo single-station short-term forecasting case. Further validation using data from more stations, seasons, and aerosol conditions is still needed.

Keywords: lidar; attenuated backscatter coefficient; time-series forecasting; fast Fourier transform; frequency-domain decomposition; multiscale modeling



Academic Editor: Matthew McGill

Received: 30 June 2026

Revised: 31 July 2026

Accepted: 4 August 2026

Published: 7 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The attenuated backscatter coefficient (ABSC) is an important observable in atmospheric light detection and ranging (lidar) remote sensing for characterizing the vertical distribution of aerosols [1,2]. It can support analyses of boundary-layer evolution, aerosol transport, and local atmospheric disturbances [3,4]. Continuous lidar observations produce time sequences of vertical ABSC profiles, providing high-temporal-resolution information on atmospheric boundary-layer development, vertical pollutant transport, and regional aerosol variability [5,6]. Forecasting future ABSC profiles may provide useful information

for the early warning of pollution episodes, the analysis of aerosol transport, and atmospheric environmental risk assessment. In this study, we focus on a short-term forecasting case based on single-station AD-Net observations from Tokyo.

Time-series forecasting methods have been widely applied to meteorological, environmental remote sensing, and Earth observation data [7,8]. With the development of deep learning, recurrent and temporal convolutional networks have increasingly been used for multivariate forecasting. Long short-term memory (LSTM) networks model temporal dependencies through gated recurrent units [9]. Temporal convolutional networks (TCNs) use dilated causal convolutions to extract local and long-range temporal features, whereas CNN-LSTM models combine convolutional feature extraction with recurrent state modeling to represent local patterns and temporal dependencies [9,10]. Although these models can learn nonlinear temporal variations, high-dimensional multistep forecasting may still be affected by accumulated errors and excessive smoothing of local structures.

Transformer-based models have further advanced time-series forecasting [11,12]. Autoformer models trends and periodic variations through series decomposition and an autocorrelation mechanism [12]. PatchTST [13] uses patch representations to extract features from local temporal segments, while DLinear [14] combines series decomposition with linear mappings to construct a low-complexity forecasting model. These methods represent temporal information through convolution, linear trends, series decomposition, or local segments. However, they do not assign branch-specific prediction structures to temporal components occupying different frequency ranges in ABSC sequences.

In an ABSC sequence, each time step corresponds to a vertical profile with a continuous structure along the height dimension. Along the temporal dimension, the sequence contains slowly varying background signals, adjustments in layered structures, rapid disturbances, and observational noise. The resulting data are high-dimensional and exhibit short-term temporal continuity, a long-tailed distribution, and variability across multiple temporal scales. Processing the original sequence with a single prediction structure may make it difficult to distinguish and learn scale-specific patterns. It is therefore useful to organize information from different temporal scales explicitly and model each frequency band according to its statistical characteristics. The frequency bands in this study represent statistical components at different temporal scales and are not assumed to correspond one-to-one with specific atmospheric physical processes.

Based on this analysis, we propose an FFT-based frequency decomposition network (FFT-FDNet) for short-term forecasting of lidar ABSC profiles in the current single-station setting. The model first applies the FFT to decompose the original sequence into low-, mid-, and high-frequency components. Three prediction branches then model the slowly varying background, local structural variations, and rapid disturbances, respectively. Finally, the branch outputs are fused to generate complete vertical ABSC profiles for multiple future time steps. Unlike a single prediction structure applied directly to the original sequence, FFT-FDNet explicitly separates the frequency components and uses branch-specific structures to organize prediction information at different temporal scales.

The main contributions of this study are summarized as follows:

1. We analyze the long-tailed distribution, short-term temporal correlation, and multiscale spectral structure of Tokyo single-station lidar ABSC profiles for short-term forecasting.
2. We develop a frequency decomposition-driven framework that explicitly separates ABSC sequences into low-, mid-, and high-frequency components and uses branch-specific predictors to organize information at different temporal scales.

3. We evaluate the proposed model on the Tokyo single-station AD-Net dataset at forecast horizons of 2, 4, and 6 h, and compare it with representative time-series forecasting models under the same experimental protocol.
4. We conduct ablation and frequency band sensitivity experiments to assess the effects of three-band decomposition, branch-specific prediction structures, and frequency cutoff selection on forecasting performance.

2. Materials and Methods

2.1. Dataset and Preprocessing

We use 532 nm attenuated backscatter coefficient data collected at the Tokyo station of the Asian Dust and Aerosol Lidar Observation Network (AD-Net), which is operated by the National Institute for Environmental Studies of Japan [15]. The observations have undergone instrument calibration and quality control. Their vertical and temporal resolutions are 30 m and 15 min, respectively, allowing continuous variations in the backscattered signal to be represented along both the temporal and height dimensions. The original AD-Net observations are stored in Network Common Data Form (NetCDF).

For this study, we extract the 532 nm attenuated backscatter coefficient measurements recorded at the Tokyo station between 3 February and 29 June 2023. The resulting data contain 11,424 observation times and 265 consecutive height levels measured upward from the ground. These retained levels are indexed from 0 to 264. Therefore, at a vertical resolution of 30 m, the highest retained level is $264 \times 30 \text{ m} = 7.92 \text{ km}$, and the data cover an approximate altitude range of 0–7.92 km. After quality control, temporal alignment, and selection of valid height levels, the observations are organized as a multivariate time-series in which each time step corresponds to one vertical ABSC profile.

For the model-input construction, the processed file contains only the timestamp and the 265 retained ABSC height levels. No additional cloud mask, signal-to-noise ratio (SNR) variable, or quality control flag is included in this modeling file. Therefore, we do not apply an extra flag-based removal of cloud-contaminated or low-SNR profiles after building the modeling file. The near-ground bins within the retained 0–7.92 km range are also kept because they contain boundary-layer and low-altitude aerosol information. The processed ABSC values used for modeling are finite values. Missing observations are represented by gaps in the time sequence rather than by interpolation. During sliding-window construction, any input-target window crossing a time gap larger than 20 min is removed. Thus, strong backscatter layers, possible cloud-affected returns, and low-signal periods remain part of the forecasting problem, and their effects are reflected in the long-tailed value distribution and the height-wise error analysis.

The processed ABSC sequence is denoted by

$$\mathcal{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]^\top \in \mathbb{R}^{T \times C}, \quad (1)$$

where T is the number of time steps, $\mathbf{x}_t \in \mathbb{R}^C$ is the vertical ABSC profile at time t , and C is the number of retained height levels.

The observations are divided chronologically into training, validation, and test sets at proportions of 70%, 10%, and 20%, containing 7996, 1144, and 2284 time steps, respectively. The training period extends from 00:00 on 3 February to 06:45 on 25 May 2023. The validation period extends from 07:00 on 25 May to 04:45 on 6 June, and the test period extends from 05:00 on 6 June to 23:45 on 29 June. To prevent data leakage, the mean and standard deviation at each height level are calculated using only the training set and then applied unchanged to the validation and test sets.

A continuity filter is applied when constructing sliding-window samples to prevent the model from learning artificial temporal dependencies caused by observation gaps. Because the regular sampling interval is 15 min, one missing observation produces a 30 min gap between adjacent valid observations. Adjacent observations separated by more than 20 min are therefore treated as discontinuous, and a sample is removed if either its input window or forecast window crosses such a discontinuity. Historical observations preceding a dataset boundary may be used as input context for the validation and test sets, but all forecast targets remain strictly within their respective subsets. With an input sequence length of 72, forecast horizons of 8, 16, and 24 steps produce 7522, 7474, and 7426 valid training samples, respectively.

The main statistical characteristics of the ABSC dataset are summarized in Table 1.

Table 1. Main statistical characteristics of the ABSC dataset.

Statistic	Value
Temporal resolution	15 min
Vertical resolution	30 m
Number of height levels	265
Altitude range	0–7.92 km
Autocorrelation at a 15 min lag	0.9759
Autocorrelation at a 1 h lag	0.7121
Autocorrelation at a 2 h lag	0.4014
Autocorrelation at a 6 h lag	0.1192
Main data characteristics	High-dimensional profiles, short-term continuity, long-tailed distribution, and multiscale variability

The relatively high autocorrelations at lags of 15 min and 1 h indicate strong short-term temporal continuity. ABSC values are generally small but exhibit a pronounced right tail caused by localized strong-scattering events. These properties motivate the separate treatment of slowly varying background signals, local structural variations, and rapid disturbances.

2.2. Characteristics of the ABSC Data

Each time step in an ABSC sequence corresponds to a complete vertical profile. The data therefore contain both temporal evolution and layered structures along the height dimension. ABSC sequences generally exhibit high-dimensional vertical profiles, short-term temporal continuity, long-tailed value distributions, and variability across multiple temporal scales. To characterize the difficulty of ABSC profile forecasting and provide empirical support for the frequency-domain design of FFT-FDNet, we analyze the training set in terms of value distribution, temporal autocorrelation, and spectral energy distribution.

First, ABSC observations have a high-dimensional vertical profile structure. Atmospheric lidar systems continuously measure backscattered signals along the vertical direction [16,17]. Aerosol layers, the top of the atmospheric boundary-layer, and cloud–aerosol mixed layers generally appear as layered structures extending over a range of altitudes [18,19]. The prediction target at each time step is therefore not a single scalar but a complete ABSC profile containing observations from multiple height levels.

Second, the distribution of ABSC values is strongly right-skewed and long-tailed. Figure 1 shows the histogram of the training data. To prevent a small number of extreme values from dominating the displayed range, only samples between the 0.1th and 99.9th percentiles are shown, and the sample counts are plotted on a logarithmic scale. Most ABSC observations are concentrated near zero, while a long tail extends toward larger values. This pattern is consistent with weak backscattering under clear-sky or relatively clean

atmospheric conditions and locally enhanced values during aerosol transport, pollution accumulation, cloud occurrence, or strong scattering events [4,20]. A forecasting model must therefore represent the low-value background while remaining responsive to localized variations and less frequent enhancement events.

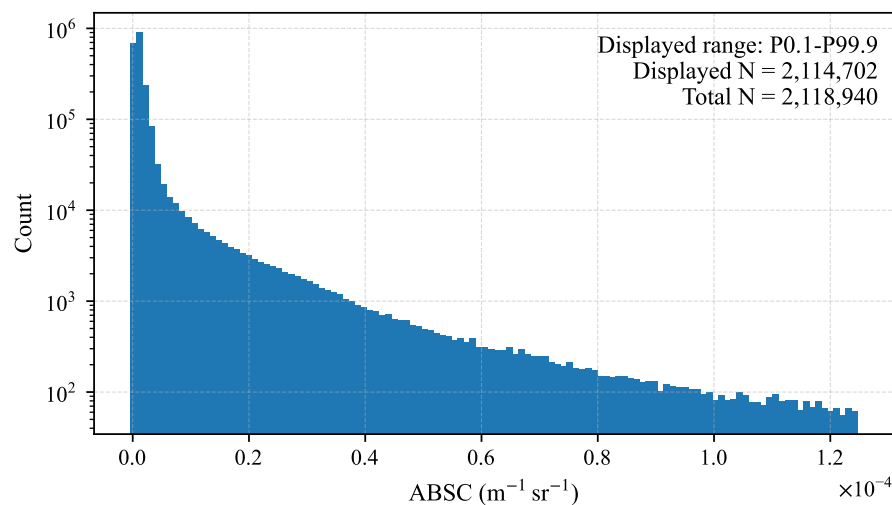


Figure 1. Histogram of ABSC values in the training set. Samples between the 0.1th and 99.9th percentiles are shown. The vertical axis represents the sample count on a logarithmic scale.

Third, ABSC sequences exhibit strong continuity over short time scales, while their temporal correlation decreases as the lag increases. Figure 2 presents the mean autocorrelation coefficients for lags of up to 24 h. Autocorrelation is calculated separately for each height level within continuous observation segments and then averaged while accounting for the number of overlapping samples. At the 15 min sampling interval, lags of 1, 4, 8, and 24 steps correspond to 0.25, 1, 2, and 6 h, respectively. Their mean autocorrelation coefficients are approximately 0.9759, 0.7121, 0.4014, and 0.1192. These results indicate strong short-term continuity, but the rapid decrease in correlation makes forecasting increasingly difficult as the forecast horizon extends.

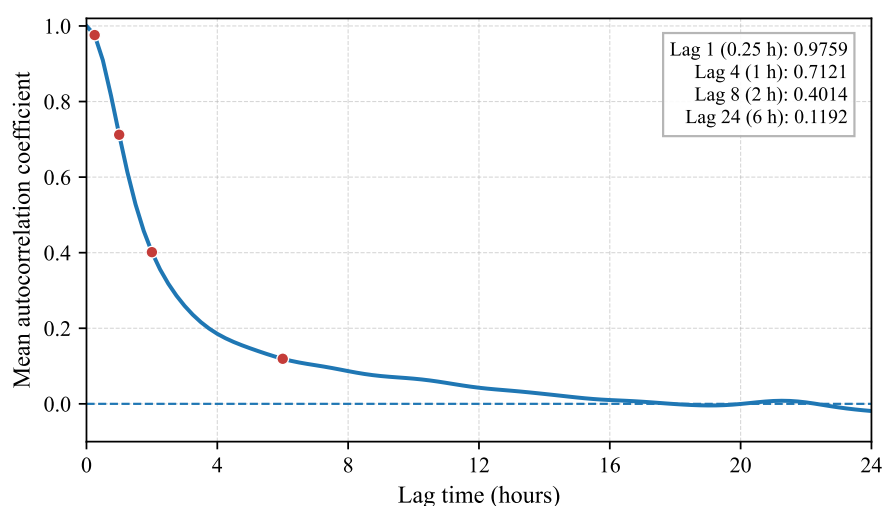


Figure 2. Mean autocorrelation coefficients of the training ABSC sequences for lags from 0 to 24 h. The statistics are calculated only within continuous observation segments at a sampling interval of 15 min; the coefficients at lags of 1, 4, 8, and 24 steps are highlighted and listed in the figure.

ABSC sequences also contain slowly varying background signals, local structural variations, and rapid disturbances [21,22]. Figure 3 shows the normalized spectral energy distribution. Continuous 72-step windows from the training set are standardized by height level and normalized within each window before applying the FFT along the temporal dimension. Spectral energy is aggregated across windows and height levels and then normalized so that the energy over all discrete frequency bins sums to one. A logarithmic vertical axis is used to display frequency bins with low energy.

Most of the energy is concentrated in the low-frequency region. Components at frequencies no higher than 0.10 cycles/sample account for approximately 94.08% of the total energy. Mid-frequency components between 0.10 and 0.45 cycles/sample account for approximately 5.72%, while high-frequency components above 0.45 cycles/sample account for approximately 0.21%. Thus, ABSC profile variations are dominated by slowly evolving background signals but retain a measurable proportion of mid-frequency structural variation. High-frequency components have considerably lower overall energy and are more closely associated with rapid local disturbances or observational noise.

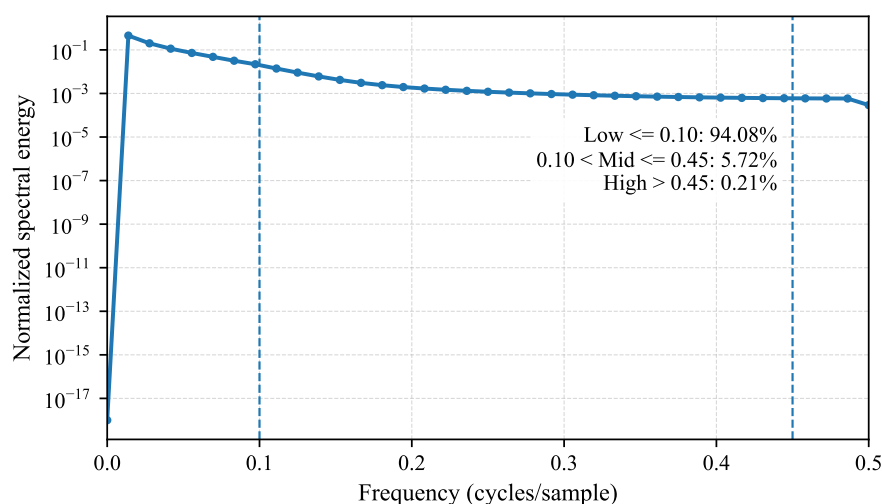


Figure 3. Normalized spectral energy distribution of the training ABSC sequences. Frequency is expressed in cycles/sample, the Nyquist frequency is 0.5 cycles/sample, and the frequency cutoffs are 0.10 and 0.45. Normalized spectral energy is shown on a logarithmic scale.

Actual atmospheric lidar observations may contain discontinuities or local outliers caused by instrument maintenance, weather-related obstruction, quality control procedures, or data filtering. Forecasting samples must therefore be constructed chronologically, and neither the input nor forecast window should cross a clear discontinuity. This ensures that the temporal dependencies learned by the model originate from continuous observation periods.

Taken together, ABSC profile forecasting requires a model to represent vertical profile structures, capture short-term temporal evolution, accommodate a long-tailed value distribution, and handle variability across multiple temporal scales. The statistical results suggest that directly modeling mixed-scale information in the time domain may introduce interference among slowly varying background signals, local structural variations, and rapid disturbances. We therefore use the FFT to separate the original ABSC sequence into low-, mid-, and high-frequency components and employ branch-specific structures to organize prediction information at different temporal scales.

2.3. Problem Formulation

The task considered in this study is multivariate multistep time-series forecasting. Given a historical observation window of length L ,

$$\mathbf{X}_{t-L+1:t} = [\mathbf{x}_{t-L+1}, \mathbf{x}_{t-L+2}, \dots, \mathbf{x}_t]^T \in \mathbb{R}^{L \times C}, \quad (2)$$

the model predicts a future ABSC profile sequence of length P ,

$$\mathbf{Y}_{t+1:t+P} = [\mathbf{x}_{t+1}, \mathbf{x}_{t+2}, \dots, \mathbf{x}_{t+P}]^T \in \mathbb{R}^{P \times C}. \quad (3)$$

The forecasting model is expressed as

$$\hat{\mathbf{Y}}_{t+1:t+P} = f_{\theta}(\mathbf{X}_{t-L+1:t}), \quad (4)$$

where $f_{\theta}(\cdot)$ denotes the model parameterized by θ , and $\hat{\mathbf{Y}}_{t+1:t+P} \in \mathbb{R}^{P \times C}$ is its output. We adopt a multivariate-input, multivariate-output setting in which both the input and forecast windows contain ABSC observations from all retained height levels.

During training, the model parameters are optimized by minimizing the error between the predicted and observed future sequences. Given the training set $\mathcal{D}$, the objective is

$$\theta^* = \arg \min_{\theta} \frac{1}{|\mathcal{D}|} \sum_{(\mathbf{X}_{t-L+1:t}, \mathbf{Y}_{t+1:t+P}) \in \mathcal{D}} \mathcal{L}_{\text{MAE}}(f_{\theta}(\mathbf{X}_{t-L+1:t}), \mathbf{Y}_{t+1:t+P}). \quad (5)$$

For one forecasting sample, the sample-level mean absolute error loss is

$$\mathcal{L}_{\text{MAE}} = \frac{1}{PC} \sum_{p=1}^P \sum_{c=1}^C |\hat{Y}_{p,c} - Y_{p,c}|. \quad (6)$$

2.4. Proposed FFT-FDNet

2.4.1. Overall Architecture

FFT-FDNet is designed to account for the short-term temporal continuity, high-dimensional vertical profiles, long-tailed value distribution, and multiscale variability of ABSC sequences. As shown in Figure 4, the model consists of three main modules: FFT-based decomposition, branch-specific forecasting, and output fusion. Here, B denotes the batch size, L the input sequence length, P the forecast horizon, and C the number of height levels.

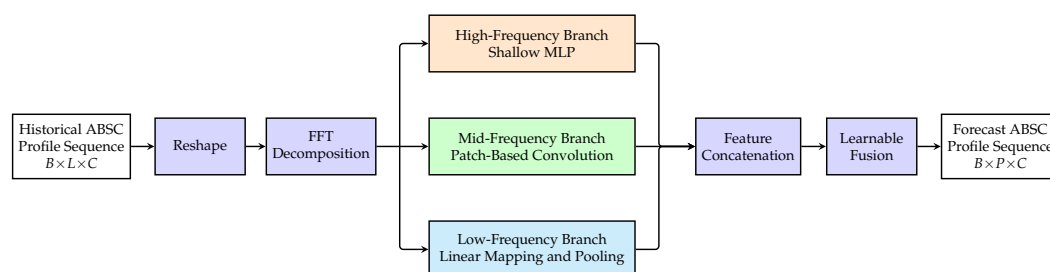


Figure 4. Overall architecture of FFT-FDNet. The historical ABSC profile sequence is separated along the temporal dimension into three frequency band components using the FFT. The components are processed by branch-specific predictors and subsequently fused to produce the future ABSC profile sequence.

FFT-FDNet first transforms the historical ABSC profile sequence into the frequency domain along the temporal dimension and separates it into low-, mid-, and high-frequency components using predefined cutoffs [23,24]. The low-frequency branch employs linear

mappings and average pooling, the mid-frequency branch combines temporal patching with convolutional operations, and the high-frequency branch uses a shallow multi-layer perceptron. The three branch outputs are then concatenated and passed through a learnable fusion layer to generate complete ABSC profiles for the next P time steps.

2.4.2. FFT-Based Frequency Decomposition

Given a batch of historical observations $\mathbf{X} \in \mathbb{R}^{B \times L \times C}$, a one-dimensional real-valued FFT is independently applied to the temporal sequence at each height level. Because the ABSC observations are real-valued, the Fourier coefficients exhibit conjugate symmetry. Only the $K = \lfloor L/2 \rfloor + 1$ non-negative-frequency coefficients are retained, yielding $\mathbf{Z} \in \mathbb{C}^{B \times K \times C}$:

$$Z_{b,k,c} = \sum_{n=0}^{L-1} X_{b,n,c} \exp\left(-j \frac{2\pi kn}{L}\right), \quad k = 0, 1, \dots, \left\lfloor \frac{L}{2} \right\rfloor. \quad (7)$$

Here, j is the imaginary unit, and $Z_{b,k,c}$ is the complex Fourier coefficient at frequency index k for height level c of sample b . The normalized frequency corresponding to index k is

$$\nu_k = \frac{k}{L}, \quad k = 0, 1, \dots, \left\lfloor \frac{L}{2} \right\rfloor. \quad (8)$$

Its unit is cycles/sample, and its range is $[0, \lfloor L/2 \rfloor / L] \subseteq [0, 0.5]$. When L is even, $\nu_k = 0.5$ corresponds to the Nyquist frequency.

To separate statistical components associated with different temporal scales, two normalized-frequency cutoffs satisfying $0 < \nu_1 < \nu_2 < 0.5$ are introduced [25]. The low-, mid-, and high-frequency masks are defined as

$$M_{\text{low}}[k] = \begin{cases} 1, & \nu_k \leq \nu_1, \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

$$M_{\text{mid}}[k] = \begin{cases} 1, & \nu_1 < \nu_k \leq \nu_2, \\ 0, & \text{otherwise,} \end{cases} \quad (10)$$

$$M_{\text{high}}[k] = \begin{cases} 1, & \nu_k > \nu_2, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

Each mask satisfies $\mathbf{M}_q \in \{0, 1\}^K$. The masks do not overlap and jointly cover all non-negative frequencies. During computation, they are broadcast along the batch and height dimensions. Frequency band selection is performed by

$$\mathbf{Z}_q = \mathbf{Z} \odot \mathbf{M}_q, \quad q \in \{\text{low}, \text{mid}, \text{high}\}, \quad (12)$$

where $\odot$ denotes element-wise multiplication. Each frequency band component is then reconstructed in the time domain using a one-dimensional inverse real FFT with output length L :

$$\mathbf{X}_q = \mathcal{F}_r^{-1}(\mathbf{Z}_q), \quad q \in \{\text{low}, \text{mid}, \text{high}\}. \quad (13)$$

The reconstructed sequences $\mathbf{X}_{\text{low}}$, $\mathbf{X}_{\text{mid}}$, and $\mathbf{X}_{\text{high}}$ have the same dimensions as the original input and satisfy

$$\mathbf{X} = \mathbf{X}_{\text{low}} + \mathbf{X}_{\text{mid}} + \mathbf{X}_{\text{high}} \quad (14)$$

up to numerical precision. These components mainly represent slowly varying background signals, local structural variations, and rapid disturbances, respectively.

The decomposition is applied independently to each finite historical input window of length L , not to the full observation record. No additional tapering window is used before the real FFT; the input window is therefore treated as a rectangular temporal segment. As a result, the reconstructed band components may contain finite-window boundary effects, especially near the first and last time steps of the input window. In this study, the same deterministic decomposition rule and fixed cutoffs are used for all samples, and only the historical input window is transformed. Therefore, the decomposition does not use future target information, while the possible endpoint artifacts are treated as a limitation of the finite-window frequency representation.

2.4.3. Branch-Specific Forecasting Network

After FFT-based decomposition, the original ABSC sequence is represented by three components with different temporal-scale characteristics. Before branch-specific processing, the height dimension is merged with the batch dimension:

$$\tilde{\mathbf{X}}_q = \text{Reshape}(\mathbf{X}_q) \in \mathbb{R}^{BC \times L}, \quad q \in \{\text{low}, \text{mid}, \text{high}\}. \quad (15)$$

The same branch parameters are therefore shared across height levels when forecasting their temporal evolution.

The branch assignment follows the statistical temporal-scale characteristics described in Section 2.2. The low-frequency component contains most of the spectral energy and mainly describes slowly varying background changes, so it is processed by lightweight linear mappings with pooling-based smoothing. The mid-frequency component contains local structural variations over shorter temporal ranges, so patching and convolution are used to preserve neighboring temporal context. The high-frequency component contains rapid variations and possible measurement noise, so a shallow branch is used to limit unnecessary model complexity. These assignments are intended to match the statistical properties of the decomposed time-scale components and are not treated as one-to-one mappings to specific atmospheric physical processes. The current branch-specific configuration should be regarded as one reasonable implementation of the frequency decomposition framework, rather than as a uniquely optimal design. Other branch designs, such as using the same architecture with different hyperparameters for different frequency bands, were not systematically explored in this study.

Low-Frequency Branch

The low-frequency component primarily represents the slowly varying background state of the ABSC sequence. A lightweight branch consisting of three linear mappings is used for this component because slowly varying components can often be represented by smooth temporal transformations without a high-capacity nonlinear module. Average pooling and layer normalization follow the first two mappings to form a smoothed bottleneck representation and stabilize training:

$$\mathbf{H}_1 = \text{LayerNorm}\left(\text{AvgPool}\left(\tilde{\mathbf{X}}_{\text{low}} \mathbf{W}_1 + \mathbf{b}_1\right)\right). \quad (16)$$

Here, $\mathbf{W}_1 \in \mathbb{R}^{L \times 4P}$ and $\mathbf{b}_1 \in \mathbb{R}^{4P}$. Average pooling reduces the last dimension to $2P$, giving $\mathbf{H}_1 \in \mathbb{R}^{BC \times 2P}$. The second stage is

$$\mathbf{H}_2 = \text{LayerNorm}(\text{AvgPool}(\mathbf{H}_1 \mathbf{W}_2 + \mathbf{b}_2)), \quad (17)$$

where $\mathbf{W}_2 \in \mathbb{R}^{2P \times P}$ and $\mathbf{b}_2 \in \mathbb{R}^P$. After pooling, $\mathbf{H}_2 \in \mathbb{R}^{BC \times (P/2)}$. All forecast horizons considered in this study contain an even number of time steps. The low-frequency output is

$$\mathbf{O}_{\text{low}} = \mathbf{H}_2 \mathbf{W}_3 + \mathbf{b}_3, \quad (18)$$

where $\mathbf{W}_3 \in \mathbb{R}^{(P/2) \times P}$ and $\mathbf{b}_3 \in \mathbb{R}^P$. Thus, $\mathbf{O}_{\text{low}} \in \mathbb{R}^{BC \times P}$.

Mid-Frequency Branch

The mid-frequency component mainly contains local temporal patterns and structural variations. To preserve the local context between neighboring time steps, the mid-frequency branch combines temporal patching with nonlinear convolutional processing. Given a patch length L_{patch} and stride S , the number of patches without end padding is

$$N_{\text{patch}} = \left\lfloor \frac{L - L_{\text{patch}}}{S} \right\rfloor + 1. \quad (19)$$

The resulting patch tensor is $\mathbf{X}_{\text{patch}} \in \mathbb{R}^{BC \times N_{\text{patch}} \times L_{\text{patch}}}$. Each patch is first projected through a linear layer, followed by a GELU activation and batch normalization:

$$\mathbf{E}_1 = \text{BatchNorm}\left(\text{GELU}\left(\mathbf{X}_{\text{patch}} \mathbf{W}_{m1} + \mathbf{b}_{m1}\right)\right). \quad (20)$$

Here, $\mathbf{W}_{m1} \in \mathbb{R}^{L_{\text{patch}} \times L_{\text{patch}}^2}$ and $\mathbf{b}_{m1} \in \mathbb{R}^{L_{\text{patch}}^2}$, yielding $\mathbf{E}_1 \in \mathbb{R}^{BC \times N_{\text{patch}} \times L_{\text{patch}}^2}$. A depth-wise convolution is then used to extract local features:

$$\mathbf{H}_{\text{conv}} = \text{BatchNorm}(\text{GELU}(\text{DepthwiseConv}(\mathbf{E}_1))). \quad (21)$$

A residual projection is introduced to retain information from the input representation:

$$\mathbf{H}_{\text{res}} = \mathbf{E}_1 \mathbf{W}_{m2} + \mathbf{b}_{m2}, \quad (22)$$

where $\mathbf{W}_{m2} \in \mathbb{R}^{L_{\text{patch}}^2 \times L_{\text{patch}}}$ and $\mathbf{b}_{m2} \in \mathbb{R}^{L_{\text{patch}}}$. Both paths produce tensors in $\mathbb{R}^{BC \times N_{\text{patch}} \times L_{\text{patch}}}$ and are combined by element-wise addition:

$$\mathbf{E}_2 = \mathbf{H}_{\text{conv}} + \mathbf{H}_{\text{res}}. \quad (23)$$

A pointwise convolution subsequently combines representations across patch positions:

$$\mathbf{E}_3 = \text{BatchNorm}(\text{GELU}(\text{PointwiseConv}(\mathbf{E}_2))), \quad (24)$$

where $\mathbf{E}_3 \in \mathbb{R}^{BC \times N_{\text{patch}} \times L_{\text{patch}}}$. The resulting features are flattened and mapped to the forecast horizon:

$$\mathbf{E}_{\text{flat}} = \text{Flatten}(\mathbf{E}_3) \in \mathbb{R}^{BC \times (N_{\text{patch}} L_{\text{patch}})}, \quad (25)$$

$$\mathbf{H}_m = \text{GELU}(\mathbf{E}_{\text{flat}} \mathbf{W}_{m3} + \mathbf{b}_{m3}), \quad \mathbf{O}_{\text{mid}} = \mathbf{H}_m \mathbf{W}_{m4} + \mathbf{b}_{m4}. \quad (26)$$

Here, $\mathbf{W}_{m3} \in \mathbb{R}^{(N_{\text{patch}} L_{\text{patch}}) \times 2P}$, $\mathbf{b}_{m3} \in \mathbb{R}^{2P}$, $\mathbf{W}_{m4} \in \mathbb{R}^{2P \times P}$, and $\mathbf{b}_{m4} \in \mathbb{R}^P$. Therefore, $\mathbf{O}_{\text{mid}} \in \mathbb{R}^{BC \times P}$.

High-Frequency Branch

The high-frequency component mainly contains rapid short-term variations and may also include high-frequency noise introduced by the measurement process. To reduce the

risk of fitting random fluctuations with an unnecessarily complex model, this branch uses a shallow multi-layer perceptron composed of two fully connected layers:

$$\mathbf{H}_{\text{high}} = \text{LayerNorm}\left(\text{GELU}\left(\tilde{\mathbf{X}}_{\text{high}}\mathbf{W}_{h1} + \mathbf{b}_{h1}\right)\right). \quad (27)$$

Here, $\mathbf{W}_{h1} \in \mathbb{R}^{L \times 2P}$ and $\mathbf{b}_{h1} \in \mathbb{R}^{2P}$, giving $\mathbf{H}_{\text{high}} \in \mathbb{R}^{BC \times 2P}$. The activation and normalization operations reduce the influence of extreme disturbances on the feature distribution. The high-frequency output is then obtained as

$$\mathbf{O}_{\text{high}} = \mathbf{H}_{\text{high}}\mathbf{W}_{h2} + \mathbf{b}_{h2}, \quad (28)$$

where $\mathbf{W}_{h2} \in \mathbb{R}^{2P \times P}$ and $\mathbf{b}_{h2} \in \mathbb{R}^P$. Therefore, $\mathbf{O}_{\text{high}} \in \mathbb{R}^{BC \times P}$.

2.4.4. Fusion and Output

The three branches produce $\mathbf{O}_{\text{low}}$, $\mathbf{O}_{\text{mid}}$, and $\mathbf{O}_{\text{high}}$. Their outputs are concatenated along the forecast-representation dimension:

$$\mathbf{O}_{\text{concat}} = \text{Concat}\left(\mathbf{O}_{\text{low}}, \mathbf{O}_{\text{mid}}, \mathbf{O}_{\text{high}}\right) \in \mathbb{R}^{BC \times 3P}. \quad (29)$$

Compared with element-wise addition, concatenation preserves the representation produced by each branch. A learnable linear mapping then combines information from the three frequency bands:

$$\mathbf{O}_{\text{final}} = \mathbf{O}_{\text{concat}}\mathbf{W}_{\text{out}} + \mathbf{b}_{\text{out}}, \quad (30)$$

where $\mathbf{W}_{\text{out}} \in \mathbb{R}^{3P \times P}$ and $\mathbf{b}_{\text{out}} \in \mathbb{R}^P$. Thus, $\mathbf{O}_{\text{final}} \in \mathbb{R}^{BC \times P}$. The output is reshaped and transposed to obtain $\hat{\mathbf{Y}} \in \mathbb{R}^{B \times P \times C}$, which contains the forecast ABSC profiles for the next P time steps. The fusion layer uses a fixed set of learned parameters and should not be interpreted as input-dependent attention weights.

2.5. Experimental Setup

2.5.1. Baseline Methods and Evaluation Scale

To evaluate FFT-FDNet on the lidar ABSC profile forecasting task, we compare it with Autoformer, FEDformer, NLinear, DLinear, PatchTST, LSTM, TCN, and CNN-LSTM. These baselines cover representative forecasting paradigms, including recurrent modeling, temporal convolution, linear prediction, series decomposition, autocorrelation mechanisms, frequency-enhanced attention, and patch-based representations [9,10]. Autoformer is implemented as an encoder–decoder model with series decomposition and the autocorrelation mechanism. FEDformer is implemented using the Fourier version with frequency-enhanced attention and series decomposition [23]. NLinear is included as a simple last-value-normalized linear baseline. It subtracts the last observed profile from the input sequence, applies a linear mapping from the input length to the forecast horizon, and adds the last observed profile back to the prediction.

All methods use the same dataset partitions, continuity-filtering rules, training set standardization parameters, input sequence lengths, forecast horizons, and random seeds. Model-specific configurations are selected according to validation performance, without using the test set for model selection.

The original ABSC observations are standardized separately at each height level using the mean and standard deviation calculated from the training set. Training, validation, and testing are performed on this standardized scale. Unless otherwise stated, MAE_{std} , MSE_{std} , and R^2 are calculated on the training set-standardized scale. MAE_{std} and MSE_{std}

are dimensionless and are used primarily for balanced relative comparisons among models across height levels, rather than as direct measures of errors in the physical units of ABSC.

2.5.2. Task Configuration and Implementation Details

This study focuses on short-term forecasting of ABSC profiles. Given the 15 min temporal resolution, forecast horizons of $P = 8, 16$, and 24 correspond to the next 2, 4, and 6 h, respectively. The 2 h task evaluates near-term profile evolution under strong short-term continuity. The 4 h task examines forecasting during adjustments in layered structures and local transport, while the 6 h task evaluates error accumulation and structural drift over an extended short-term horizon. These settings reflect short-term forecasting requirements in continuous lidar monitoring and avoid treating the rapidly varying ABSC task as long-term time-series forecasting.

The input sequence length is fixed at $L = 72$, corresponding to 18 h of historical observations. This window retains recent profile evolution while limiting the inclusion of older observations that are less closely related to the current state. The same setting is used in the main comparison, ablation study, and frequency cutoff experiments. Within each experiment, the input length is not adjusted separately for individual forecast horizons or test samples.

All models use the same dataset partition, standardization procedure, input/output construction, batch size, maximum number of training epochs, learning rate schedule, evaluation metrics, and training-loss type. Each experiment is independently repeated with random seeds 2021, 2022, and 2023. The mean and standard deviation over the three runs are reported. The common experimental settings are listed in Table 2.

Table 2. Common experimental settings for all models.

Parameter	Value
Input sequence length	72
Forecast horizons	8, 16, and 24 steps, corresponding to 2, 4, and 6 h
Dataset partition	Fixed training, validation, and test sets
Standardization	Per-height standardization using training set statistics
Continuity filtering	Input and forecast windows do not cross clear temporal discontinuities
Batch size	32
Initial learning rate	0.0005 for most models; 0.0001 for FEDformer
Learning rate schedule	Sigmoid
Maximum number of epochs	500
Early-stopping patience	7
Optimizer	AdamW for FFT-FDNet; Adam for the baseline implementations
Random seeds	2021, 2022, and 2023

The baseline configurations are fixed within each implementation and selected according to validation performance. The test set is not used for hyperparameter selection. The main baseline hyperparameters are listed in Table 3.

The model-specific configuration of FFT-FDNet is summarized in Table 4. The normalized-frequency cutoffs $(\nu_1, \nu_2) = (0.10, 0.45)$ are selected based on the spectral characteristics of the training set and the sensitivity experiments on the validation set. They remain fixed in all subsequent test set evaluations.

Table 3. Main hyperparameters of the baseline models.

Model	Main Hyperparameters
Autoformer	Model dimension: 256; attention heads: 4; encoder layers: 2; decoder layer: 1; feed-forward dimension: 1024; moving-average window: 4
FEDformer	Fourier version; model dimension: 256; attention heads: 4; encoder layers: 2; decoder layer: 1; feed-forward dimension: 1024; moving-average window: 4; Fourier modes: 32
NLinear	Last-value normalization; shared linear projection from the input length of 72 to the forecast horizon
DLinear	Moving-average window: 25; shared linear prediction heads for the trend and residual components
PatchTST	Model dimension: 256; encoder layers: 2; attention heads: 4; feed-forward dimension: 1024; patch length: 16; stride: 8; reversible instance normalization enabled
LSTM	Hidden dimension: 256; recurrent layers: 2; dropout: 0.05
TCN	Hidden channels: 256; residual blocks: 2; kernel size: 3; dilation rates: 1 and 2
CNN-LSTM	Two one-dimensional convolutional layers; kernel size: 3; LSTM hidden dimension: 256; recurrent layers: 2

Table 4. Main architecture and training parameters of FFT-FDNet.

Parameter	Value
Normalized-frequency cutoffs (ν_1, ν_2)	(0.10, 0.45)
Number of frequency branches	3
Patch length	16
Patch stride	8
Within-branch normalization	BatchNorm and LayerNorm
Activation function	GELU

2.5.3. Computing Environment

The main software and hardware configurations are summarized in Table 5. All methods use the same dataset partitions and training protocol, and their evaluation metrics are calculated from predictions on the test set.

Table 5. Main computing environment used in the experiments.

Resource	Configuration
Operating system	Ubuntu 22.04
Programming language	Python 3.11.15
Processor	Intel Core i5-14600K
Memory	64 GB
Deep-learning framework	PyTorch 2.7.0+cu128
GPU memory	16 GB
Storage	1 TB
GPU	NVIDIA GeForce RTX 5080

To provide a complete comparison of computational cost, we also measure the number of trainable parameters, the training time for one full epoch, and the inference time on the complete test loader for FFT-FDNet and all evaluated baseline models. The benchmark is conducted with $L = 72$, $P \in \{8, 16, 24\}$, and a batch size of 32 on the same GPU. The runtime values in Table 6 are averaged over the three forecast horizons.

Table 6. Computational cost of FFT-FDNet and all evaluated baseline models under the common setting.

Model	Trainable Parameters	Train/Epoch (s)	Inference (s)
FFT-FDNet	15.8–32.6 K	2.57	0.12
PatchTST	1.60–1.64 M	18.08	2.52
DLinear	1.2–3.5 K	0.41	0.05
NLinear	0.6–1.8 K	0.38	0.05
Autoformer	3.31 M	2.22	0.27
FEDformer	4.06–4.45 M	6.03	0.38
LSTM	1.61–2.70 M	0.98	0.10
CNN-LSTM	2.00–3.09 M	1.13	0.11
TCN	1.41–2.50 M	11.54	3.22

Note: Runtime is measured on the complete training or test loader and averaged over $P = 8, 16$, and 24 . The parameter count is reported as a range when it changes with the forecast horizon.

The computational cost comparison is used only to describe the practical tradeoff among the evaluated models. NLinear and DLinear have the lowest computational cost, but their forecasting accuracy is lower than that of FFT-FDNet in the main comparison. FFT-FDNet has more parameters and higher runtime than these two simple linear baselines. However, it uses far fewer trainable parameters than PatchTST, Autoformer, FEDformer, LSTM, CNN-LSTM, and TCN. Its training and inference times are also lower than those of PatchTST, FEDformer, and TCN under the same benchmark setting. This indicates that the accuracy improvement over PatchTST is not obtained at the cost of a larger model or longer runtime in the present experiment.

2.6. Evaluation Metrics

Predictions and observations from the test set are flattened across samples, forecast steps, and height levels. Let $\hat{y}_i$ and y_i denote the i th predicted and observed values, respectively, and let N be the total number of scalar observations. The mean absolute error (MAE) and mean squared error (MSE) are defined as

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (31)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (32)$$

The coefficient of determination (R^2) is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (33)$$

where $\bar{y} = N^{-1} \sum_{i=1}^N y_i$ denotes the mean of all observed values used in the calculation. Lower MAE and MSE values indicate smaller forecasting errors, whereas an R^2 value closer to one indicates that the predictions explain a larger proportion of the observed variability.

When y_i and $\hat{y}_i$ are expressed on the scale standardized using the training set, the corresponding error metrics are denoted by MAE_{std} and MSE_{std} . These dimensionless metrics are used for the main relative comparison because standardization reduces the dominance of height levels with larger ABSC magnitudes. After inverse standardization to the original physical scale of ABSC, the same definitions yield MAE_{phy} and MSE_{phy} , with $\text{RMSE}_{\text{phy}} = \sqrt{\text{MSE}_{\text{phy}}}$. MAE_{phy} and RMSE_{phy} are reported to interpret the absolute error magnitude in the original observational units.

3. Results

3.1. Overall Quantitative Comparison

Following the task configuration described in Section 2.5, the main comparison uses a fixed input sequence length of $L = 72$ and forecast horizons of $P \in \{8, 16, 24\}$. Each model is independently trained using three random seeds. Table 7 reports the mean and standard deviation of MAE_{std} , MSE_{std} , and R^2 as standardized-scale metrics for relative model comparison across the vertical profile.

Table 7. Comparison of forecasting performance at different forecast horizons.

Model	Forecast Horizon	$\text{MAE}_{\text{std}} \downarrow$	$\text{MSE}_{\text{std}} \downarrow$	$R^2 \uparrow$
Autoformer	8	0.4523 ± 0.0017	1.5878 ± 0.0263	-2.6558 ± 0.1060
	16	0.4862 ± 0.0019	1.9947 ± 0.0177	-1.8394 ± 0.0123
	24	0.4989 ± 0.0016	2.0926 ± 0.0191	-1.5505 ± 0.0082
CNN-LSTM	8	0.2762 ± 0.0023	1.3611 ± 0.0381	0.3083 ± 0.0245
	16	0.3185 ± 0.0014	1.7594 ± 0.0238	0.1056 ± 0.0083
	24	0.3287 ± 0.0013	1.8758 ± 0.0328	0.0543 ± 0.0174
DLinear	8	0.1808 ± 0.0024	0.7159 ± 0.0112	0.6000 ± 0.0067
	16	0.2605 ± 0.0026	1.2403 ± 0.0128	0.3456 ± 0.0084
	24	0.2911 ± 0.0010	1.4597 ± 0.0054	0.2294 ± 0.0039
FEDformer	8	0.4730 ± 0.0038	1.9333 ± 0.0103	-2.7514 ± 0.1039
	16	0.4908 ± 0.0027	2.0528 ± 0.0063	-1.8111 ± 0.0882
	24	0.5000 ± 0.0018	2.1054 ± 0.0044	-1.5770 ± 0.0409
FFT-FDNet	8	0.1221 ± 0.0017	0.4116 ± 0.0057	0.7505 ± 0.0041
	16	0.2108 ± 0.0011	1.0317 ± 0.0072	0.4844 ± 0.0028
	24	0.2509 ± 0.0006	1.3112 ± 0.0051	0.3435 ± 0.0014
LSTM	8	0.2922 ± 0.0029	1.3797 ± 0.0160	0.2261 ± 0.0299
	16	0.3218 ± 0.0016	1.7786 ± 0.0211	0.1086 ± 0.0046
	24	0.3297 ± 0.0014	1.8920 ± 0.0158	0.0502 ± 0.0151
NLinear	8	0.1934 ± 0.0028	0.8453 ± 0.0217	0.4998 ± 0.0132
	16	0.3070 ± 0.0026	1.5138 ± 0.0505	0.1004 ± 0.0219
	24	0.3648 ± 0.0026	1.7962 ± 0.0621	-0.0929 ± 0.0224
PatchTST	8	<u>0.1348 ± 0.0039</u>	<u>0.4740 ± 0.0212</u>	<u>0.7221 ± 0.0105</u>
	16	<u>0.2224 ± 0.0035</u>	<u>1.1094 ± 0.0193</u>	<u>0.4468 ± 0.0083</u>
	24	<u>0.2596 ± 0.0023</u>	<u>1.3841 ± 0.0159</u>	<u>0.3140 ± 0.0070</u>
TCN	8	0.2455 ± 0.0032	0.9965 ± 0.0120	0.4806 ± 0.0084
	16	0.2911 ± 0.0025	1.4115 ± 0.0151	0.3040 ± 0.0116
	24	0.3082 ± 0.0016	1.5792 ± 0.0154	0.2179 ± 0.0103

Note: Values are reported as the mean $\pm$ standard deviation over three runs with different random seeds. Forecast horizons of 8, 16, and 24 steps correspond to 2, 4, and 6 h, respectively. MAE_{std} and MSE_{std} are dimensionless metrics calculated on the scale standardized using the training set statistics and are used for relative comparison across height levels. Bold and underlined values indicate the best and second-best results, respectively, for the same forecast horizon. Downward and upward arrows indicate that lower and higher values are better, respectively.

As shown in Table 7, FFT-FDNet achieves the lowest mean MAE_{std} and MSE_{std} and the highest mean R^2 at all three forecast horizons. For $P = 8$, FFT-FDNet obtains an MAE_{std} of 0.1221, an MSE_{std} of 0.4116, and an R^2 of 0.7505. The corresponding results of the second-best method, PatchTST, are 0.1348, 0.4740, and 0.7221. Under the current Tokyo single-station setting, these results indicate that explicit frequency-domain decomposition can reduce forecasting errors in the 2-h task.

As P increases from 8 to 24, MAE_{std} and MSE_{std} generally increase for all methods, while R^2 decreases. This pattern indicates increasing forecasting uncertainty as the temporal

range extends and is consistent with the combined effects of layered-structure adjustments, local transport, and rapid disturbances on ABSC profiles. For $P = 24$, FFT-FDNet achieves an MAE_{std} of 0.2509, an MSE_{std} of 1.3112, and an R^2 of 0.3435, retaining the best mean values among the evaluated methods. This result supports its relative performance on the current 6 h task but should not be interpreted as evidence of long-horizon forecasting capability.

PatchTST ranks second for all three forecast horizons, indicating that patch-based representations are useful for capturing local continuity in ABSC sequences. Compared with PatchTST, FFT-FDNet reduces MAE_{std} by approximately 9.42%, 5.22%, and 3.35% at $P = 8, 16$, and 24, respectively. The corresponding MSE_{std} reductions are approximately 13.16%, 7.00%, and 5.27%. Thus, the advantage of FFT-FDNet is clearer for the 2 h task, while the difference becomes smaller at the 4 and 6 h horizons. This narrowing gap suggests that PatchTST remains a strong baseline and that forecasting uncertainty increases as the horizon extends. However, PatchTST does not explicitly separate temporal components from different frequency ranges, which may limit its ability to model mixed-scale dynamics using different structures. DLinear and NLinear generally perform better than LSTM and CNN-LSTM. This result suggests that part of the smooth variation in ABSC sequences can be represented using linear mappings, although a purely linear structure remains limited in its ability to capture local nonlinear disturbances.

The standard Autoformer and FEDformer obtain negative R^2 values at all three forecast horizons. This result does not indicate a problem with the ABSC data. It suggests a mismatch between these standard long-term forecasting architectures and the present short-term, high-dimensional ABSC profile forecasting task. Autoformer was mainly designed for long-term forecasting with clearer trend and periodic components, while FEDformer further introduces frequency-enhanced attention in a decomposed Transformer framework. In the current data, rapid layer changes, local disturbances, cloud or aerosol variability, and 265-dimensional profile outputs make stable trend, autocorrelation, and selected Fourier-mode patterns harder to learn. Therefore, the poor results of Autoformer and FEDformer are treated as evidence of model–task mismatch under the tested configuration, not as a general conclusion about these models. TCN performs better than LSTM and CNN-LSTM but remains less accurate than FFT-FDNet, PatchTST, and DLinear. Temporal convolution can capture part of the local dynamics, but it does not explicitly organize information from multiple frequency bands.

Overall, the performance of FFT-FDNet is consistent with the statistical structure of the ABSC data. Each time step contains a high-dimensional vertical profile, while the temporal sequence combines slowly varying background signals, local structural variations, and rapid disturbances. Frequency-domain decomposition assigns these mixed-scale components to separate forecasting branches, reducing interference that may arise when they are processed by a single structure. This interpretation is further examined through the subsequent ablation study rather than being based solely on the differences in the overall metrics.

To examine whether the results on the standardized scale remain consistent at the actual observational magnitude, we calculate MAE_{phy} and $RMSE_{phy}$ after transforming the predictions and observations back to the original ABSC scale. These physical-scale metrics are used to interpret the absolute error magnitude in units of ABSC. As shown in Table 8, FFT-FDNet achieves the lowest mean MAE_{phy} and $RMSE_{phy}$ at all three forecast horizons, while PatchTST remains the second-best method. The relative ranking is therefore consistent between the standardized and original physical scales.

Table 8. Forecasting errors of the evaluated models on the original physical scale of ABSC.

Model	$P = 8$		$P = 16$		$P = 24$	
	MAE _{phy} ↓	RMSE _{phy} ↓	MAE _{phy} ↓	RMSE _{phy} ↓	MAE _{phy} ↓	RMSE _{phy} ↓
Autoformer	3.188 ± 0.006	8.789 ± 0.062	3.400 ± 0.013	9.728 ± 0.057	3.480 ± 0.007	9.950 ± 0.029
CNN-LSTM	2.008 ± 0.028	8.266 ± 0.304	2.284 ± 0.015	9.329 ± 0.017	2.325 ± 0.010	9.478 ± 0.026
DLinear	1.283 ± 0.025	5.590 ± 0.070	1.851 ± 0.013	7.521 ± 0.030	2.074 ± 0.003	8.225 ± 0.019
FEDformer	3.306 ± 0.027	9.533 ± 0.020	3.435 ± 0.023	9.845 ± 0.012	3.503 ± 0.017	9.985 ± 0.003
LSTM	2.111 ± 0.045	8.232 ± 0.044	2.302 ± 0.003	9.326 ± 0.047	2.326 ± 0.006	9.485 ± 0.005
NLinear	1.378 ± 0.021	6.070 ± 0.088	2.172 ± 0.014	8.408 ± 0.038	2.585 ± 0.005	9.274 ± 0.132
PatchTST	<u>0.954 ± 0.049</u>	<u>4.565 ± 0.205</u>	<u>1.558 ± 0.013</u>	<u>7.158 ± 0.027</u>	<u>1.826 ± 0.015</u>	<u>8.117 ± 0.038</u>
TCN	1.753 ± 0.008	6.769 ± 0.017	2.086 ± 0.010	8.257 ± 0.066	2.184 ± 0.019	8.710 ± 0.034
FFT-FDNet	0.858 ± 0.011	4.194 ± 0.021	1.492 ± 0.007	6.858 ± 0.018	1.773 ± 0.003	7.829 ± 0.003

Note: Values are reported as the mean ± standard deviation over three runs with different random seeds. The input sequence length is fixed at 72. MAE_{phy} and RMSE_{phy} are calculated after inverse standardization to the original physical scale of ABSC. All values are expressed in units of $10^{-6} \text{ m}^{-1} \text{ sr}^{-1}$. Bold and underlined values indicate the best and second-best results, respectively, for the same forecast horizon and metric. The downward arrow indicates that lower values are better.

For $P = 8$, FFT-FDNet obtains an MAE_{phy} of $0.858 \times 10^{-6} \text{ m}^{-1} \text{ sr}^{-1}$ and an RMSE_{phy} of $4.194 \times 10^{-6} \text{ m}^{-1} \text{ sr}^{-1}$. These values are approximately 10.05% and 8.12% lower than those of PatchTST, respectively. This agrees with the standardized-scale results, for which FFT-FDNet reduces MAE_{std} and MSE_{std} by approximately 9.42% and 13.16%, respectively. Thus, the lower errors at the 2 h horizon are observed on both evaluation scales.

Errors on the physical scale also increase as the forecast horizon extends. For FFT-FDNet, MAE_{phy} increases from 0.858×10^{-6} to $1.773 \times 10^{-6} \text{ m}^{-1} \text{ sr}^{-1}$, while RMSE_{phy} increases from 4.194×10^{-6} to $7.829 \times 10^{-6} \text{ m}^{-1} \text{ sr}^{-1}$. This trend is consistent with the increase in MAE_{std} and MSE_{std} and the decrease in R^2 as P increases.

The difference between FFT-FDNet and PatchTST becomes smaller at the longer horizons. For $P = 16$, FFT-FDNet reduces MAE_{phy} and RMSE_{phy} by approximately 4.27% and 4.19%, respectively. For $P = 24$, the corresponding reductions are approximately 2.89% and 3.56%. A similar pattern appears on the standardized scale, where the reduction in MAE_{std} relative to PatchTST decreases from approximately 9.42% at $P = 8$ to 3.35% at $P = 24$. This narrowing difference indicates that uncertainty increases with the forecast horizon, although FFT-FDNet retains lower mean errors in the evaluated settings.

FFT-FDNet also obtains lower mean physical-scale errors than DLinear. At $P = 8, 16$, and 24, its MAE_{phy} is approximately 33.15%, 19.39%, and 14.53% lower than that of DLinear, respectively. The corresponding reductions in RMSE_{phy} are approximately 24.97%, 8.82%, and 4.82%. This direction of improvement is consistent with the standardized-scale results and suggests that linear mappings alone remain limited in representing local structural variations and nonlinear disturbances in the present task.

The physical-scale analysis complements the main quantitative comparison. Metrics on the standardized scale describe the balanced forecasting performance across multiple height levels, whereas metrics on the original physical scale quantify absolute errors at the observed ABSC magnitude. Both sets of metrics show that FFT-FDNet achieves the lowest mean errors at all three forecast horizons under the current data partition, indicating a consistent relative improvement across the two evaluation scales.

To further examine whether the forecasting errors are concentrated only at a few selected height levels, we calculate height-wise error profiles on the complete test set. For each model, MAE_{phy} and RMSE_{phy} are first calculated at each height level by averaging over all test samples and forecast steps, and the resulting profiles are then averaged over the three random seeds. Figure 5 shows the vertical distribution of physical-scale

errors for selected representative methods. In the Results Section, the tables provide the complete comparison of all evaluated models, whereas the diagnostic figures use selected representative methods for readability. Therefore, Figures 5–8 consistently show FFT-FDNet, PatchTST, DLinear, and NLinear, whose error ranges are closer and allow detailed differences to be observed more clearly.

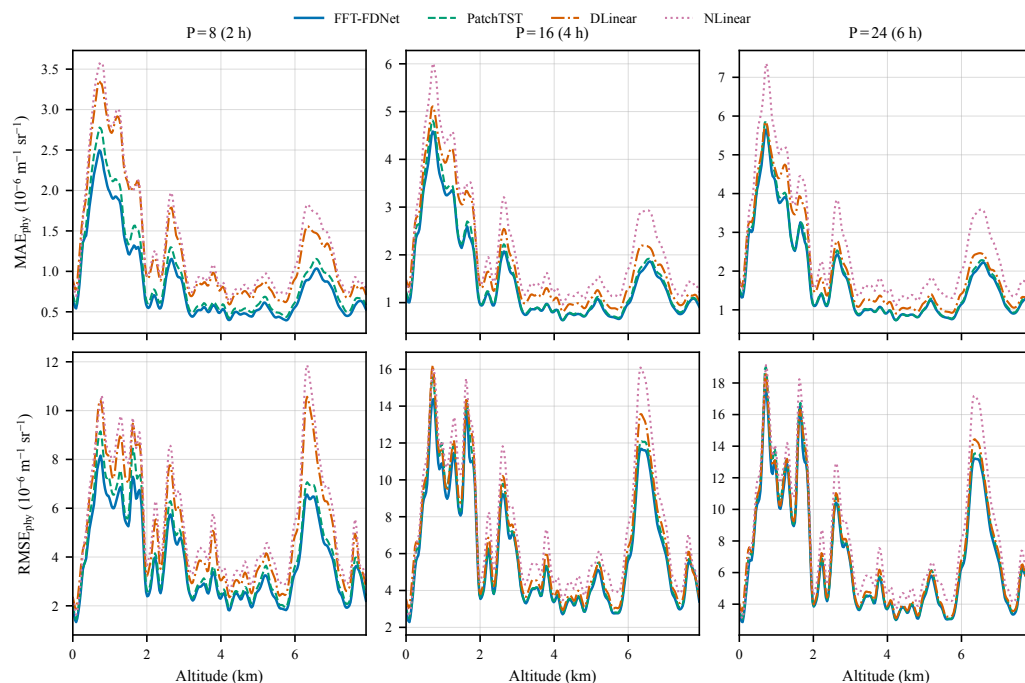


Figure 5. Height-wise MAE_{phy} and $RMSE_{phy}$ profiles on the complete test set for $P = 8, 16$, and 24 . The plotted methods are FFT-FDNet, PatchTST, DLinear, and NLinear. Errors are calculated after inverse standardization to the original ABSC scale and are reported in units of $10^{-6} m^{-1} sr^{-1}$. Each curve represents the mean profile over three random seeds.

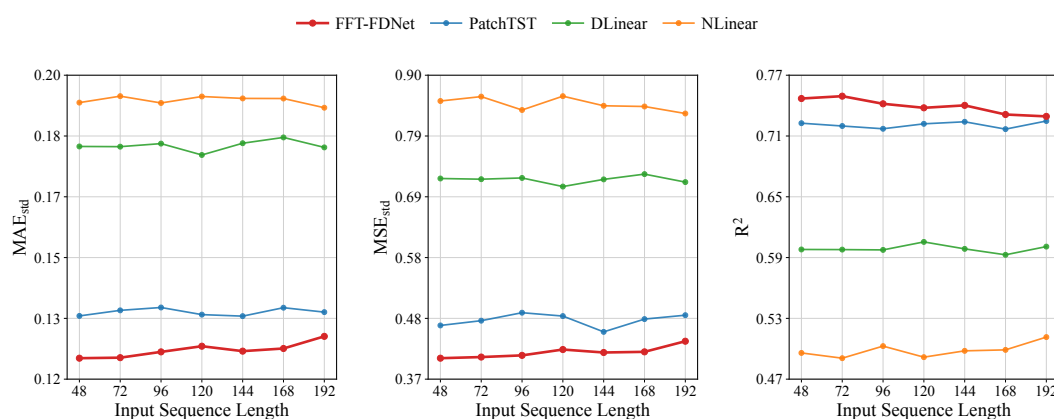


Figure 6. Forecasting performance at different input sequence lengths for $P = 8$ (2 h). The plotted methods are FFT-FDNet, PatchTST, DLinear, and NLinear. The horizontal axis is the input sequence length L , and the left, middle, and right panels show MAE_{std} , MSE_{std} , and R^2 , respectively.

As shown in Figure 5, the error magnitude varies with altitude for all models. Larger errors appear mainly in the lower troposphere and around several elevated layers, where ABSC values and structural changes are stronger. This pattern is consistent with the vertical profile nature of the data and indicates that the forecasting difficulty is not uniform across height. FFT-FDNet generally keeps lower height-wise MAE_{phy} than the competing methods across most altitude levels and forecast horizons. For $RMSE_{phy}$, the advantage is also

visible across the three horizons, although the differences among FFT-FDNet, PatchTST, and DLinear become smaller at some high-error layers, especially when the forecast horizon increases. These results support the overall metric comparison and show that the improvement of FFT-FDNet is not caused only by isolated height levels.

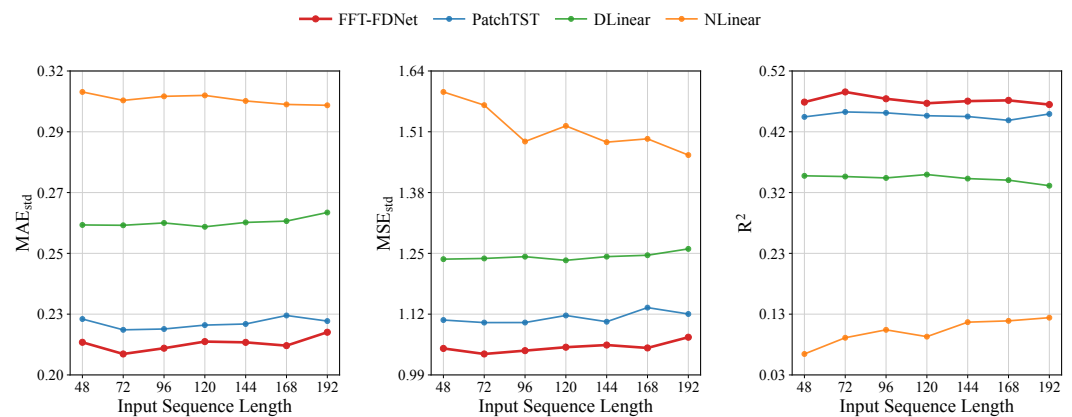


Figure 7. Forecasting performance at different input sequence lengths for $P = 16$ (4 h). The plotted methods are FFT-FDNet, PatchTST, DLinear, and NLinear. The horizontal axis is the input sequence length L , and the left, middle, and right panels show MAE_{std}, MSE_{std}, and R², respectively.

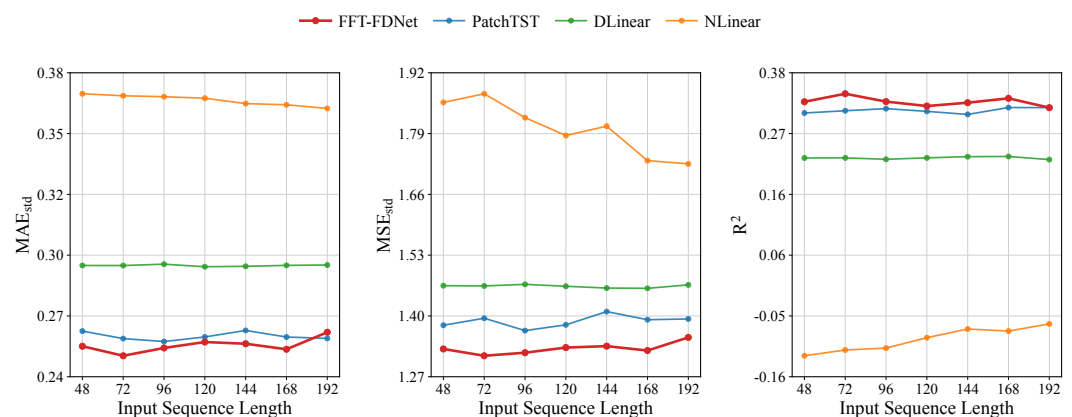


Figure 8. Forecasting performance at different input sequence lengths for $P = 24$ (6 h). The plotted methods are FFT-FDNet, PatchTST, DLinear, and NLinear. The horizontal axis is the input sequence length L , and the left, middle, and right panels show MAE_{std}, MSE_{std}, and R², respectively.

3.2. Ablation Study

To examine the effects of FFT-based frequency-domain decomposition and the individual frequency branches on ABSC profile forecasting, we conduct ablation experiments with a fixed input sequence length of 72 and forecast horizons of $P \in \{8, 16, 24\}$. FFT-FDNet denotes the complete model. FFT-FDNet *w/o* High, FFT-FDNet *w/o* Mid, and FFT-FDNet *w/o* Low remove the high-, mid-, and low-frequency branches, respectively. FDNet *w/o* FFT (same capacity) removes the FFT decomposition but keeps the same three-branch prediction structure by feeding the original input sequence into all three branches. This control is used to isolate the effect of the FFT decomposition from the effect of model capacity. In addition, the original single-branch no-FFT implementation is treated only as a time-domain single-branch baseline and is not used as the main evidence for the same-capacity ablation.

Table 9 presents the results of the ablation variants at the three forecast horizons. The complete FFT-FDNet and FFT-FDNet *w/o* High consistently occupy the top two positions. Removing the high-frequency branch produces the best mean values at $P = 8$, whereas the complete model achieves the best mean MAE_{std}, MSE_{std}, and R² at $P = 16$

and $P = 24$. In contrast, removing the low-frequency branch, the mid-frequency branch, or the FFT decomposition leads to a consistent deterioration in performance. These results support the contribution of explicit frequency-domain decomposition and complementary modeling of different frequency components.

Table 9. Ablation results at different forecast horizons.

Model	Forecast Horizon	MAE _{std} ↓	MSE _{std} ↓	R ² ↑
FFT-FDNet	8	<u>0.1221 ± 0.0017</u>	<u>0.4116 ± 0.0057</u>	<u>0.7505 ± 0.0041</u>
	16	0.2108 ± 0.0011	1.0317 ± 0.0072	0.4844 ± 0.0028
	24	0.2509 ± 0.0006	1.3112 ± 0.0051	0.3435 ± 0.0014
FFT-FDNet <i>w/o</i> Mid	8	0.1792 ± 0.0018	0.6393 ± 0.0072	0.5902 ± 0.0027
	16	0.2487 ± 0.0014	1.1737 ± 0.0037	0.3747 ± 0.0052
	24	0.2759 ± 0.0001	1.4011 ± 0.0018	0.2761 ± 0.0013
FFT-FDNet <i>w/o</i> High	8	0.1213 ± 0.0002	0.4092 ± 0.0034	0.7521 ± 0.0013
	16	<u>0.2115 ± 0.0017</u>	<u>1.0352 ± 0.0055</u>	<u>0.4819 ± 0.0051</u>
	24	<u>0.2512 ± 0.0004</u>	<u>1.3122 ± 0.0029</u>	<u>0.3426 ± 0.0010</u>
FFT-FDNet <i>w/o</i> Low	8	0.1895 ± 0.0033	0.5184 ± 0.0109	0.3295 ± 0.0292
	16	0.2610 ± 0.0010	1.1172 ± 0.0048	0.2565 ± 0.0044
	24	0.2898 ± 0.0006	1.3729 ± 0.0039	0.2028 ± 0.0056
FDNet <i>w/o</i> FFT (same capacity)	8	0.1287 ± 0.0014	0.4619 ± 0.0090	0.7328 ± 0.0023
	16	0.2184 ± 0.0010	1.0844 ± 0.0086	0.4594 ± 0.0044
	24	0.2586 ± 0.0015	1.3656 ± 0.0126	0.3226 ± 0.0032

Note: Values are reported as the mean ± standard deviation over three runs with different random seeds. Forecast horizons of 8, 16, and 24 steps correspond to 2, 4, and 6 h, respectively. MAE_{std} and MSE_{std} are calculated on the scale standardized using the training set statistics. Bold and underlined values indicate the best and second-best results, respectively, for the same forecast horizon. Downward and upward arrows indicate that lower and higher values are better, respectively.

At $P = 8$, FFT-FDNet *w/o* High obtains an MAE_{std} of 0.1213, an MSE_{std} of 0.4092, and an R² of 0.7521. Relative to the complete model, MAE_{std} and MSE_{std} are reduced by approximately 0.63% and 0.58%, while R² increases by 0.0016. These differences are small and do not establish a consistent advantage from removing the high-frequency branch. For the 2 h task, the model appears to rely mainly on low-frequency background information and mid-frequency structural variations, while the marginal contribution of the high-frequency branch to the average metrics is limited.

At $P = 16$ and $P = 24$, the complete model achieves better mean values for all three metrics. For example, at $P = 24$, the complete model obtains an MAE_{std} of 0.2509, an MSE_{std} of 1.3112, and an R² of 0.3435, compared with 0.2512, 1.3122, and 0.3426 for FFT-FDNet *w/o* High. The high-frequency branch therefore provides a small additional contribution at some forecast horizons but should not be regarded as the primary source of the overall performance improvement. This interpretation is consistent with the cutoff energy analysis in Section 3.3.2, where the selected cutoff pair assigns only a very small fraction of the training-window energy to the high-frequency band.

The low- and mid-frequency branches have a larger effect on forecasting performance. After removing the mid-frequency branch, MAE_{std} increases by 46.80%, 17.98%, and 9.95% at $P = 8$, 16, and 24, respectively. The corresponding increases in MSE_{std} are 55.31%, 13.76%, and 6.86%. These results support the interpretation that the mid-frequency component contains useful information about local structural variations.

Removing the low-frequency branch increases MAE_{std} by 55.24%, 23.81%, and 15.51% at $P = 8$, 16, and 24, respectively. The corresponding decreases in R² are 0.4211, 0.2279, and 0.1406. The low-frequency component therefore has a substantial influence on the

representation of the background baseline and the dominant temporal variation in the current task.

Removing the FFT decomposition also degrades all metrics at every forecast horizon, even when the same three-branch capacity is retained. Compared with the complete model, FDNet *w/o* FFT (same capacity) increases MAE_{std} by 3.04–5.42% and MSE_{std} by 4.15–12.21%, while reducing R^2 by 0.0178–0.0250. This result indicates that the performance gain is not only caused by using more branch capacity. Although the same-capacity time-domain network can still extract useful patterns from the original sequence, processing mixed-scale information without explicit frequency separation may increase interference among slowly varying signals, local structural variations, and rapid disturbances. These results support the role of explicit frequency-domain decomposition in separating components associated with different temporal scales.

In summary, the low- and mid-frequency branches have a substantial influence on the forecasting results, whereas the high-frequency branch provides a smaller contribution at some forecast horizons. All variants are evaluated using the same dataset partition, training protocol, and random seeds, allowing their relative effects to be compared under consistent conditions. The complete model shows a consistent improvement across all three metrics at $P = 16$ and $P = 24$, supporting the complementary roles of different frequency components in ABSC profile forecasting.

3.3. Parameter Sensitivity Analysis

To examine the effects of key parameter choices on forecasting performance, we analyze the input sequence length and the frequency band cutoffs used in the FFT decomposition. The input length determines how much recent historical information is available to the model, while the frequency cutoffs control how temporal components are assigned to the low-, mid-, and high-frequency branches. All other training settings are kept unchanged during the sensitivity experiments. The frequency cutoffs are selected using the spectral characteristics of the training set and the validation results; after they are fixed, the test set is used only for final performance evaluation.

3.3.1. Input Sequence Length

To evaluate sensitivity to the amount of historical context, we vary the input sequence length L while keeping the model architecture, training strategy, and forecast horizon unchanged. Performance is evaluated using MAE_{std} , MSE_{std} , and R^2 . A short input window may not capture the continuous evolution of layered structures, whereas an excessively long window may introduce historical information that is less closely related to the current state.

Figures 6–8 show how FFT-FDNet, PatchTST, DLinear, and NLinear respond to changes in input length at the three forecast horizons. Consistent with Figure 5, these diagnostic figures show selected representative methods with closer error ranges for readability. The complete quantitative comparison of all evaluated models, including Autoformer, FEDformer, LSTM, CNN-LSTM, and TCN, is provided in Tables 7 and 8. This separation avoids expanding the vertical axis in the diagnostic figures and keeps the detailed differences among the main comparable methods visible. Increasing the amount of historical context does not produce a monotonic improvement. Short windows may omit part of the recent structural evolution, while long windows may include observations that are weakly related to the current state. FFT-FDNet maintains relatively favorable performance across most input lengths and forecast horizons, indicating a degree of robustness to changes in the historical window.

At $P = 8$, the overall errors are lower than those at the longer horizons, which is consistent with the strong continuity of the dominant ABSC structures over 2 h. FFT-FDNet achieves the lowest MAE_{std} and MSE_{std} and the highest R^2 at every evaluated input length, with particularly favorable results around $L = 48$ and $L = 72$. Its errors increase slightly at $L = 168$ and $L = 192$, indicating that a longer historical window does not provide a consistent additional benefit.

At $P = 16$, FFT-FDNet achieves the best mean values across the evaluated input lengths. Its lowest MAE_{std} and MSE_{std} and highest R^2 occur at $L = 72$. Under the current data and experimental settings, 18 h of historical observations therefore provide an effective representation of recent profile evolution. PatchTST, DLinear, and NLinear also perform favorably at some input lengths, but their mean metrics remain less favorable overall than those of FFT-FDNet.

At $P = 24$, the errors of the four plotted methods increase further and their R^2 values generally decrease. FFT-FDNet achieves relatively favorable mean values at most input lengths. At $L = 72$, it obtains an MAE_{std} of 0.2509, an MSE_{std} of 1.3112, and an R^2 of 0.3435. At $L = 192$, the differences between FFT-FDNet and PatchTST narrow for some metrics, again showing that an excessively long input window does not yield a consistent improvement.

Across the three forecast horizons, $L = 72$ provides a favorable balance, while further increases in input length do not consistently improve performance. This pattern suggests that short-term forecasting under the current setting depends more strongly on recent profile states and their evolution. The input length sensitivity experiment is used to characterize model behavior and is not treated as independent evidence of generalization.

3.3.2. Frequency Cutoff Selection and Sensitivity

FFT-FDNet uses two normalized-frequency cutoffs, ν_1 and ν_2 , to separate the input sequence into low-, mid-, and high-frequency components. The cutoff values directly affect how information from different temporal scales is assigned to the forecasting branches. To avoid selecting these values based on test performance, we use only the training and validation sets. Candidate ranges are first identified from the mean spectral energy distribution of the training data. Different cutoff combinations are then evaluated on the validation set, and the setting with favorable and relatively stable validation performance is selected. The test set does not participate in this selection process.

Let Δt denote the sampling interval and let ν denote normalized frequency in cycles/sample. The corresponding physical frequency is

$$f(\nu) = \frac{\nu}{\Delta t}, \quad (34)$$

and its approximate temporal period is

$$\tau(\nu) = \frac{1}{f(\nu)} = \frac{\Delta t}{\nu}. \quad (35)$$

For the 15 min sampling interval, $\Delta t = 0.25$ h. With $\nu_1 = 0.10$ and $\nu_2 = 0.45$, the corresponding periods are

$$\tau(0.10) = \frac{0.25}{0.10} = 2.5 \text{ h}, \quad (36)$$

$$\tau(0.45) = \frac{0.25}{0.45} \approx 0.56 \text{ h}. \quad (37)$$

The low-frequency component, $\nu \leq 0.10$, mainly represents slowly varying signals with periods of at least 2.5 h. The mid-frequency component, $0.10 < \nu \leq 0.45$, represents local structural variations over approximately 0.56–2.5 h. The high-frequency component, $\nu > 0.45$, contains rapid variations near the Nyquist frequency and is more likely to reflect local short-duration disturbances, brief strong-scattering changes, and observational noise. These temporal ranges provide a statistical interpretation of the frequency components and are not assumed to correspond one-to-one with specific atmospheric processes.

To evaluate sensitivity to the frequency cutoffs, we fix the input sequence length at 72 and consider five combinations: (0.10, 0.45), (0.02, 0.10), (0.05, 0.20), (0.10, 0.30), and (0.05, 0.40). Each combination is evaluated at $P \in \{8, 16, 24\}$. All other model and training settings remain unchanged, and each result is calculated from three runs with different random seeds.

Table 10 reports the corresponding band energy fractions calculated from the standardized training input windows. The values are obtained using the same one-dimensional real FFT and the same frequency masks as those used in FFT-FDNet.

Table 10. Training set energy fractions under different frequency cutoff combinations.

(ν_1, ν_2)	Low-Frequency (%)	Mid-Frequency (%)	High-Frequency (%)
(0.10, 0.45)	97.19	2.74	0.07
(0.02, 0.10)	54.81	42.38	2.81
(0.05, 0.20)	82.94	16.43	0.63
(0.10, 0.30)	97.19	2.48	0.33
(0.05, 0.40)	82.94	16.90	0.15

Note: Energy fractions are calculated from all valid standardized training input windows with $L = 72$. The one-sided real FFT spectrum is weighted according to Parseval's relation before summing the band energy.

As shown in Table 11, $(\nu_1, \nu_2) = (0.10, 0.45)$ achieves the best mean values for all metrics at the three forecast horizons. At $P = 8$, this setting obtains an MAE_{std} of 0.1225, an MSE_{std} of 0.4125, and an R^2 of 0.7499. At $P = 24$, the corresponding values are 0.2511, 1.3113, and 0.3434. As shown in Table 10, this setting assigns 97.19% of the training-window energy to the low-frequency band, 2.74% to the mid-frequency band, and only 0.07% to the high-frequency band, which is the smallest high-frequency fraction among the evaluated cutoff pairs. This result indicates that most predictable information in the current ABSC sequence is concentrated in slowly varying components, while the part near the Nyquist frequency is very limited. Therefore, the selected cutoff pair gives a conservative high-frequency allocation. It keeps most slowly varying and structurally meaningful energy in the low-frequency branch, while leaving only a small residual part to the high-frequency branch. This supports the interpretation that the high-frequency branch has a modest and horizon-dependent role rather than being the main source of improvement.

The (0.10, 0.30) setting ranks second at all three forecast horizons. At $P = 16$, its MAE_{std} and MSE_{std} are 0.2142 and 1.0508, approximately 1.61% and 1.83% higher than those obtained with (0.10, 0.45), while R^2 is lower by 0.0076. The two settings use the same ν_1 , but reducing ν_2 from 0.45 to 0.30 narrows the mid-frequency range and increases the high-frequency energy fraction from 0.07% to 0.33%. This means that more components with approximate temporal scales of 0.56–0.83 h are assigned to the high-frequency branch. The resulting performance change suggests that this range still contains useful local structural information and should not be assigned too early to the high-frequency branch.

The (0.02, 0.10) setting produces the least favorable results at all three forecast horizons. At $P = 8$, its MAE_{std} and MSE_{std} are approximately 12.49% and 24.36% higher than those of the best setting, while R^2 is lower by 0.0382. In this configuration, $\nu_1 = 0.02$ corresponds

to variations with periods of approximately 12.5 h or longer, and $\nu_2 = 0.10$ corresponds to a period of approximately 2.5 h. The low-frequency energy fraction decreases from 97.19% to 54.81%, while the mid-frequency and high-frequency fractions increase to 42.38% and 2.81%, respectively. This redistribution indicates that many slowly varying or structurally meaningful components are no longer processed by the low-frequency branch. The intermediate results of (0.05, 0.20) and (0.05, 0.40) further show that the cutoffs affect the information received by the individual branches.

Table 11. Validation set sensitivity results for different frequency cutoff combinations.

(ν_1, ν_2)	Forecast Horizon	MAE _{std} ↓	MSE _{std} ↓	R ² ↑
(0.10, 0.45)	8	0.1225 ± 0.0023	0.4125 ± 0.0084	0.7499 ± 0.0055
	16	0.2108 ± 0.0010	1.0319 ± 0.0038	0.4841 ± 0.0073
	24	0.2511 ± 0.0006	1.3113 ± 0.0051	0.3434 ± 0.0017
(0.02, 0.10)	8	0.1378 ± 0.0015	0.5130 ± 0.0073	0.7117 ± 0.0040
	16	0.2240 ± 0.0002	1.0939 ± 0.0006	0.4469 ± 0.0008
	24	0.2621 ± 0.0003	1.3618 ± 0.0016	0.3123 ± 0.0020
(0.05, 0.20)	8	0.1313 ± 0.0012	0.4821 ± 0.0093	0.7292 ± 0.0023
	16	0.2192 ± 0.0003	1.0742 ± 0.0006	0.4629 ± 0.0025
	24	0.2567 ± 0.0003	1.3393 ± 0.0027	0.3284 ± 0.0007
(0.10, 0.30)	8	<u>0.1275 ± 0.0015</u>	<u>0.4456 ± 0.0047</u>	<u>0.7383 ± 0.0029</u>
	16	<u>0.2142 ± 0.0001</u>	<u>1.0508 ± 0.0015</u>	<u>0.4765 ± 0.0005</u>
	24	<u>0.2534 ± 0.0007</u>	<u>1.3222 ± 0.0025</u>	<u>0.3363 ± 0.0040</u>
(0.05, 0.40)	8	0.1316 ± 0.0008	0.4621 ± 0.0054	0.7243 ± 0.0020
	16	0.2172 ± 0.0003	1.0671 ± 0.0034	0.4642 ± 0.0005
	24	0.2566 ± 0.0014	1.3447 ± 0.0072	0.3264 ± 0.0047

Note: Values are reported as the mean ± standard deviation over three runs with different random seeds. Forecast horizons of 8, 16, and 24 steps correspond to 2, 4, and 6 h, respectively. MAE_{std} and MSE_{std} are calculated on the scale standardized using the training set statistics. Bold and underlined values indicate the best and second-best results, respectively, for the same forecast horizon. Downward and upward arrows indicate that lower and higher values are better, respectively.

Based on the validation results across the three forecast horizons, we use $(\nu_1, \nu_2) = (0.10, 0.45)$ as the common frequency band configuration for FFT-FDNet. This setting is kept unchanged in the main quantitative comparison, input length analysis, and ablation study. The results in Table 11 are used only for cutoff selection and sensitivity analysis on the validation set. After the model configuration is fixed, the test set is used only to evaluate generalization performance. The cutoffs are therefore neither selected retrospectively from test results nor adjusted separately for individual forecast horizons.

3.4. Qualitative Analysis

To complement the quantitative evaluation with an examination of vertical-layer preservation and temporal continuity, we select the same test sample, sample_00952, and visualize the forecasts at $P = 8, 16$, and 24 , corresponding to the next 2, 4, and 6 h, respectively. This sample contains several clearly defined high-value layers whose intensity and structure change during the forecast period. It is therefore suitable for illustrating differences among the evaluated methods, but it is not assumed to represent all test samples.

The three forecasting tasks use the same starting point for the historical input. The ground-truth sequences for $P = 8$ and $P = 16$ correspond to the first 8 and 16 steps of the $P = 24$ ground-truth sequence, respectively. This arrangement allows the forecast deviations to be examined as the horizon increases. In Figures 9–11, the panels show the ground-truth, FFT-FDNet, PatchTST, DLinear, and NLinear in that order. A common color scale is used across the three figures. These qualitative examples illustrate structural

differences for the selected sample and do not replace the overall quantitative evaluation on the test set.

As shown in Figure 9, the ground-truth sequence contains clear high-value layers at altitudes of approximately 1.8, 2.7, and 4.1 km. Weaker transient layers are also present at higher altitudes. At the 2 h horizon, all methods identify the main layers, but differences appear in their intensity, thickness, and rate of decay. FFT-FDNet reproduces the continuous layered structures near 1.8 and 2.7 km and captures the gradual weakening of the high-value region near 4.1 km as the forecast progresses. PatchTST preserves the locations of the main layers, although their local boundaries are smoother and some horizontal banding is visible. DLinear and NLinear tend to extend the initially strong layers into later forecast steps, resulting in a slower decay than that observed in the ground-truth sequence.

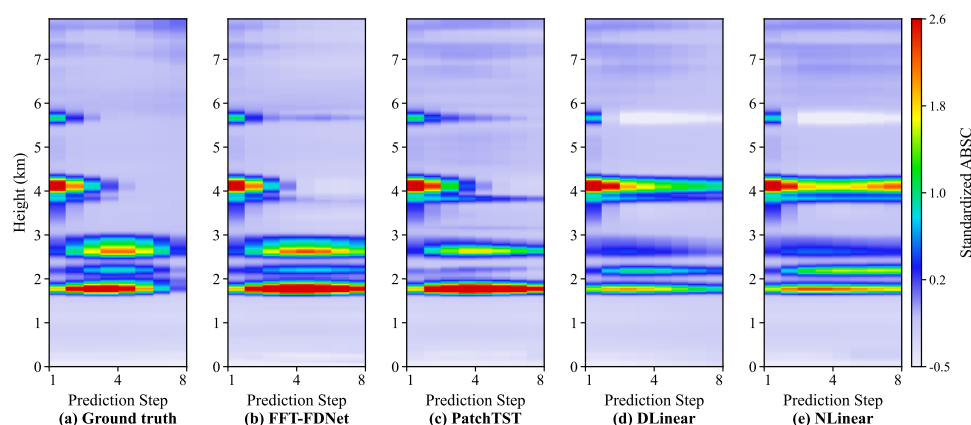


Figure 9. Standardized ABSC profile forecasts for sample_00952 at $P = 8$ (2 h) using random seed 2022. The horizontal axis is the prediction step, the vertical axis is height in km, and the colorbar shows standardized ABSC. The same color limits are used for all panels in this figure and are set from the 1st–99th percentiles of the ground-truth sample. Because standardization is applied separately at each height level, the color-to-physical-scale conversion is height-dependent; physical-scale errors are reported in Table 8 and Figure 5. Panel order: (a) ground-truth; (b) FFT-FDNet; (c) PatchTST; (d) DLinear; and (e) NLinear.

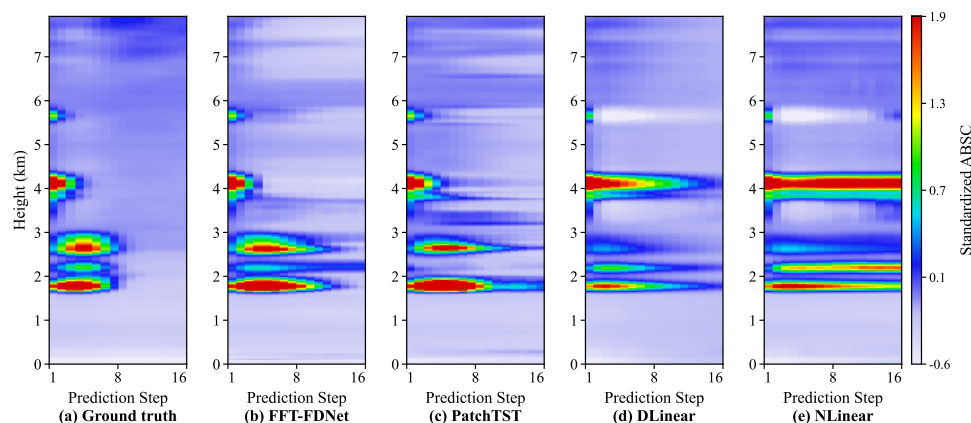


Figure 10. Standardized ABSC profile forecasts for sample_00952 at $P = 16$ (4 h) using random seed 2022. The horizontal axis is the prediction step, the vertical axis is height in km, and the colorbar shows standardized ABSC. The same color limits are used for all panels in this figure and are set from the 1st–99th percentiles of the ground-truth sample. Because standardization is applied separately at each height level, the color-to-physical-scale conversion is height dependent; physical-scale errors are reported in Table 8 and Figure 5. Panel order: (a) ground-truth; (b) FFT-FDNet; (c) PatchTST; (d) DLinear; and (e) NLinear.

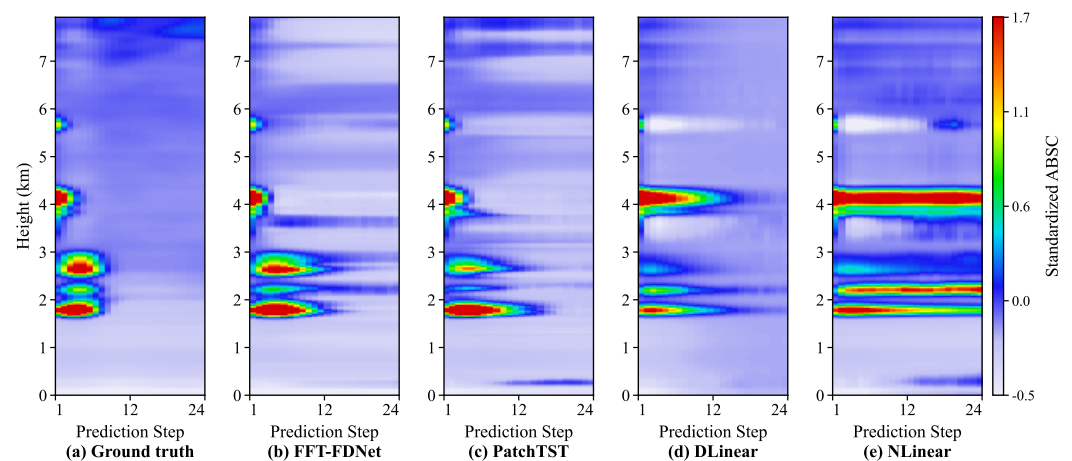


Figure 11. Standardized ABSC profile forecasts for sample_00952 at $P = 24$ (6 h) using random seed 2022. The horizontal axis is the prediction step, the vertical axis is height in km, and the colorbar shows standardized ABSC. The same color limits are used for all panels in this figure and are set from the 1st–99th percentiles of the ground-truth sample. Because standardization is applied separately at each height level, the color-to-physical-scale conversion is height dependent; physical-scale errors are reported in Table 8 and Figure 5. Panel order: (a) ground-truth; (b) FFT-FDNet; (c) PatchTST; (d) DLinear; and (e) NLinear.

At $P = 16$, the main high-value layers gradually weaken after the first several forecast steps, while the weaker layers also change over time. As shown in Figure 10, FFT-FDNet preserves the relative positions of the multiple layers and allows the high-value regions to contract as the forecast progresses. PatchTST provides a reasonable reconstruction of the main layers near 1.8 and 2.7 km but exhibits more pronounced smoothing and banding in the middle and upper parts of the profile. DLinear and NLinear maintain the high-value layer near 4.1 km for longer than observed in the ground-truth. For this sample, trend extrapolation or a simple linear mapping alone does not fully capture the rapid decay of the layered structure.

At the longer horizon of $P = 24$, the structural differences among the methods become more apparent. Figure 11 shows that the high-value layers in the ground-truth sequence are concentrated mainly in the early forecast steps and then weaken rapidly. FFT-FDNet preserves the approximate altitude positions of the main layers and produces a temporal decay pattern similar to the ground truth. PatchTST continues to recover the dominant layers, but some structures are extended with stronger smoothing. The horizontal high-value bands generated by DLinear and NLinear persist into later forecast steps, indicating a weaker response to rapid structural changes. The behavior observed in this sample is consistent with the intended role of explicit three-band decomposition and branch-specific modeling. However, the qualitative results are used only to illustrate differences among the methods for this particular sample.

LSTM, CNN-LSTM, TCN, the standard Autoformer, and FEDformer are not included in the qualitative comparison because their overall quantitative performance is lower across the three forecast horizons. The visualization focuses on PatchTST, DLinear, and NLinear, whose quantitative results are closer to those of FFT-FDNet. This selection keeps the color-scale comparison readable and avoids obscuring the structural differences among the main comparable methods. The overall conclusions remain based on the test set statistics obtained from three random seeds. All three visualizations use random seed 2022 to maintain comparability across forecast horizons.

4. Discussion

The results show that FFT-FDNet achieves favorable mean performance at the 2, 4, and 6 h forecast horizons under the current Tokyo single-station dataset and chronological data partition. The ablation results indicate that the low- and mid-frequency branches account for most of the performance improvement, whereas the high-frequency branch provides only a small contribution at some horizons. As the forecast horizon increases, the errors of all models increase and their performance differences narrow, reflecting greater uncertainty in future ABSC states. The height-wise error profiles further show that the advantage of FFT-FDNet becomes modest at the longer horizon of $P = 24$ in several layers, and the RMSE profiles of the representative methods tend to converge in the upper troposphere. This suggests that the benefit of frequency decomposition may diminish when the forecast horizon becomes longer and the backscatter signal is weaker.

Because the experiments are based on one AD-Net station and one observation period from February to June 2023, the results should be interpreted as a single-site case study. They should not be regarded as evidence of general performance across different lidar stations, seasons, or aerosol regimes. Seasonal aerosol conditions, cloud occurrence, local emissions, transport patterns, and site characteristics may affect the temporal-frequency structure of ABSC sequences. Therefore, cross-season, cross-year, and multi-station validation is needed before the conclusions are extended to broader lidar forecasting applications.

5. Conclusions

This study proposed FFT-FDNet for short-term forecasting of lidar ABSC profiles and evaluated it as a Tokyo single-station case study. The model separates the input sequence along the temporal dimension into low-, mid-, and high-frequency components. These components are processed using a lightweight linear structure, a patch-based convolutional structure, and a shallow multi-layer perceptron, respectively. Their outputs are subsequently combined through a learnable linear fusion layer to produce future ABSC profiles.

Within this single-site Tokyo setting, FFT-FDNet achieved the best average performance among the evaluated methods under the current data partition and the 2, 4, and 6 h forecasting settings. The standardized-scale and physical-scale metrics exhibit a consistent direction of relative improvement. The ablation results show consistent performance degradation when the FFT decomposition, low-frequency branch, or mid-frequency branch is removed. The high-frequency branch provides a smaller additional contribution at some forecast horizons. The input length experiments characterize the response of the model to different amounts of historical context, while the frequency cutoff experiments support the use of $(\nu_1, \nu_2) = (0.10, 0.45)$ within the evaluated candidate range.

These findings are limited to the current single-site dataset and experimental configuration. They should not be directly generalized to other lidar stations, seasons, climatic regions, or aerosol conditions.

This study has two main limitations. First, the experiments use data from a single station and a single observation period. The results may therefore be affected by the local climatic conditions and sample distribution. Second, the model relies only on historical ABSC profiles and does not explicitly incorporate external drivers such as wind fields, temperature, humidity, or boundary-layer height. Future work will extend the evaluation to cross-season, cross-year, and multi-station AD-Net observations. Leave-station-out or cross-station tests will also be considered when suitable data are available. Reanalysis meteorological fields and spatial-neighborhood information will also be considered to assess model generalization under different weather processes and regional transport conditions.

Author Contributions: Conceptualization, H.C. and Z.Z.; methodology, Z.Z.; software, Z.Z.; validation, H.C. and Z.Z.; formal analysis, H.C.; investigation, H.C.; resources, J.L.; data curation, Z.R.; writing—original draft preparation, Z.Z.; writing—review and editing, H.C.; visualization, H.C.; supervision, F.G.; project administration, F.G.; funding acquisition, H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (NSFC) under Grant 61805194, Grant 42375137 and Grant 42465007, and by the Key Program of the NSFC under Grant 42530112.

Data Availability Statement: The 532 nm attenuated backscatter coefficient data used in this study are publicly available from the Asian Dust and Aerosol Lidar Observation Network (AD-Net) Tokyo station, operated by the National Institute for Environmental Studies, Japan. The data can be accessed at <https://www-lidar.nies.go.jp/AD-Net/> (accessed on 30 June 2026).

Acknowledgments: Lidar data were provided by courtesy of AD-Net (<https://www-lidar.nies.go.jp/AD-Net> (accessed on 30 June 2026)).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABSC	Attenuated backscatter coefficient
AD-Net	Asian Dust and Aerosol Lidar Observation Network
CNN	Convolutional neural network
FFT	Fast Fourier transform
LSTM	Long short-term memory
MAE	Mean absolute error
MSE	Mean squared error
RMSE	Root mean squared error
TCN	Temporal convolutional network

References

- Li, D.; Wu, Y.; Gross, B.; Moshary, F. Capabilities of an Automatic Lidar Ceilometer to Retrieve Aerosol Characteristics within the Planetary Boundary Layer. *Remote Sens.* **2021**, *13*, 3626. [CrossRef]
- Li, M.; Cao, X.; Zhang, Z.; Ji, H.; Zhang, M.; Guo, Y.; Tian, P.; Liang, J. Optical Properties and Vertical Distribution of Aerosols Using Polarization Lidar and Sun Photometer over Lanzhou Suburb in Northwest China. *Remote Sens.* **2023**, *15*, 4927. [CrossRef]
- Kotthaus, S.; Bravo-Aranda, J.A.; Collaud Coen, M.; Guerrero-Rascado, J.L.; Costa, M.J.; Cimini, D.; O'Connor, E.J.; Hervo, M.; Alados-Arboledas, L.; Jimenez-Portaz, M.; et al. Atmospheric Boundary Layer Height from Ground-Based Remote Sensing: A Review of Capabilities and Limitations. *Atmos. Meas. Tech.* **2023**, *16*, 433–479. [CrossRef]
- Chazette, P.; Totems, J. Lidar Profiling of Aerosol Vertical Distribution in the Urbanized French Alpine Valley of Annecy and Impact of a Saharan Dust Transport Event. *Remote Sens.* **2023**, *15*, 1070. [CrossRef]
- Liu, X.; Alves Gouveia, D.; Henzing, B.; Apituley, A.; Hensen, A.; van Dinter, D.; Huang, R.; Dusek, U. Aerosol Optical Properties within the Atmospheric Boundary Layer Predicted from Ground-Based Observations Compared to Raman Lidar Retrievals during RITA-2021. *Atmos. Chem. Phys.* **2024**, *24*, 9597–9614. [CrossRef]
- Jing, D.; He, Y.; Yin, Z.; Huang, K.; Liu, F.; Yi, F. Evolution of Tropospheric Aerosols over Central China during 2010–2024 as Observed by Lidar. *Atmos. Chem. Phys.* **2025**, *25*, 17047–17067. [CrossRef]
- Fu, Y.; Zhu, Z.; Liu, L.; Zhan, W.; He, T.; Shen, H.; Zhao, J.; Liu, Y.; Zhang, H.; Liu, Z.; et al. Remote Sensing Time Series Analysis: A Review of Data and Applications. *J. Remote Sens.* **2024**, *4*, 0285. [CrossRef]
- Miller, L.; Pelletier, C.; Webb, G.I. Deep Learning for Satellite Image Time-Series Analysis: A Review. *IEEE Geosci. Remote Sens. Mag.* **2024**, *12*, 81–124. [CrossRef]
- Lim, B.; Zohren, S. Time-Series Forecasting with Deep Learning: A Survey. *Philos. Trans. R. Soc. A* **2021**, *379*, 20200209. [CrossRef] [PubMed]
- Moskolai, W.R.; Abdou, W.; Dipanda, A.; Kolyang. Application of Deep Learning Architectures for Satellite Image Time Series Prediction: A Review. *Remote Sens.* **2021**, *13*, 4822. [CrossRef]

11. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [\[CrossRef\]](#)
12. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
13. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. In Proceedings of the International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
14. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? *Proc. AAAI Conf. Artif. Intell.* **2023**, *37*, 11121–11128. [\[CrossRef\]](#)
15. National Institute for Environmental Studies. Asian Dust and Aerosol Lidar Observation Network (AD-Net). Available online: <https://www-lidar.nies.go.jp/AD-Net/> (accessed on 30 June 2026).
16. Lolli, S. Machine Learning Techniques for Vertical Lidar-Based Detection, Characterization and Classification of Aerosols and Clouds: A Comprehensive Survey. *Remote Sens.* **2023**, *15*, 4318. [\[CrossRef\]](#)
17. Yin, Z.; Wang, L.; He, Y.; Wang, X. Understanding the Roles of Aerosols and Clouds in Environment, Meteorology and Climate with Advanced Lidar Remote Sensing Techniques. *Remote Sens.* **2024**, *16*, 593. [\[CrossRef\]](#)
18. Rieutord, T.; Aubert, S.; Machado, T. Deriving Boundary Layer Height from Aerosol Lidar Using Machine Learning: KABL and ADABL Algorithms. *Atmos. Meas. Tech.* **2021**, *14*, 4335–4353. [\[CrossRef\]](#)
19. Wijnands, J.S.; Apituley, A.; Alves Gouveia, D.; Noteboom, J.W. Deep-Pathfinder: A Boundary Layer Height Detection Algorithm Based on Image Segmentation. *Atmos. Meas. Tech.* **2024**, *17*, 3029–3045. [\[CrossRef\]](#)
20. Zhang, H.; Huang, W.; Shen, X.; Ramisetty, R.; Song, J.; Kiseleva, O.; Holst, C.C.; Khan, B.; Leisner, T.; Saathoff, H. Aerosol Composition, Air Quality, and Boundary Layer Dynamics in the Urban Background of Stuttgart in Winter. *Atmos. Chem. Phys.* **2024**, *24*, 10617–10637. [\[CrossRef\]](#)
21. Zhang, X.; Zheng, Y.; Che, H.; Gui, K.; Li, L.; Zhao, H.; Liang, Y.; Yao, W.; Zhang, X.; Zhao, H.; et al. Seasonal and Diurnal Characteristics of the Vertical Profile of Aerosol Optical Properties in Urban Beijing, 2017–2021. *Remote Sens.* **2023**, *15*, 475. [\[CrossRef\]](#)
22. Xie, H.; Wang, Z.; Luo, T.; Yang, K.; Zhang, D.; Zhou, T.; Yang, X.; Liu, X.; Fu, Q. Seasonal Variation of Dust Aerosol Vertical Distribution in Arctic Based on Polarized Micropulse Lidar Measurement. *Remote Sens.* **2022**, *14*, 5581. [\[CrossRef\]](#)
23. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; Volume 162, pp. 27268–27286.
24. Yi, K.; Zhang, Q.; Fan, W.; Wang, S.; Wang, P.; He, H.; An, N.; Lian, D.; Cao, L.; Niu, Z. Frequency-Domain MLPs Are More Effective Learners in Time Series Forecasting. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 76656–76679. [\[CrossRef\]](#)
25. Fan, W.; Yi, K.; Ye, H.; Ning, Z.; Zhang, Q.; An, N. Deep Frequency Derivative Learning for Non-Stationary Time Series Forecasting. In Proceedings of the 33rd International Joint Conference on Artificial Intelligence, Jeju, Republic of Korea, 3–9 August 2024; pp. 3944–3952. [\[CrossRef\]](#) [\[PubMed\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.