Wind Energy

WILEY

RESEARCH ARTICLE OPEN ACCESS

A Chebyshev Polynomial-Based Wind Speed Profile Characterization Framework: Applications in Mesoscale Model Evaluation

Harish Baki[1] | Sukanta Basu[1,2]

[1]Atmospheric Sciences Research Center, University at Albany, Albany, New York, USA | [2]Department of Environmental and Sustainable Engineering, University at Albany, Albany, New York, USA

Correspondence: Harish Baki (hbaki@albany.edu)

Received: 11 April 2025 | **Revised:** 20 November 2025 | **Accepted:** 24 November 2025

Keywords: Chebyshev polynomial | low-level jet | wind speed profile

ABSTRACT

This study presents a polynomial-based wind speed profile characterization framework that approximates full vertical profiles using five physically meaningful Chebyshev polynomial coefficients. The framework captures key morphological features of wind profiles, including mean wind speed, vertical shear, curvature, inflection-related behavior, and higher order fluctuations. Its accuracy is demonstrated using well-documented wind regimes such as well-mixed, shear-dominant, logarithmic-shaped, and low-level jet (LLJ) profiles, showing consistent performance across diverse atmospheric conditions. To evaluate its practical utility by moving beyond conventional error metrics, the framework was applied to compare simulated wind profiles from a mesoscale model-generated dataset against lidar observations. Results show that the mesoscale model reliably captured mean and shear structures, especially at coastal sites, while underrepresenting curvature and higher order variations at inland locations. The framework also facilitated a spatial and temporal assessment of model behavior, distinguishing coastal from inland site characteristics and capturing diurnal and seasonal variability, including LLJs driven by sea-breeze circulation.

1 | Introduction

Wind energy has undergone significant global expansion over the past decade as individuals, businesses, and governments pursue decarbonization targets in response to climate change, consistent with the objectives of the Paris Agreement. This trend is evidenced by the global installed capacity reaching 945.5 GW onshore and 75.2 GW offshore as of 2023 [1]. A key aspect of deploying wind power systems at any scale is the accurate assessment of available wind resources and a clear understanding of prevailing wind conditions, both of which are essential for optimal system design and performance [2].

Modern turbines have grown substantially in scale. The tallest onshore turbine to date, the SANY SI-270150 15 MW, has a hub height of 143 m and a rotor diameter of 270 m, reaching a total height of more than 275 m [3]. Offshore turbines are even larger; for instance, the MySE18.X 20 MW turbine features a 160-m hub height and a 292-m rotor diameter, extending up to 306 m [4]. These turbines penetrate deep into the atmospheric boundary layer (ABL) and experience varying wind speed and shear conditions across the rotor plane. Moreover, Akhtar et al. [5] demonstrated that the influence of tall turbines can extend up to and beyond 500 m into the ABL. Similarly, Krishnamurthy et al. [6] showed that low-level jets (LLJs) occurring between 250 and 500 m, contribute to increased momentum transfer and enhanced wake recovery. These studies emphasize that the effects of modern wind turbines span the full depth of the ABL, underscoring the need for accurate wind speed and wind shear information throughout the rotor-swept layer and beyond.

Remote sensing instruments, such as Doppler lidars and sodars, can measure wind profiles (horizontal wind speed at multiple vertical levels) within and above the ABL, often reaching heights beyond 500 m [7–13]. Occasionally, tall metmasts can also reach heights beyond 300 m [14]. However, multiyear site measurements are costly, and commercial developers typically collect data for only a few years, often with significant gaps.

A practical and potentially low-cost alternative is wind data generated by Numerical Weather Prediction (NWP) models. Several organizations worldwide have developed long-term wind profile datasets using NWP models. These datasets typically span multiple decades and altitudes reaching several hundred meters, with temporal resolutions ranging from hourly to every 6 h [15]. Notable examples include the fifth generation of ECMWF atmospheric reanalyses (ERA5) [16], copernicus regional reanalysis for Europe (CERRA) [17], New European Wind Atlas (NEWA) [18, 19], Met Éireann high-resolution reanalysis for Ireland (MÉRA) [20], Dutch Offshore Wind Atlas (DOWA) [21], Norwegian Hindcast archive (NORA3) [22], and the 2023 National Offshore Wind data set (NOW23) [23]. Not just for long-term wind resource assessment, the NWP models were also used in simulating wind profiles for understanding key meteorological phenomena, such as nocturnal LLJs [24], land-sea breeze circulations [25], frontal low-level jets (FLLJs) [26], and coastal low-level jets (CLLJ) [27, 28], to name a few, that contribute to improving design and operation of wind turbines.

Since Lorenz's early work on atmospheric predictability [29], it has been well-established that NWP models are highly sensitive to inaccuracies in their initial and boundary conditions [30], as well as to limitations in their physical parameterizations. Al-Yahyai et al. [31] and references therein further emphasize that this sensitivity extends across various modeling configurations, particularly in the context of wind speed estimation. Therefore, evaluating NWP model-simulated wind profiles is essential.

The majority of the studies in the wind energy literature have evaluated the performance of NWP models using conventional verification methods, primarily by comparing simulated wind speeds against observations at discrete height levels. Quantitative evaluations typically rely on metrics such as bias, root-mean-square error (RMSE), centered RMSE (cRMSE), mean absolute error (MAE), Pearson's correlation coefficient r, and Earth movers distance (EMD), computed at individual heights. For qualitative assessment, visual comparison of vertical profiles of mean wind speed or error metrics is commonly employed. Some of these studies are summarized in Table 1, illustrating the widespread use of height-based evaluation methods across various models, datasets, and validation periods.

Some researchers evaluated the NWP model-simulated wind speed by comparing wind speed shear exponent (α) [32, 42], which plays a critical role in turbine loading and power output. Krogsaeter et al. [43] noted modeling sensitivity in capturing α computed between 40- and 100-m levels, along with deviation from a nominal design value of 0.1, in stable conditions. Holstlag et al. [44] further noted that an inaccurate wind shear profile will result in an overestimation of turbine blade root fatigue loads by a factor of 2. Mata et al. [45] showed that an overestimation of 31% in wind power estimation is possible if one considers a turbine power curve for a given hub height, without any information about wind shear, in an onshore utility-scale wind turbine. Given the fact that wind turbines are becoming bigger and taller, complex wind shear can occur within the rotor area, such as positive and negative shear occurring simultaneously over the rotor area. If one considered just two levels across the rotor-plane in estimating the wind shear exponent and evaluate with respect to observations, it could substantially overestimate/underestimate the actual shear.

Based on these studies, it is evident that the shape of the wind speed profile is more important, as it could offer the characteristics of wind speed shear profile and occurrences of local wind speed maxima. Thus, evaluating the wind speed profiles by quantifying the difference between shapes of the profiles is necessary, rather than evaluating wind speed at specific height levels. There has been a study from Durán et al. [46], who proposed an approach for automated classification of wind profiles using machine learning-based self-organizing maps (SOM). The methodology can be employed to categorize simulated and observed wind profiles and qualitatively evaluate their difference, based on the “patterns” or “shapes”.

In the current study, we propose a different methodology that can quantitatively differentiate two wind profile datasets in the lowest 500 m of our atmosphere, along with providing characteristics of the wind profile, such as the mean of the wind profile, wind speed shear, curvature, occurrence of LLJs, and higher order fluctuations. In turn, this approach can quantify whether NWP models tend to produce specific weather phenomena at a specific geographical location. Such findings are highly valuable for model developers, offering insights into biases associated with parameterization schemes, initial and boundary conditions, domain configurations, and so on.

The heart of our methodology lies in approximating wind speed profiles using a set of polynomials. In the wind energy community, the Legendre polynomials [47–49] and Chebyshev polynomials [50–53] have been used in this regard. For example, Otte et al. [47] used the Legendre polynomials in developing a new single column PBL parameterization scheme, while Jonkman et al. [52] used the Chebyshev polynomials to model an LLJ wind profile for the TurbSim module. In the present work, we propose a novel **C**hebyshev polynomials-based **Wi**nd **S**peed **Pr**ofile characteriation (C-WiSPr) framework, which offers a physically interpretable framework for evaluating wind speed profiles that goes beyond traditional height-based metrics and addresses the increasing complexity of wind resource characterization in the age of tall turbines. We provide an in-depth technical background of the framework in Section 2.

In order to represent the NWP-simulated wind profiles, we used the publicly available NOW23 dataset, which has been generated using the Weather Research and Forecasting (WRF) model by the National Renewable Energy Lab (NREL). For evaluation purposes, we obtained the New York State Mesonet (NYSM) lidar observed wind profiles data. More details of the datasets are provided in Section 3. The primary goal of this study is to compare the NYSM measured and the NOW23-simulated wind speed vertical profiles using the C-WiSPr framework. This framework quantifies the shape differences between the observed and

TABLE 1 | Summary of wind profile evaluation studies in the literature. Quantitative metrics are typically computed at individual height levels, while qualitative evaluations are based on vertical profile comparisons.

Study	Model(s)	Focus	Quantitative Evaluation	Qualitative Evaluation
Storm et al. [32]	WRF	Sensitivity to PBL and radiation schemes	Wind shear exponent computed between 10 m and 127 m heights	Visual comparison of full wind profile shape
Floors et al. [33]	WRF	Sensitivity to surface roughness, PBL schemes, forcing data, and vertical resolution during coastal flow regimes	Bias, R^2 at 10, 100, 650 m	Visual comparison of full wind profile shape
Nunalee et al. [27]	WRF	Sensitivity to PBL schemes, initial and boundary conditions, and vertical resolution	Mean wind speed, wind power density, annual wind power, and Weibull parameters computed at turbine hub height (179 m)	Visual comparison of wind profile shape
Gryning et al. [34]	WRF	Comparison of wind profile simulations in analysis and forecasting	Bias, normalized bias, slope coefficient, RMSE, and Weibull parameters computed at 100, 300, and 500 m heights	Visual comparison of vertical profiles of mean wind speed and Weibull shape parameter
Gevorgyan et al. [35]	WRF	Simulating topographically-induced LLJs		Visual comparison of LLJ structure in profiles
Kalverla et al. [36]	ECMWF-IFS, HARMONIE-AROME, WRF	Model intercomparison over 30 diverse cases	Bias, cRMSE, at 27, 115, 215, and 315 m; wind shear between lowest and top most levels	Visual inspection of wind speed profiles
Hahmann et al. [19]	WRF (NEWA)	Best configuration for pan-European wind atlas	Bias, RMSE, EMD at various levels up to 200+ m	Mean wind profile comparisons
Brune et al. [37]	ERA5, MERRA2, COSMO-REA6	Wind energy potential across Central Europe	Bias, RMSE, Pearson's r at multiple levels up to 200 m	Visual inspection vertical profiles of mean wind speed and evaluation metrics
Pronk et al. [38]	WRF, ERA5	Annual offshore vs onshore wind comparison	Mean, bias, cRMSE, r, EMD, std. dev. at multiple levels	Vertical profiles of wind and metric values
Sward et al. [39]	WRF Ensemble	Offshore wind simulation across seasons	Bias, RMSE at multiple levels	Mean profile comparison, profile-averaged RMSE
Pentikäinen et al. [40]	ECMWF-IFS	Forecasting wind profiles with global lidars	MAE of wind vector within 117-1000 m	Diurnal and seasonal MAE variation
Bodini et al. [23]	WRF	Offshore wind resource development for US	Bias, cRMSE, r^2 at individual levels, rotor equivalent wind speed (REWS) error	Vertical profiles of wind metrics and REWS
Cheynet et al. [41]	NORA3, NEWA, ERA5	Validation of tall wind resource datasets	Bias, RMSE, r^2, EMD at individual heights heights up to 500 m	Visual comparison of full error profiles

simulated profiles using Chebyshev polynomial coefficients. In doing so, the method captures key physical features such as vertical shear, curvature, and jet-like structures, offering novel metrics for model evaluation that complement and extend traditional height-based error assessments. In Section 4, we have reported a quantitative comparison between the NWP model-simulated and observed wind profile dataset. Lastly, concluding remarks and further applications of the proposed methodology are mentioned in Section 5.

2 | Wind Speed Profile Characterization Using Chebyshev Polynomials

In this section, we introduce the C-WiSPr framework and provide more insights with the help of some well-documented cases.

2.1 | Chebyshev Polynomial Approximation

Chebyshev polynomials are a set of orthogonal functions, which have the capability to approximate any given function, with minimum truncation error, due the minimax property [54]. The Chebyshev polynomials of the first kind are defined in the interval of [−1,1] and can be estimated using a simple recurrence relation as

$$T_0(z) = 1, \tag{1}$$

$$T_1(z) = z, \tag{2}$$

$$T_{n+1}(z) = 2zT_n(z) - T_{n-1}(z). \tag{3}$$

Here, T_0 is the 0^{th} order, T_1 is the 1^{st} order, and T_n is the n^{th} order polynomials, which are functions of the independent variable z.

Following the recurrence relation, polynomials up to the fourth-order are given below and are illustrated in Figure 1a.

$$T_2(z) = 2z^2 - 1 \tag{4}$$

$$T_3(z) = 4z^3 - 3z \tag{5}$$

$$T_4(z) = 8z^4 - 8z^2 + 1 \tag{6}$$

Once knowing these polynomials, any function can be approximated as, $U(z) = \sum_{i=0}^{n} C_i T_i(z)$, where C_i is the Chebyshev coefficient of i^{th} order, while U represents wind speed as a function of height z. Since the polynomial approximation is applicable within the limits of [−1, 1], the height coordinate [0 m, 500 m] is normalized between these limits. With the polynomials in hand, the Chebyshev coefficients can be estimated by solving a system of linear equations, provided wind speed profiles (wind speed as function of height).

The product of Chebyshev polynomials and the corresponding coefficients are known as projections (e.g., $C_k T_k(z)$), represent individual modes of the expansion, with each mode contributing a specific structural component to the overall wind speed profile. The coefficient C_k governs the amplitude of its corresponding mode. Figure 1b,c illustrates the individual projections and their cumulative contribution to the total expansion, for a representative set of Chebyshev coefficients.

The zeroth-order projection $C_0 T_0(z)$ represents a constant mode, establishing the baseline of the expansion (Figure 1b). The coefficient C_0 dictates the magnitude of this baseline and serves as an approximation of the mean of the wind speed profile. Around this baseline, higher order projections introduce fluctuations, as shown in the cumulative contributions in Figure 1c.

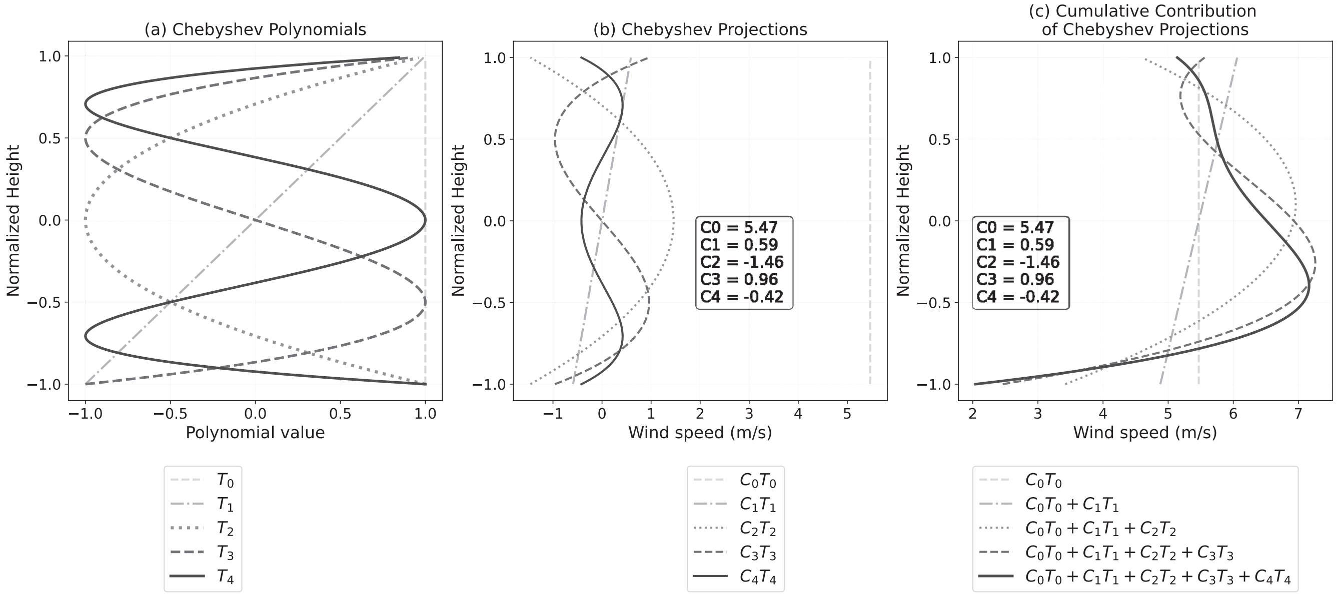


FIGURE 1 | (a) An illustration of the Chebyshev polynomials of order 4: T_0, T_1, T_2, T_3, and T_4. The Chebyshev projections for a set of Chebyshev coefficients are illustrated in (b), while (c) illustrates the cumulative contribution of each projection. The set of Chebyshev coefficients are obtained from an observed wind speed profile, which is infact the final outcome of cumulative contributions of all projections (black line in (c)).

The first-order projection $C_1T_1(z)$ adds a linear trend (Figure 1b), contributing to the overall shear in the wind speed profile (Figure 1c). A positive C_1 indicates increasing wind speed with height, while a negative C_1 indicates a decreasing trend. Similarly, the second-order projection $C_2T_2(z)$ introduces quadratic curvature to the expansion, shaping the convexity or concavity of the profile. The magnitude of the curvature is dictated by the coefficient C_2.

Higher order projections, such as $C_3T_3(z)$ and $C_4T_4(z)$, add cubic and quartic structures, respectively (Figure 1b). These terms enhance the ability of the expansion to capture complex vertical features, such as localized shear gradients, inflection points, and sharp curvature changes. In particular, C_3 contributes to both curvature and linear shear, as seen in the overall wind speed difference across the domain: $U(1) - U(-1) = 2C_1 + 2C_3$.

Importantly, the inclusion of C_3 and C_4 can shift the position of the local wind speed maximum (jet nose) downward in the profile (Figure 1c). This behavior is critical for realistically representing LLJs, where the jet peak often occurs in the lower part of the ABL. Thus, while the first three modes (C_0 to C_2) capture the bulk structure of the wind profile, such as mean, linear shear, and curvature, higher order projections are essential for accurately modeling localized features.

The following section offers more details regarding the Chebyshev coefficients governing specific characteristics of wind speed profiles with the help of some well-documented cases.

2.2 | Well-Documented Wind Speed Profiles

To better understand the contribution of individual and pairwise combinations of Chebyshev coefficients, we approximated wind speed profiles from well-documented Leipzig [55] and Wangara campaigns [56], and lidar observations at Høvsøre [57] and Wantagh [58]. These profiles are shown in Figure 2, along with the Chebyshev approximations, exemplify diverse atmospheric conditions.

The famous "Leipzig wind profile" shown in Figure 2a was measured by Mildner more than 90 years ago using pilot balloons. The profile resembles a shear-dominant structure, with winds increasing with height, and showing insignificant curvature. In this case, the coefficient C_1 has a large positive value, indicating strong vertical shear, while C_2 is relatively low and negative. The higher order coefficients contribute minimally. Although the profile appears to be a pure-shear case with monotonically increasing winds, wind shear itself decreases with height, which is effectively captured by the low negative C_2 value.

In contrast, Wangara-1 (Figure 2b) and Høvsøre-9 (Figure 2n) display nearly uniform wind speeds with height, appearing almost as vertical lines. These profiles were observed during turbulent afternoon periods. Both are primarily characterized by the coefficient C_0, which closely matches the mean wind speeds (5.32 m/s for Wangara-1, with $C_0 = 5.00$, and 5.03 m/s for Høvsøre-9, with $C_0 = 4.98$). The remaining coefficients are relatively low, confirming the well-mixed and nonsheared nature of these profiles. These cases show that under well-mixed conditions, the wind profile can be sufficiently represented by C_0 alone.

Several profiles exhibit LLJ characteristics, with the wind speed maximum (or jet nose) occurring between 150 and 250 m. Wangara-2 (Figure 2c) was recorded at night under stably stratified conditions with weak turbulent mixing. These conditions result in atmospheric decoupling and the development of a low-level wind maximum, driven by inertial oscillation. A similar behavior is observed in Høvsøre-1 (Figure 2f), recorded in the early morning during very stable surface conditions and low atmospheric forcing. This LLJ also stems from inertial oscillation but appears during its diminishing phase. Wantagh-1 and Wantagh-2 (Figure 2d–e) show coastal LLJ features observed in the late evening over New York Bight. These jets are linked to sea breeze circulation during warm-season conditions, triggered by strong temperature gradients caused by coastal upwelling. These profiles are well-represented by Chebyshev polynomials, with significant contributions from C_0, C_2, and C_3, while C_1 and C_4 contribute marginally. Here, C_2 majorly contributes to the overall curvature, while C_3 is responsible for the inflection point along with reduced jet nose height, positive overall shear, and increased shear near the ground. The low C_1 accounts for fraction of overall shear, while the minor role of C_4 reflects minor fluctuations at higher levels (400–500 m).

Høvsøre-3 (Figure 2h) was observed during midnight hours (2200–0130) and also shows a LLJ structure. The profile has strong near-surface shear and a wind speed maximum at around 400 m. This LLJ is associated with inertial oscillation [57]. The high C_1 indicates strong shear, and the high negative C_2 captures the jet curvature. On the other hand, negative C_3 contributes to reduced overall shear and increased jet height, without any sign of inflection point. Notably, the combination of high C_1 and large negative C_2 tends to produce LLJs with jet maxima located above 250 m.

Høvsøre-2 (Figure 2g) was recorded in the early morning (0600–0730) under neutral to stable surface conditions. The profile shows a wind speed maximum at 400 m, followed by a decrease in wind speed above. This deceleration is attributed to baroclinic effects acting on the u component. The strong wind shear is reflected by a high C_1, and the curvature around 400 m is captured by a high negative C_2.

The Høvsøre-4 profile (Figure 2i) was measured in the morning under stable surface conditions and high atmospheric forcing. This combination produces a profile with strong wind shear (high C_1) and a high mean wind speed (15.57 m/s), resulting in a high C_0 value. The profile's curvature is also evident in a moderate negative C_2.

Høvsøre-5 (Figure 2j) and Høvsøre-7 (Figure 2l) were observed from afternoon to evening under neutral surface conditions. Both show noticeable surface-layer shear (moderately high C_1) and moderate negative C_2, indicating curvature. The higher order coefficients C_3 and C_4 are minimal, suggesting a lack of inflection or fine-scale structure. According to Peña et al. [57], these profiles resemble the classical logarithmic wind profile

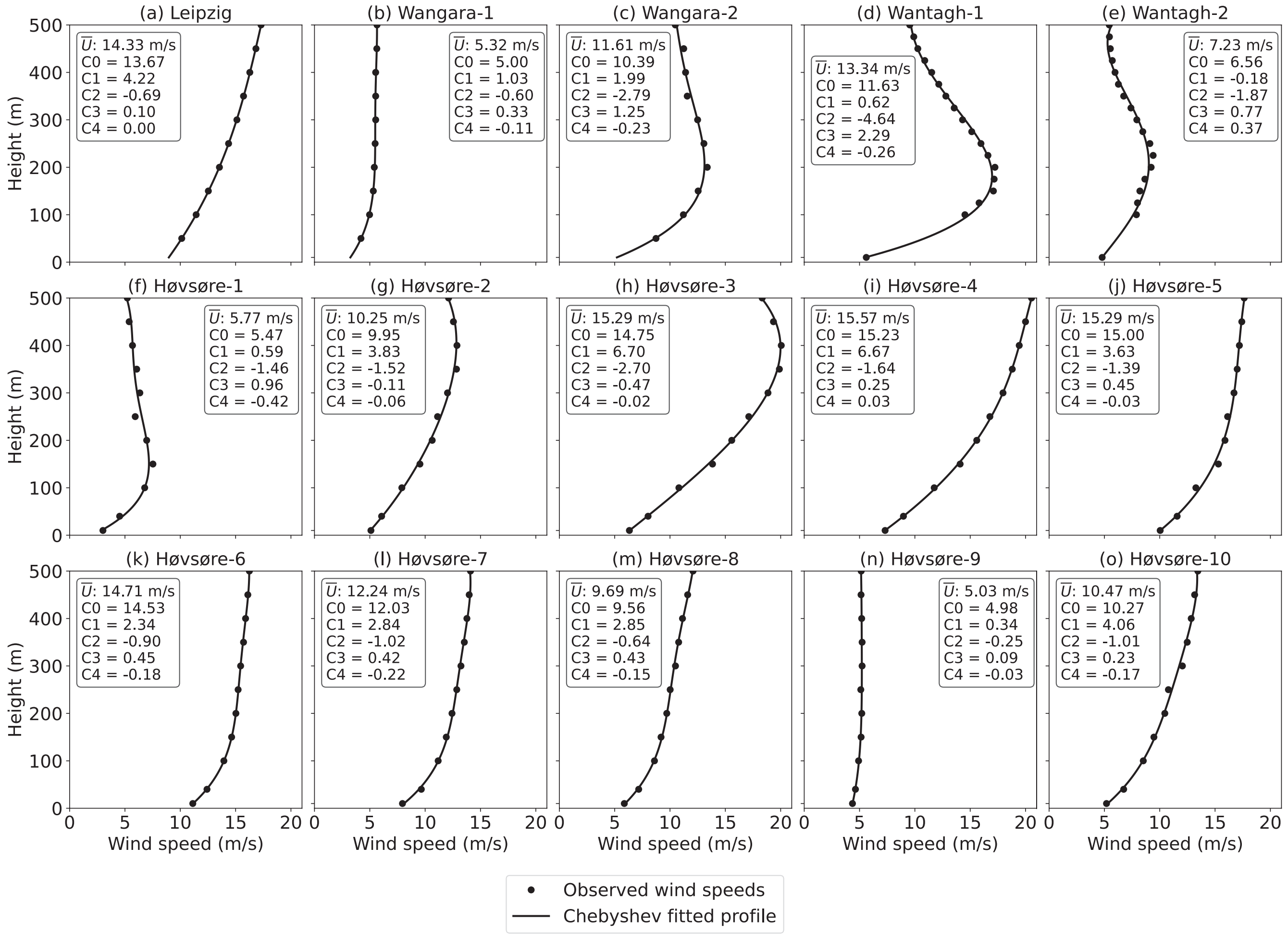


FIGURE 2 | Observed wind profiles at Leipzig, Wangara, Wantagh, and Høvsøre. The profiles are approximated by Chebyshev polynomials of order 4, which are overlaid using a solid black line, and the corresponding coefficients (C_0 to C_4) are shown in the text box. In addition, the mean of the observed wind speed profile ($\bar{U}$) is also shown in the text box.

and can be effectively characterized by a combination of moderate C_1 and negative C_2.

Høvsøre-6 (Figure 2k) was observed during afternoon hours (0830–1640) under slightly unstable but high-forcing conditions. The profile has a significantly high mean wind speed (14.71 m/s), reflected in a high C_0 value. The moderate values of C_1 and C_2 indicate moderate shear and curvature, while the higher order coefficients contribute little.

Finally, Høvsøre-8 (Figure 2m) and Høvsøre-10 (Figure 2o) were observed under slightly stable conditions during 0200–0330 and 0820–0950, respectively. These profiles show strong shear (high C_1) and moderate curvature (moderate negative C_2), with very low contributions from higher order coefficients. The lack of significant C_3 and C_4 implies that fluctuations and complex structures are minimal, resulting in smooth, shear-driven profiles.

From these profiles, it is clearly evident that the Chebyshev coefficients effectively capture the key features of wind speeds. In addition, there will be a unique profile for a unique set of coefficients, implying that a one-on-one comparison is possible between two different wind speed profiles, just by comparing their corresponding coefficients.

3 | Description of Data

3.1 | Wind Speed Profiles in the NOW23 Dataset

The 2023 National Offshore Wind dataset (NOW23) [59] was implemented to enhance the Wind Integration National Dataset (WIND) Toolkit by providing high-resolution wind data specifically for offshore environments. This dataset was generated using the Weather Research and Forecasting (WRF) model version 4.2.1, forced with ERA5 reanalysis data for atmospheric variables and the Operational Sea Surface Temperature and Ice Analysis (OSTIA) for surface temperature inputs. Spatially, a two-domain nested approach was employed, with resolutions of 6 km for the outer domain and 2 km for the inner domains. The outer domain spans the entire United States, while eight inner domains focus on key coastal regions: the North Pacific, South Pacific, Hawaii, Gulf of Mexico, South Atlantic, Mid-Atlantic, North Atlantic, and the Great Lakes. In this study, we utilized data from the Great Lakes and Mid-Atlantic regions, which effectively cover the New York State. The dataset provides wind speed at height levels: 10 m, every 20 m from 20 to 300, 400, and 500 m. With a temporal resolution of 5 min, the dataset spans from 2000 to 2020 across these two regions. For computing the Chebyshev coefficients respective to each wind profile, the height coordinate is normalized between [0 m, 500 m] to [-1, 1]. Since the Chebyshev polynomials are time invariant, the Chebyshev coefficients for each wind profile, varying with time, are computed by solving a system of linear equations.

3.2 | Wind Speed Profiles in the NYSM Profiler Dataset

New York State has established a network of 18 wind profiling stations, collectively known as the New York State Mesonet (NYSM) Profiler Network, continuously monitoring atmospheric conditions [9, 10, 60]. Each station is equipped with a Leosphere–Vaisala Scanning Windcube Doppler lidar (DL) 100S, a state-of-the-art remote sensing instrument operating at a near-infrared wavelength of 1540 nm. The lidars operates in Doppler Beam Swinging (DBS) mode, scanning in five directions: four scans at a 75° elevation in the cardinal directions and one vertical scan at 90°. By averaging these scans, the system derives three-dimensional wind components. Measurements cover altitudes from 100 to 7000 m, with a vertical spacing of 25 m below 1000 and 50 m above 1000 m.

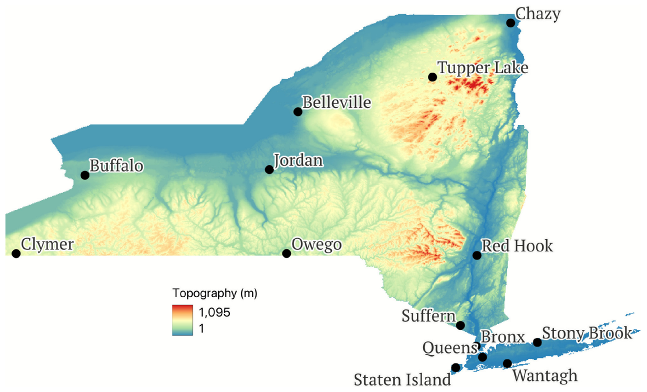


FIGURE 3 | New York State Mesonet profiler station locations, satisfying our selection criteria, overlaid on orography.

In this study, we focus on characterizing wind profiles up to 500 m. Since the Doppler lidars begin measurements at 100 m, we incorporate wind speed measurements from nearby NYSM Standard Network sites, which provide wind speeds at 10 m [60]. We utilized profiler data from 2018 to 2020, at 5-min temporal resolution, aligning with the temporal availability of the NOW23 dataset. Moreover, not all NYSM profiler stations are located within the NOW23 domains, and not all stations are close to NYSM Standard Network sites. Based on a collocation analysis, we identified 13 stations that meet the selection criteria These selected sites are shown in Figure 3, and their details are provided in Table 2. Here as well, the height coordinate is normalized between [0 m, 500 m] and [–1, 1], and the Chebyshev coefficients of each profile, varying with time, are computed.

3.3 | Data Processing

In this study, we opted Chebyshev polynomials up to the fourth-order. A detailed analysis on how and why we opt for the fourth-order polynomials is given in Appendix A. The NOW23 wind profile dataset is complete, with no missing values in time or height, as it is generated from continuous WRF model simulations. In contrast, the lidar wind profile observations contain missing instances and gaps at various height levels. These gaps arise due to limitations such as aerosol type and concentration, precipitation, atmospheric refractive turbulence, and humidity. To retain as much usable data as possible from the lidar records, a conditional threshold was applied before computing the Chebyshev coefficients. Since the Chebyshev polynomial approximation does not require values at every height level, but rather a sufficient number of points for a stable fit, the following criteria were used: valid wind speed observation at 10-m level, at least two valid wind speed observations between 100 and 200 m,

TABLE 2 | Details of NYSM profiler stations.

stid	name	Llatitude (degrees)	Longitude (%)	Elevation (m)
BUFF	Buffalo	42.993590	–78.794610	185.39
BELL	Belleville	43.788230	–76.117650	152.10
BRON	Bronx	40.872481	–73.893522	59.31
STAT	Staten Island	40.604014	–74.148499	34.43
STON	Stony Brook	40.919579	–73.133284	55.10
QUEE	Queens	40.734335	–73.815856	52.89
WANT	Wantagh	40.650250	–73.505400	18.25
TUPP	Tupper Lake	44.224256	–74.441052	525.20
SUFF	Suffern	41.133034	–74.085979	191.87
CLYM	Clymer	42.021430	–79.627460	457.45
JORD	Jordan	43.068747	–76.469993	129.46
OWEG	Owego	42.024938	–76.253072	464.45
REDH	Red Hook	41.999830	–73.884120	72.85

at least four between 225 and 375 m, and at least two between 400 and 500 m. A detailed analysis on the validity of the conditional search is provided in Appendix B. If an instance satisfied these conditions, Chebyshev coefficients were computed from the lidar data. The same time instances were then extracted from the NOW23 dataset, and their corresponding Chebyshev coefficients were calculated to ensure consistent comparison. We would like to point out that all the datasets considered in this study are recorded in UTC.

With the lidar and NOW23 wind speed profiles pre-processed and their corresponding Chebyshev coefficients computed at matched time instances, the datasets are now prepared for evaluation. In the next section, we assess the degree of agreement between the modeled (NOW23) and observed (NYSM lidar) wind profiles using two complementary approaches. First, we adopt conventional height-based evaluation metrics to compare wind speeds at specific altitude levels. Second, we apply the proposed C-WiSPr framework to compare the morphological characteristics of the profiles via their Chebyshev coefficients. This dual-stage evaluation enables a more comprehensive understanding of both level-wise and shape-based differences in wind profile representations.

4 | Results

The following results present the two-stage evaluation of NOW23 wind profiles against NYSM lidar observations. First, in the conventional height-based evaluation, evaluation metrics, including mean, bias, cRMSE, and r, are computed at common height levels between the two datasets. These metrics are plotted as a function of height to provide insight into the vertical agreement between the simulated and observed wind profiles. Second, a direct comparison is performed between the coefficients derived from the NOW23 simulations and lidar observations, allowing for evaluation of how well the simulations capture key wind profile features: mean wind speed (C_0), shear (C_1), curvature (C_2), and higher order fluctuations (C_3, C_4). Bias, cRMSE, and r are computed for each coefficient to quantitatively assess the level of agreement. In addition, bivariate histograms, quantile-quantile (Q-Q) plots, and probability density function (PDF) comparisons are used for qualitative evaluation of the distributional similarity between the two datasets. To explore temporal dynamics, the diurnal and seasonal variability of the five coefficients are analyzed, providing further insights into agreement or divergence in characteristic patterns. Finally, r values are computed across stations and coefficients to investigate geographical similarity in wind profile characteristics.

The NOW23 dataset is primarily developed to support offshore wind resource assessments. To evaluate its performance in this context, the analysis categorizes the stations into inland and coastal groups. The inland stations include Clymer, Owego, Suffern, Tupper, Red Hook, and Jordan. The coastal stations are further divided based on their location: those situated along the Atlantic Ocean are Staten Island, Stony Brook, Queens, Wantagh, and Bronx, and those along the Great Lakes are Buffalo and Belleville. This classification allows for a focused assessment of how well the NOW23 simulations align with lidar observations across different geographic settings, particularly to determine whether the dataset effectively serves its intended purpose in coastal and offshore environments.

We acknowledge that lidar-based wind measurements come with inherent limitations. As height increases, the measurement volume of the lidar also expands, which means that small-scale turbulence can influence the 10-minute mean values differently across range gates [61]. Several external factors such as terrain complexity, surface roughness, forest cover, measurement height, atmospheric stability, and the lidar's half-cone opening angle also contribute to uncertainty in the observations [62]. Given these considerations, we treat the lidar dataset as a reference rather than an absolute ground truth. Our primary aim is not to assess the accuracy of the NOW23 dataset, but to demonstrate the strength of the newly proposed C-WiSPr framework used for wind profile evaluation. Although we use terms like “overestimation” and “underestimation” throughout the paper, these refer solely to deviations from lidar-based estimates, but not from a verified ground truth.

4.1 | Conventional Evaluation

Figures 4 and 5 show the vertical profiles of wind speed statistics and evaluation metrics, computed at each height level, comparing the NOW23 simulated and lidar-observed wind datasets at coastal and inland sites, respectively.

The mean profiles are well matching at the coastal sites, whereas a considerable deviation is visible at the inland sites. Among all stations, the best agreement is observed at Wantagh, whereas the largest deviation is seen at Jordan. A closer look at the lidar profiles reveals that the 10-m mean wind speed is approximately 2 m/s at four inland sites, but only at one coastal site. Conversely, the 500 m mean wind speed reaches around 10 m/s at five inland sites, compared to three coastal sites. This supports that inland sites generally experience stronger vertical wind shear than coastal sites. However, beyond this general trend, the mean wind profile alone does not provide detailed insight into how shear evolves with height, unless such variations are explicitly measured or derived through higher order metrics.

The vertical profiles of bias also reflect this overall pattern, with inland stations exhibiting higher bias compared to coastal ones. For most stations, the bias increases from the 10 m level up to around 100 m and then gradually decreases with height. Among all sites, the highest positive bias is recorded at Jordan, reaching 3.3 m/s at 100 m, followed by Clymer, which also shows a notably high bias at this level. In contrast, Wantagh exhibits the lowest bias across all levels, even showing a negative bias at 10 m. At 500 m, the bias at Wantagh reduces to nearly zero, indicating excellent agreement between the simulated and observed wind speeds.

The cRMSE, computed at common height levels, represents the random error component after the bias is removed between the NOW23 and lidar wind speed datasets. Interestingly, the cRMSE tends to increase rapidly from the 10 m level up to around 100 m, then continues to rise gradually, typically peaking between 200 and 300 m, before decreasing again at higher levels. This trend suggests that the random error is lowest at 10 m, either due to the

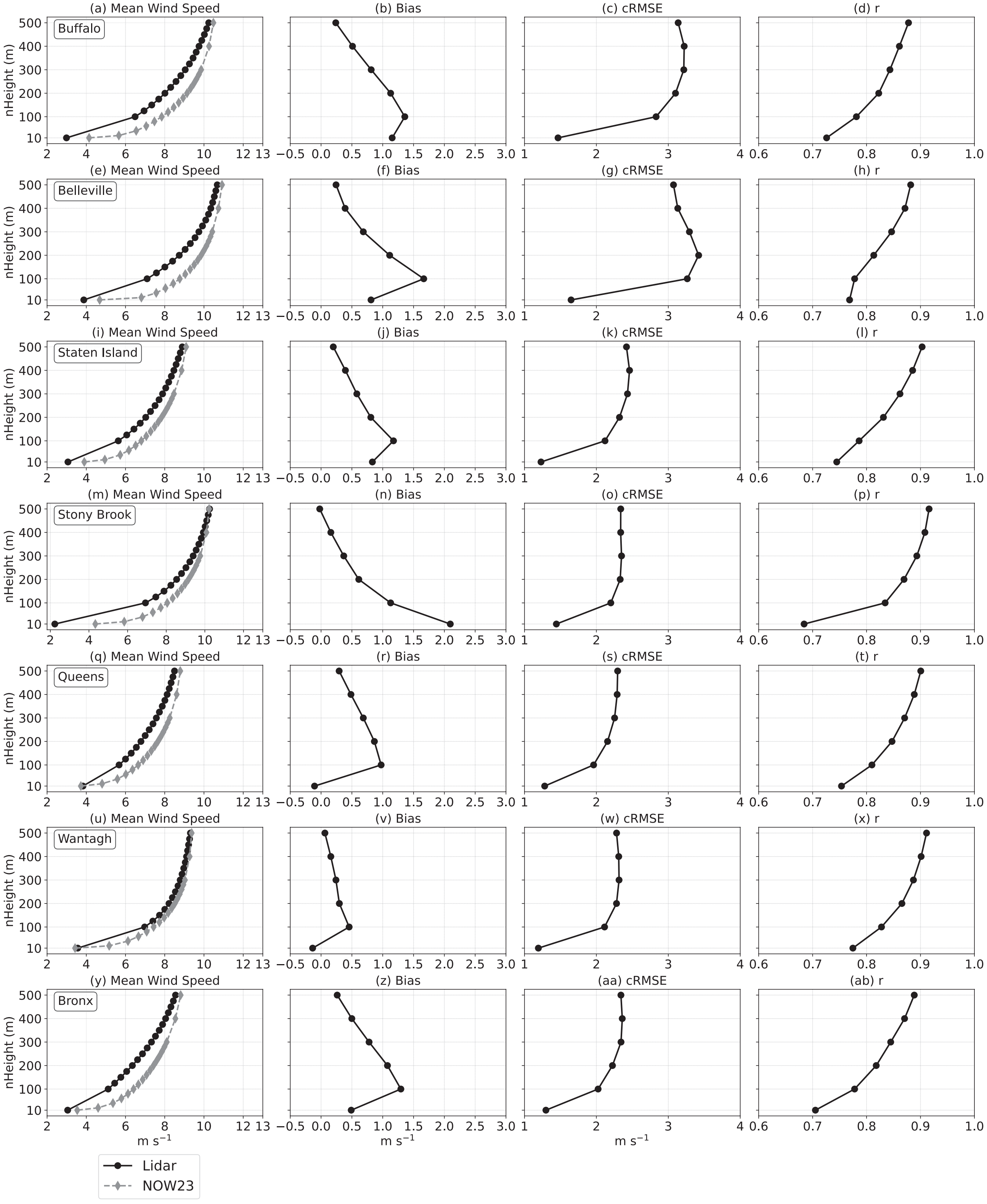


FIGURE 4 | Vertical profiles of wind speed statistics and evaluation metrics computed at each height level from NOW23 and lidar datasets, for the full analysis period (2018–2020), at coastal sites: (a–d) Buffalo, (e–h) Belleville, (i–l) Staten Island, (m–p) Stony Brook, (q–t) Queens, (u–x) Wantagh, and (y-ab) Bronx. The columns represent different variables: the first column shows mean wind speed; the second column shows bias; the third column shows centralized root mean square error (cRMSE); the fourth column shows Pearson's correlation coefficient (*r*).

lower wind speeds or better agreement between the NOW23 and lidar winds at this level. Across all stations, coastal sites exhibit lower cRMSE at the 10 m level, with four of them showing values around 1.2 m/s, while inland stations report higher errors, typically around 1.5 m/s. At higher levels, five coastal stations maintain cRMSE in the range of 2–2.5 m/s, whereas inland stations

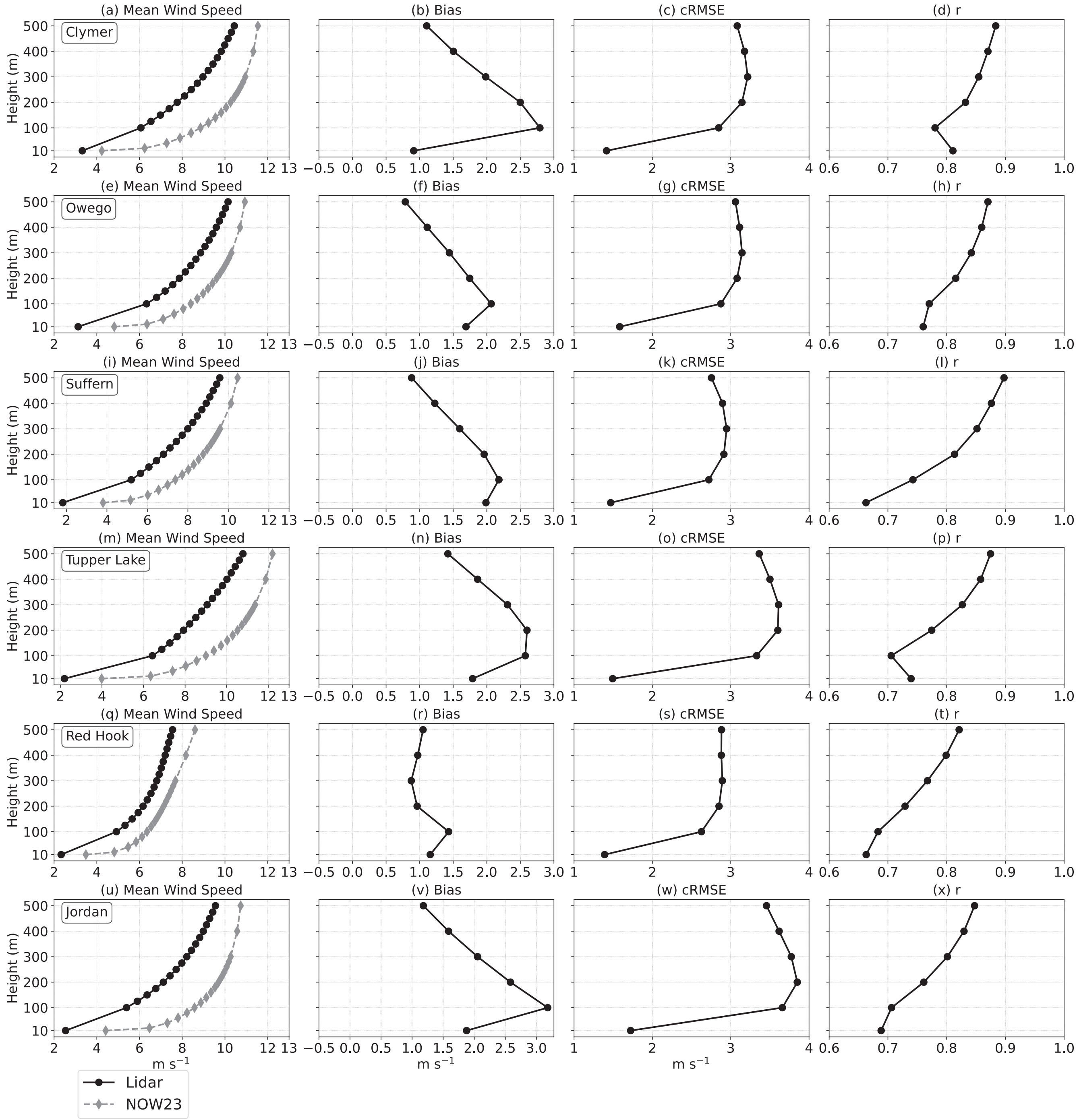


FIGURE 5 | Same as Figure 4, but for inland stations: (a–d) Clymer, (e–h) Owego, (i–l) Suffern, (m-p) Tupper Lake, (q–t) Red Hook, and (u–x) Jordan.

show higher values, reaching approximately 3 m/s. These results clearly indicate that NOW23 performs better over coastal regions, aligning more closely with lidar observations in those areas. Furthermore, a comparison within the coastal group reveals that stations located near the Great Lakes (Buffalo and Belleville) exhibit significantly higher cRMSE values (exceeding 3 m/s) compared to those situated along the Atlantic Ocean. Despite these insights, it is important to note that cRMSE alone does not provide information about the specific types of wind profiles or atmospheric conditions that contribute most to the observed discrepancies between NOW23 and lidar data.

From the vertical profiles of r, it is evident that the correlation is lowest at the 10 m level (approximately 0.7) and consistently increases with height, reaching values close to 0.9 at higher levels. Notably, across the coastal stations, the correlation coefficient exceeds 0.8 above the 200 m level, indicating strong agreement between the simulated and observed wind patterns. In contrast, only three inland stations reach correlation values above 0.8 at similar heights. This distinction highlights that the agreement between NOW23 simulations and lidar observations, in terms of the wind speed pattern correlation at respective heights, is notably stronger at coastal sites than at inland locations.

Thus far, the evaluation of the NOW23 wind profiles has primarily focused on direct wind speed comparisons at multiple height levels. However, these methods do not provide insights into critical characteristics such as the overall profile structure, wind shear, curvature in profiles, the presence of LLJs, or intermittent fluctuations associated with specific weather conditions.

4.2 | C-WiSPr Framework-Based Evaluation

4.2.1 | Aggregated Analysis

Figure 6 compares Chebyshev coefficients derived from lidar and NOW23 datasets, aggregated across 13 stations over a 3-year period. Outlier values those beyond the 0.1% and 99.9% percentiles of the lidar distributions are removed from both datasets to ensure a consistent comparison. The first row of subplots shows bivariate histograms, the second row shows quantile-quantile (Q-Q) plots, and the third row displays probability density function (PDF) comparisons. While the bivariate histograms examine concurrent times, the Q-Q and PDF plots are better suited for assessing how well the model captures the overall distribution. Agreement is indicated when points follow the 45-degree line; deviations suggest over or underestimation by NOW23. In bivariate histograms, an elliptical contour is drawn, that encapsulates points within three standard deviations of the joint distribution. In Q-Q plots, since the coefficients will have both positive and negative values, points lying above (below) the 1:1 line in the positive (negative) domain imply overestimation, and vice versa.

From Figure 6a, the coefficient C_0, representing mean wind speed, exhibits strong correlation but a clear positive bias, as indicated by the tilted elliptical contour, highlighting consistent overestimation by NOW23. For example, at a lidar value of $C_0 = 10$, NOW23 ranges from 5 to 18, while at $C_0 = 10$ in NOW23, the lidar values range more narrowly from 3 to 14. The bias, cRMSE, and correlation coefficient are 1.24, 2.38, and 0.86, respectively. The Q-Q plot (Figure 6f) further shows overestimation beginning around $C_0 = 2$, becoming more pronounced in the upper tail. The PDF comparison (Figure 6k) reveals that while both distributions center around $C_0 = 5$, the NOW23 profile is broader with heavier tails, confirming greater variability and positive bias.

Shear in the wind profile is represented by C_1, which shows strong correlation but a modest underestimation. In the bivariate histogram (Figure 6b), the elliptical contour aligns

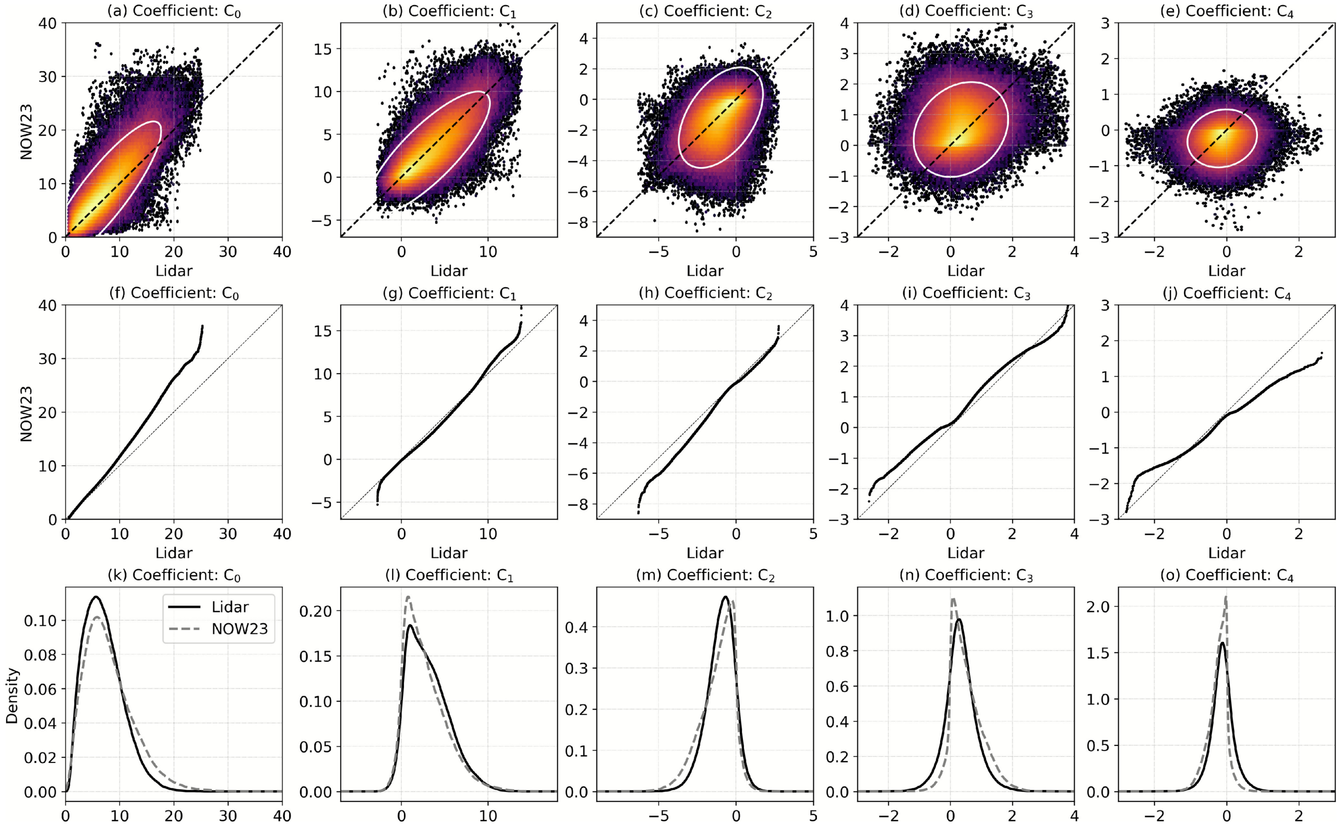


FIGURE 6 | (a-e) Bivariate histograms comparing the five Chebyshev coefficients (C_0-C_4) between lidar and NOW23 data, aggregated across all 13 stations over the 3-year period. Lidar values are shown on the abscissa and NOW23 values on the ordinate. The dashed line at 45° represents the 1:1 reference line, indicating perfect linear agreement. Brightened regions indicate higher data density, and a white ellipse in each panel represents the joint distribution, encompassing data points within three standard deviations. A tighter clustering of points along the 1:1 line signifies stronger linear correlation, while deviations highlight discrepancies. (f-j) Quantile-quantile (Q-Q) plots for the same five coefficients, with lidar values on the abscissa and NOW23 values on the ordinate. The dashed 45° line represents ideal agreement; departures from this line reflect differences in distribution between the two datasets. (k-o) Probability density function (PDF) plots comparing the distributions of each coefficient, where the black solid line represents lidar data and the black dashed line represents NOW23 data.

with the 1:1 line but is shifted downward, indicating a bias of -0.57. The cRMSE is 1.44 and the correlation is 0.80. The Q-Q plot (Figure 6g) shows consistent underestimation through the mid-range, with slight overestimation at the distribution tails. The PDF plot (Figure 6l) indicates that NOW23 underrepresents the variability in shear, producing a narrower and more sharply peaked distribution compared to the broader lidar distribution.

Curvature, captured by C_2, is moderately well represented. As seen in Figure 6c, the elliptical shape aligns roughly with the 1:1 line but is wider, indicating greater scatter. The bias is minimal (-0.03), though cRMSE is 0.94 and correlation is lower at 0.47. The Q-Q plot (Figure 6h) shows reasonable alignment at the center but deviations in the tails, especially due to NOW23 overestimation. The PDF comparison (Figure 6m) reveals the NOW23 distribution is sharply peaked at zero, while lidar distribution is peaked slightly towards negative with reduced density, suggesting that NOW23 underrepresents the full range of curvature variability.

Coefficient C_3, which describes part of the overall linear shear and cubic curvature in the profile, is weakly correlated between datasets. In the bivariate histogram (Figure 6d), the contour is centered but loosely aligned with the 1:1 line, and the scatter is more pronounced. Bias is small (0.09), with a cRMSE of 0.58 and correlation of only 0.23. The Q–Q plot (Figure 6i) shows misalignment at both extremes, indicating that NOW23 fails to capture high-inflection profiles. The PDF plot (Figure 6n) shows that NOW23 is more peaked and narrower, again reflecting the underestimation of variability present in the lidar data.

The highest order coefficient, C_4, reflects small-scale quartic curvature and shows the weakest agreement. The bivariate histogram (Figure 6e) displays a near-circular elliptical contour centered on zero, with widespread scatter. The correlation is very low (0.13), with a cRMSE of 0.41 and a negative bias of -0.17. The Q-Q plot (Figure 6j) shows consistent underestimation across the range. The PDF comparison (Figure 6o) highlights a noticeable difference: the NOW23 curve is narrowly peaked around zero and skewed slightly left, while the lidar curve is broader and centered slightly negative. This reflects a general underrepresentation of higher order variability by the NOW23 dataset.

Although the contribution of C_4 to the overall characterization and wind profile approximation is relatively small, using a fourth-order polynomial is still important for minimizing the overall approximation error, as discussed in Appendix A. We also found that 4th-order or higher polynomials are necessary to capture non-monotonic wind speed profiles with negative wind shear or multiple inflection points and local minima/maxima, as discussed in Appendix C.

From the C-WiSPr framework comparison, one would notice that the N number of height levels per profile is compressed into just five coefficients. Thus, we no longer require the wind profiles to be on the same height levels for a better comparison, which is the case with many NWP model simulations and observations.

4.2.2 | Station-Wise Analysis

To assess the spatial consistency of model performance, we analyzed the agreement between Chebyshev coefficients derived from NOW23 and lidar observations at individual stations. Figure 7 presents quantile-quantile (Q-Q) plots for coastal and inland stations, highlighting agreement across selected percentiles (0.1%, 1%, 25%, 50%, 75%, 99%, and 99.9%) to capture central and extreme distributional characteristics.

The results show a consistent overestimation of the mean coefficient C_0, particularly at inland stations, with the largest discrepancies occurring at the upper tail. Wind shear (C_1) is well represented overall, although minor underestimations appear at central quantiles, and deviations in the upper tail are more evident at coastal sites. The curvature coefficient C_2 is generally well captured within the interquartile range, though biases increase at the tails. The NOW23 tends to overestimate curvature at inland stations and underestimate it at coastal extremes. For C_3 and C_4, model performance declines, particularly in capturing distribution tails, indicating limitations in representing complex vertical structures like inflection points and higher order fluctuations.

To complement the Q-Q analysis, Figure 8 summarizes bias, cRMSE, and correlation r between lidar and NOW23 coefficients across stations. As expected, model performance is best for C_0 and C_1, with uniformly high correlation ($r > 0.8$ for C_0) and lower errors at coastal sites. Agreement diminishes progressively with coefficient order: C_2 shows moderate correlation ($r = 0.4$–0.6 at coastal stations), while C_3 and C_4 exhibit weak correlations across most sites ($r < 0.4$), reflecting challenges in reproducing inflectional and high-frequency features. Inland stations generally exhibit higher biases and lower correlations across all coefficients.

While these findings provide valuable insights into the spatial variability of model performance, we acknowledge that conclusions drawn from only 13 stations, which are further subdivided into inland and coastal groups, should be interpreted cautiously. Nonetheless, the observed trends support the core argument that NOW23 captures bulk profile features (mean and shear) more reliably than localized or transient structures, with better overall performance in coastal environments.

4.2.3 | Analysis of diurnal and seasonal variability

Figure 9 presents the seasonal and diurnal variability of Chebyshev coefficients from lidar (solid line) and NOW23 (dashed line) data, at the Wantagh station. Strong diurnal patterns are evidently visible in C_0 to C_3, from both lidar and NOW23 coefficients, across all the seasons. These coefficients generally dip during 14-18 UTC (local late-morning and early-afternoon hours), rise again afterwards, and peak either during 22-00 UTC hours (local early-evening hours) or during 04-11 UTC (local nighttime to morning hours). In contrast, coefficient C_4 shows no clear diurnal pattern in the lidar data, while NOW23 displays a distinct and systematic diurnal cycle.

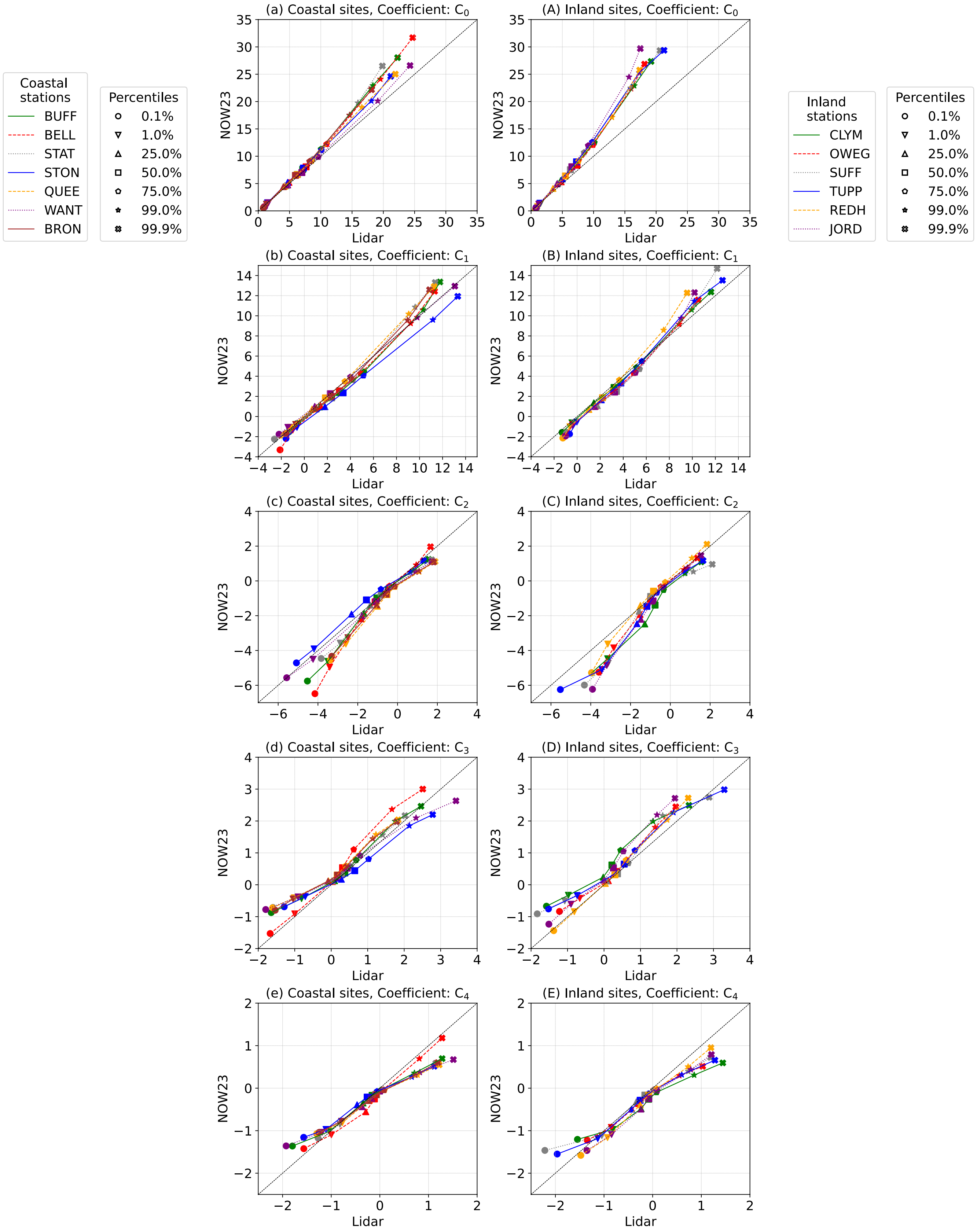


FIGURE 7 | Quantile-quantile plots comparing LiDAR and NOW23 Chebyshev coefficients at coastal sites (a-e) and inland sites (A-E). The subplots (a-A) correspond to coefficient C_0, (b-B) to C_1, (c-C) to C_2, (d-D) to C_3, and (e-E) to C_4. Percentiles are computed at 0.1%, 1%, 25%, 50%, 75%, 99%, and 99.9%. In each Q-Q plot, markers represent these percentiles.

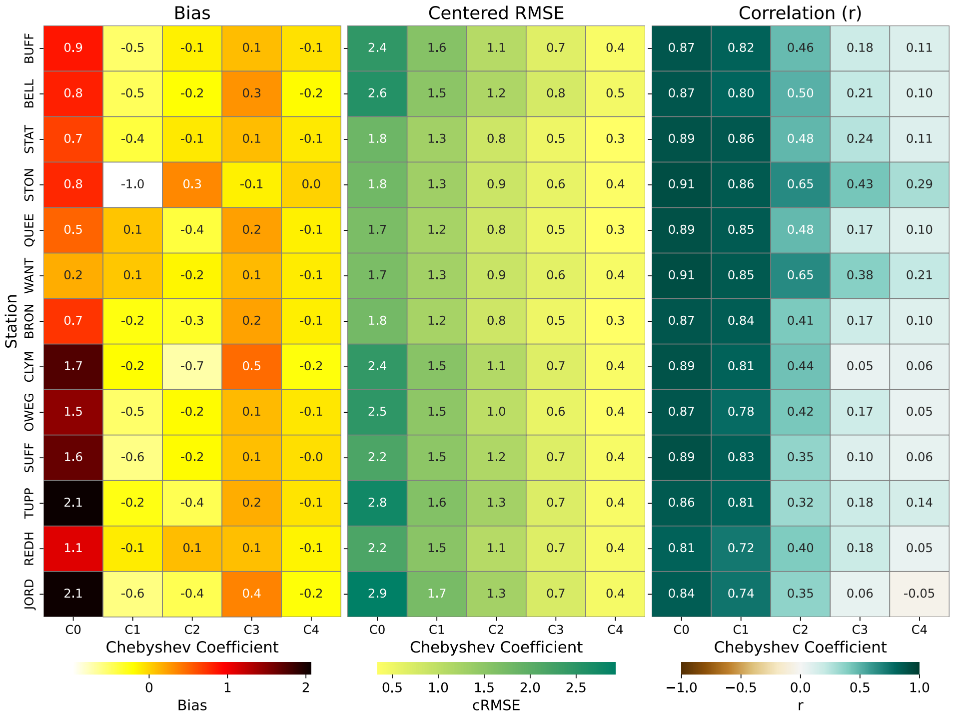


FIGURE 8 | Heatmaps comparing the bias (left panel), cRMSE (middle panel), and *r* (right panel) computed between the Chebyshev coefficients of lidar and NOW23 data, at individual sites, aggregated over a 3-year period.

Seasonal trends are evident in coefficients C_0 to C_3, with values typically lowest during summer, increasing through autumn and winter, and peaking in spring. The timing and magnitude of diurnal peaks and dips vary across seasons. Among the coefficients, C_1 shows the strongest agreement between the lidar and NOW23 datasets, particularly in autumn and summer, with only slight deviations in winter and spring. This consistency highlights that the simulated and observed shear patterns are broadly aligned.

The patterns of C_0 also align well during winter and autumn. During spring and summer, partial agreement is observed during dip hours (0800–1600 UTC), while NOW23 tends to overestimate during other periods, indicating a general overestimation of mean wind speeds. For C_2 (curvature), agreement between the datasets is limited mostly to dip hours across all seasons, while NOW23 consistently overestimates during rising and peak hours, suggesting excessive curvature in the simulated profiles.

In contrast, coefficient C_3 exhibits a persistent positive bias in NOW23, with only partial agreement during dip periods in autumn and winter. During spring and summer, a negative bias is observed during rising hours. Since C_3 interacts with both C_1 and C_2, biases in this coefficient can alter the overall linear shear, shift localized shear near the surface, and lower the jet nose height, affecting the accuracy of simulated LLJ structures.

When the linear shear coefficient C_1 is at a minimum, the influence of higher order coefficients becomes more pronounced in shaping the wind profile. Notably, summer exhibits the largest discrepancies between the lidar and NOW23 datasets in coefficients C_2 and C_3, particularly during 1600-0000 UTC hours when C_1 is unusually low compared to other seasons. To investigate the resulting wind profile structures, we extracted the coefficients during this period and reconstructed wind speed profiles, as shown in Figure 10.

Across both datasets, the reconstructed profiles show evidence of local wind speed maxima, indicative of LLJs. Between 1600 and 2200 UTC, C_2 remains similar across the datasets, while C_3 is lower in NOW23. This results in a higher jet nose height and weaker near-surface shear in the NOW23 profiles compared to those from lidar. At 0000 UTC, both C_2 and C_3 are elevated in the NOW23 dataset, producing a sharper curvature and steeper shear near the ground, along with a lower jet nose height. These subtle variations in the coefficients clearly demonstrate their ability to characterize morphological differences in wind profiles, particularly during periods when the shear is diminished.

4.2.4 | Cross-Station Correlation

When observations at a target site are limited, long-term wind characteristics can be estimated by correlating short-term measurements at the site with long-term data from a nearby reference station. This process is commonly referred to as the measure-correlate-predict (MCP) approach [63]. To assess how well the NOW23 dataset captures cross-station wind relationships, Pearson's correlation coefficient (*r*) is computed between Chebyshev coefficients from all pairs of stations. Figure 11 displays the resulting heatmaps, with lidar-based values on the left and NOW23-based values on the right.

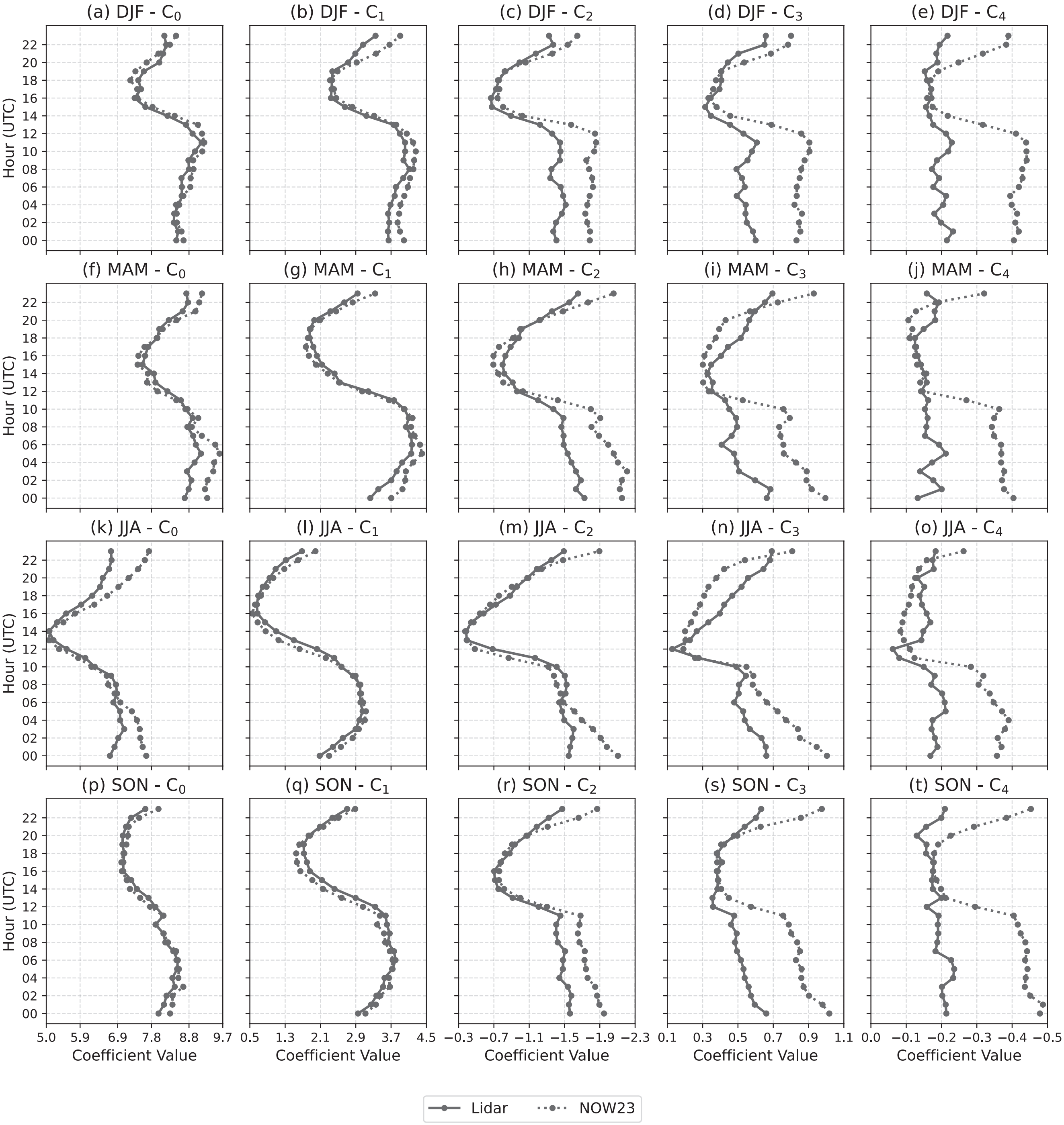


FIGURE 9 | Comparison of the diurnal and seasonal variability of Chebyshev coefficients from lidar (solid lines) and NOW23 (dashed lines) at the Wantagh station, shown for (a–e) Winter (December–February), (f–j) Spring (March–May), (k–o) Summer (June–August), and (p–t) Autumn (September–November). Each column corresponds to one Chebyshev coefficient: C_0 (mean wind speed), C_1 (shear), C_2 (curvature), C_3 (inflection), and C_4 (higher order fluctuations), from left to right. For interpretability, the horizontal axis (coefficient values) is oriented positively for C_0, C_1, and C_3, and negatively for C_2 and C_4, as negative values are more physically significant for these latter coefficients.

For the primary coefficients C_0 to C_2 (Figures 11a-f), both datasets exhibit similar spatial patterns. Strong correlations are observed among the Atlantic coastal stations: Staten Island, Stony Brook, Queens, Wantagh, and Bronx, suggesting consistent wind regimes of mean, linear shear, and curvature characteristics across these locations. Although Suffern is not categorized as a coastal site, its proximity to the ocean may explain its high correlation with these stations in both datasets.

A second group, consisting of Buffalo, Belleville, Clymer, Owego, Tupper Lake, and Jordan, also shows strong internal correlation. While Buffalo and Belleville are both located along the Great Lakes, the remaining stations are farther inland. Nonetheless, the high correlations imply similar wind profile features such as mean wind speed, shear, and curvature. One exception is Tupper Lake, where lidar-based C_2 shows weak correlation with other sites, while NOW23-based C_2 indicates strong correlation.

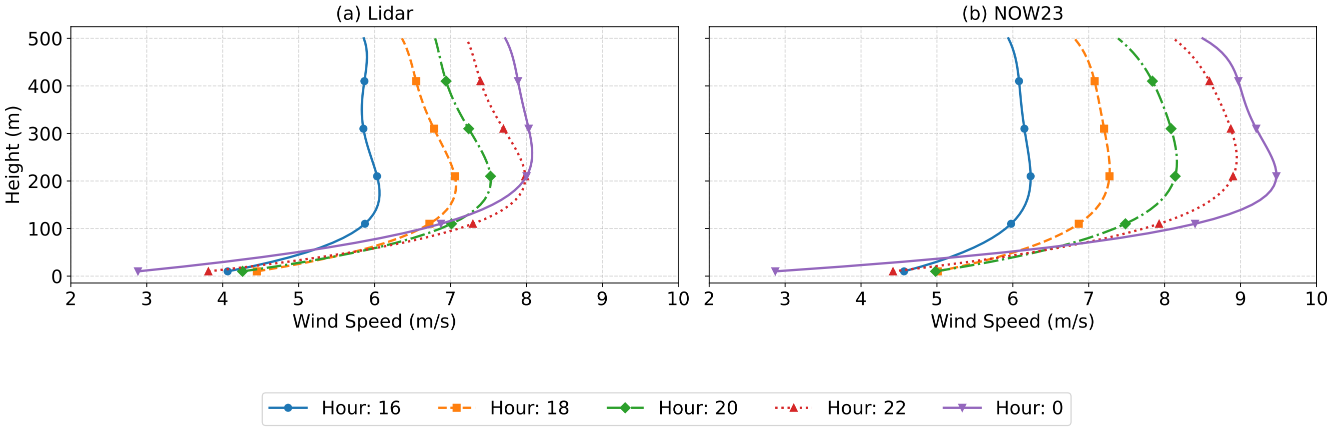


FIGURE 10 | Comparison of wind speed profiles recreated from the mean diurnal Chebyshev coefficients, during Summer, across the time range, at station Wantagh. The left column shows wind profiles from the lidar data, while the right column shows wind profiles from the NOW23 data

This may suggest that NOW23 overrepresents curvature characteristics at this station.

For the higher order coefficients C_3 and C_4 (Figure 11g–j), the coastal stations again show high interstation correlation in both datasets. However, correlation values in the NOW23 data are significantly higher, indicating that the model produces more spatially coherent patterns of cubic and quartic curvature. These features are often related to near-surface shear and LLJs. In contrast, lidar-derived coefficients for inland stations show little correlation, while NOW23 data still reflect moderate correlation. This suggests that the NOW23 may exaggerate the spatial consistency of fine-scale wind structures in these regions.

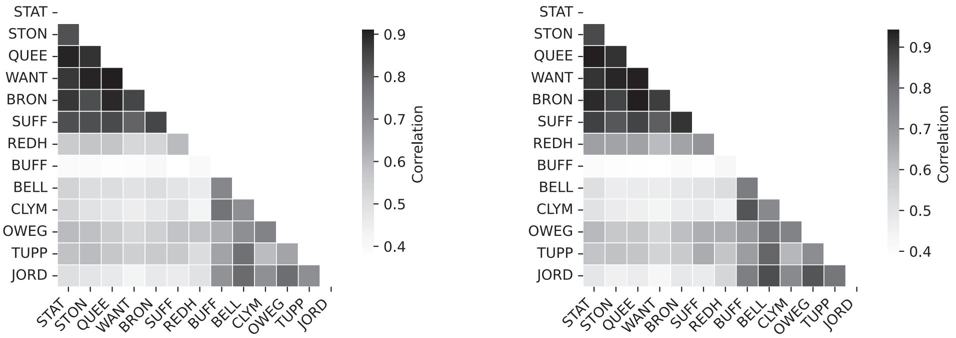
5 | Conclusions and Future Scope

In this study, we introduced a novel framework for wind profile characterization based on Chebyshev polynomials, termed C-WiSPr. This approach transforms horizontal wind speed data at multiple height levels into a compact set of five physically meaningful coefficients (C_0 to C_4), offering an insightful representation of the entire vertical wind profile structure. Each coefficient captures a distinct aspect of the wind profile: C_0 represents the overall mean wind speed, C_1 describes vertical shear, C_2 accounts for curvature, while C_3 and C_4 represent higherorder variations. We demonstrated that well-mixed profiles (prevalent in afternoon conditions) can be effectively characterized by C_0 alone; high-forcing monotonic shear profiles (such as during gale events) by C_0 and C_1; logarithmic-shaped profiles (often occurring under neutral afternoon to evening conditions) by a combination of C_0, C_1, and C_2; and LLJ profiles (frequently observed under stably stratified conditions and land-sea breeze circulations) by the interplay of C_0, C_2, and C_3. In cases with subtle inflectional features or fine-scale upperlevel fluctuations, C_4 offers marginal but relevant contributions to capturing profile shape, particularly in resolving structure near 400–500 m.

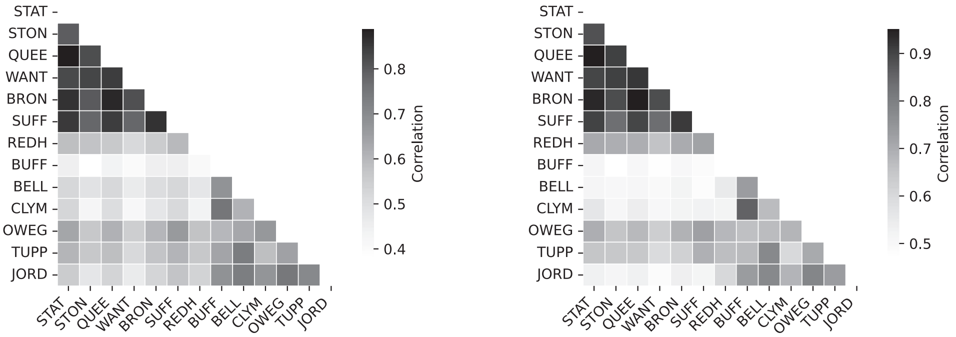
This C-WiSPr framework enables a physically interpretable way to quantitatively compare entire wind profiles, regardless of their height levels. To demonstrate its applicability, we evaluated wind profiles simulated by the WRF model (NOW23 dataset) against lidar-observed profiles from the NYSM profiler network, despite the two datasets existing at different height levels. The NOW23-derived coefficients reliably captured the mean wind speed (C_0) and the overall shear structure (C_1). However, its performance declined with increasing coefficient order, revealing limitations in representing curvature (C_2), inflection features (C_3), and fine-scale fluctuations (C_4). Bivariate distributions, Q–Q plots, and probability density comparisons highlighted a systematic overestimation of mean wind speeds and an underrepresentation of variability in both shear and higher order structural features. Station-wise analysis showed that inland sites exhibited larger discrepancies compared to coastal stations, reaffirming that the NOW23 dataset is better suited for the coastal wind resource assessment for which it was originally designed. When we examined the variability, the NOW23 dataset closely followed the observed diurnal and seasonal patterns of mean wind speed, overall shear, and curvature at the Wantagh station. At this coastal location, the occurrence of low-level jets was evident through elevated C_2 and C_3 values alongside reduced C_1 values, particularly during summer months. These features are attributed to sea breeze circulations and were reasonably well captured by the NOW23 dataset.

The C-WiSPr framework has demonstrated its potential in quantitatively validating simulated wind profiles from observations, offering a variety of morphological characteristics beyond conventional height-based metrics such as bias, cRMSE, and Pearson's correlation coefficient. This form of profile-level evaluation provides more insights into model behavior, making it especially valuable for simulation frameworks and model development. We anticipate that the framework can be rigorously applied in sensitivity studies involving modeling configurations, including physical parameterization representations, initial and boundary conditions, and domain setups. Such applications can aid in identifying optimal configurations and guide the development of next-generation physical parameterization schemes.

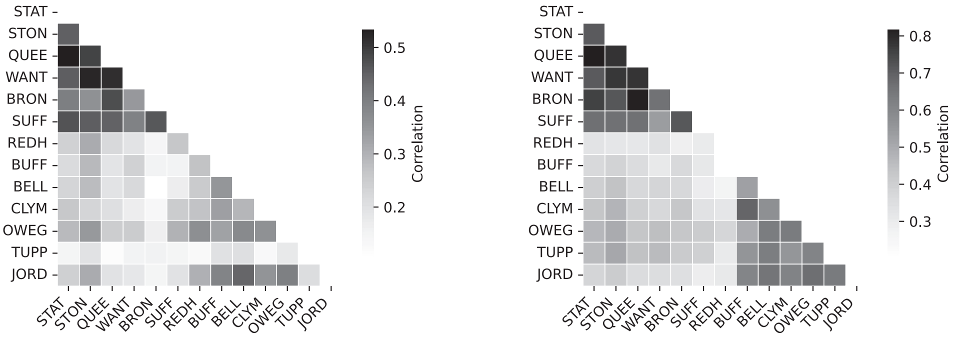
Finally, we examined the cross-station correlation of the coefficients to evaluate spatial consistency. We observed that coastal


(a) Lidar - Coefficient C0

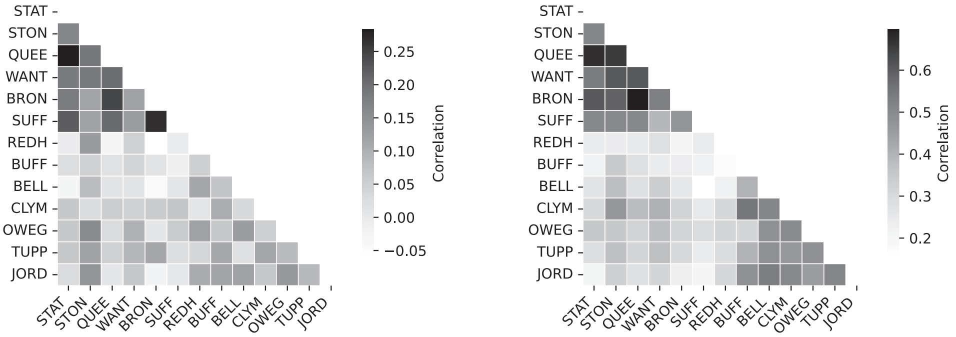
(b) NOW23 - Coefficient C0


(c) Lidar - Coefficient C1

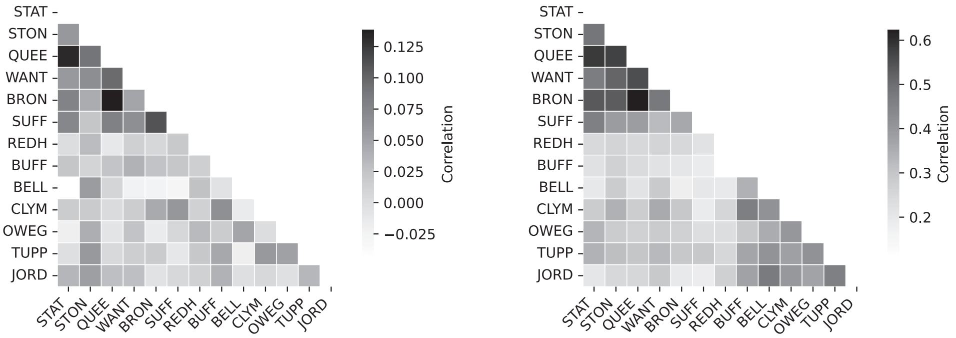
(d) NOW23 - Coefficient C1


(e) Lidar - Coefficient C2

(f) NOW23 - Coefficient C2


(g) Lidar - Coefficient C3

(h) NOW23 - Coefficient C3


(i) Lidar - Coefficient C4

(j) NOW23 - Coefficient C4

FIGURE 11 | Legend on next page.

FIGURE 11 | Comparison of heatmaps showing Pearson's correlation coefficients between pairs of stations, computed from Chebyshev coefficients derived from (left panels) lidar data and (right panels) NOW23 data. Panels (a, b) correspond to coefficient C_0, (c, d) to C_1, (e, f) to C_2, (g, h) to C_3, and (i,j) to C_4.

stations formed a distinct group, exhibiting strong correlation among themselves in the characteristics of mean wind speed, linear shear, curvature, and inflection-related features, which remained consistent across both datasets. In contrast, inland stations also formed a coherent group, showing strong correlation between the two datasets in terms of mean wind speed, linear shear, and curvature. However, the correlation for higher order fluctuations was stronger in NOW23 than in observations, indicating that the model may be overrepresenting these features. This analysis opens the way for extending the proposed framework to Measure-Correlate-Predict (MCP) approaches, such as identifying suitable long-term reference datasets [64] and selecting target stations based on measurements at reference stations [65].

The seamless capability of the proposed C-WiSPr framework to represent complete wind speed profiles using only five coefficients makes it suitable for various other practical applications. Notably, our approach can also identify nonmonotonic wind profiles, such as those exhibiting negative shear near the surface or alternating wind speed gradients with height, sometimes observed in offshore environments [66, 67]. These profiles can be effectively isolated using targeted coefficient thresholds (e.g., $C_1 < 0$, $C_2 > 0$, $C_3 < 0$, $C_4 > 0$), as shown in C1, demonstrating the framework's diagnostic potential beyond smooth or monotonic profile shapes. Future directions include machine learning-based gap-filling methods along vertical levels in observed wind profiles [68]; improved extrapolation of simulated wind speeds to buoy levels (typically at 3–4 m), which may outperform neutral log-law-based extrapolation [69]; extending wind profiles beyond the surface layer by integrating large-scale reanalysis data with machine learning techniques [70, 71]; generating and characterizing diverse inflow conditions to better understand unsteady blade aerodynamics and associated loads [72]; and ultimately improving short-term wind speed forecasting using machine learning models to support efficient wind power system operations [73].

Author Contributions

H.B. is responsible for conceptualization, data curation, formal analysis, investigation, methodology, visualization, and writing – original draft. S.B. is responsible for conceptualization, methodology, supervision, and writing – review and editing.

Acknowledgments

This research is made possible by the New York State (NYS) Mesonet. Original funding for the NYS Mesonet (NYSM) buildup was provided by Federal Emergency Management Agency grant FEMA-4085-DR-NY. The continued operation and maintenance of the NYSM is supported by National Mesonet Program, University at Albany, Federal and private grants, and others. We extend our gratitude to the NREL for publicly sharing the NOW23-simulated wind profiles data.

Funding

The authors are grateful for financial support from the State University of New York's Empire Innovation Program.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The datasets used in this study are publicly available. The NOW23-simulated wind profile data can be accessed at https://dx.doi.org/10.25984/1821404, while the NYSM profile lidar wind profile data can be requested at https://nysmesonet.org/weather/requestdata. The codes developed and used for complete analysis can be publicly accessible at https://github.com/HarishBaki/C-WiSPr.git.

References

1. Global Wind Energy Council. Global Wind Report 2024, 2024. https://www.gwec.net/. Accessed March 22, 2025.

2. P. Veers, K. Dykes, S. Basu, et al., “Grand Challenges: Wind Energy Research Needs for a Global Energy Transition,” *Wind Energy Science* 7, no. 6 (2022): 2491–2496, https://doi.org/10.5194/wes-7-2491-2022.

3. SANY SI 270 150 Wind Turbine. https://en.wind-turbine-models.com/turbines/2644-sany-si-270150; 2025. Accessed: 2025-07-21.

4. H. Qin, H. Yang, H. Li, and G. Du, “Highlights of key Advances in Chinas Wind Turbines Technology in 2024,” *Frontiers in Energy* 19 (2025): 18–27, https://doi.org/10.1007/s11708-025-0987-3.

5. N. Akhtar, B. Geyer, and C. Schrum, “Larger Wind Turbines as a Solution to Reduce Environmental Impacts,” *Scientific Reports* 14, no. 1 (2024): 6608.

6. R. Krishnamurthy, R. K. Newsom, C. M. Kaul, et al., “Observations of Wind Farm Wake Recovery at an Operating Wind Farm,” *Wind Energy Science* 10, no. 2 (2025): 361–380.

7. K. Wolz, C. Holst, F. Beyrich, E. Päschke, and M. Mauder, “Comparing Triple and Single Doppler Lidar Wind Measurements With Sonic Anemometer Data Based on a new Filter Strategy for Virtual Tower Measurements,” *Geoscientific Instrumentation, Methods and Data Systems* 13, no. 2 (2024): 205–223.

8. A. Visich and B. Conan, “Measurement and Analysis of High Altitude Wind Profiles Over the sea in a Coastal Zone Using a Scanning Doppler LiDAR: Application to Offshore Wind Energy,” *Ocean Engineering* 325 (2025): 120749.

9. B. Shrestha, J. A. Brotzge, J. Wang, et al., “Overview and Applications of the new York State Mesonet Profiler Network,” *Journal of Applied Meteorology and Climatology* 60, no. 11 (2021): 1591–1611.

10. B. Shrestha, J. A. Brotzge, and J. Wang, “Evaluation of the new York State Mesonet Profiler Network Data,” *Atmospheric Measurement Techniques* 15, no. 20 (2022): 6011–6033.

11. C. Ng and K. Hon, “Fast Dual-Doppler LiDAR Retrieval of Boundary Layer Wind Profiles,” *Weather* 77, no. 4 (2022): 134–142.

12. J. Steinheuer, F. Beyrich, and U. Löhnert, “Exploiting the Full Potential of Doppler Lidars: High-Resolution Wind-Gust Profiling in Significant Weather,” *Quarterly Journal of the Royal Meteorological Society* 151, no. 769 (2025): e4961.

13. J. He, P. Chan, Q. Li, and C. Lee, “Characterizing Coastal Wind Energy Resources Based on Sodar and Microwave Radiometer Observations,” *Renewable and Sustainable Energy Reviews* 163 (2022): 112498.

14. J. Ramon, L. Lledó, N. Pérez-Zanón, A. Soret, and F. J. Doblas-Reyes, “The Tall Tower Dataset: A Unique Initiative to Boost Wind Energy Research,” *Earth System Science Data* 12, no. 1 (2020): 429–439.

15. G. Gualtieri, “Analysing the Uncertainties of Reanalysis Data Used for Wind Resource Assessment: A Critical Review,” *Renewable and Sustainable Energy Reviews* 167 (2022): 112741.

16. H. Hersbach, B. Bell, P. Berrisford, et al., Complete ERA5 from 1940: Fifth generation of ECMWF atmospheric reanalyses of the global climate. https://doi.org/10.24381/cds.143582cf; 2017. Accessed on DD-MMM-YYYY.

17. S. Schimanke, M. Ridal, P. Le Moigne, et al., CERRA sub-daily regional reanalysis data for Europe on height levels from 1984 to present. https://doi.org/10.24381/cds.38b394e6; 2021. Accessed on DD-MMM-YYYY.

18. M. Dörenkämper, B. T. Olsen, B. Witha, et al., “The Making of the new European Wind Atlas–Part 2: Production and Evaluation,” *Geoscientific Model Development Discussions* 2020 (2020): 1–37.

19. A. N. Hahmann, T. Sile, B. Witha, et al., “The Making of the new European Wind Atlas – Part 1: Model Sensitivity,” *Geoscientific Model Development* 13, no. 10 (2020): 5053–5078.

20. E. Gleeson, E. Whelan, and J. Hanley, “Met Éireann High Resolution Reanalysis for Ireland,” *Advances in Science and Research* 14 (2017): 49–61.

21. I. Wijnant, B. Van Ulft, B. Van Stratum, et al., The Dutch Offshore Wind Atlas (DOWA): Description of the dataset. *Royal Netherlands Meteorological Institute, Ministry of Infrastructure and Water Management, De Bilt* 2019.

22. I. M. Solbrekke and A. Sorteberg, “NORA3-WP: A High-Resolution Offshore Wind Power Dataset for the Baltic, North, Norwegian, and Barents Seas,” *Scientific Data* 9, no. 1 (2022): 362.

23. N. Bodini, M. Optis, S. Redfern, et al., “The 2023 National Offshore Wind Data set (NOW-23),” *Earth System Science Data* 16, no. 4 (2024): 1965–2006.

24. S. Basu, J. F. Vinuesa, and A. Swift, “Dynamic LES Modeling of a Diurnal Cycle,” *Journal of Applied Meteorology and Climatology* 47, no. 4 (2008): 1156–1174.

25. C. Srinivas, R. Venkatesan, and A. B. Singh, “Sensitivity of Mesoscale Simulations of Land–sea Breeze to Boundary Layer Turbulence Parameterization,” *Atmospheric Environment* 41, no. 12 (2007): 2534–2548.

26. H. Baki, S. Basu, and G. Lavidas, “Modeling Frontal low-Level Jets and Associated Extreme Wind Power Ramps Over the North sea,” *Wind Energy Science* 10, no. 8 (2025): 1575–1609, https://doi.org/10.5194/wes-10-1575-2025.

27. C. G. Nunalee and S. Basu, “Mesoscale Modeling of Coastal low-Level Jets: Implications for Offshore Wind Resource Estimation,” *Wind Energy* 17, no. 8 (2014): 1199–1216.

28. H. Baki, S. Basu, and G. Lavidas, “Estimating the Offshore Wind Power Potential of Portugal by Utilizing Gray-Zone Atmospheric Modeling,” *Journal of Renewable and Sustainable Energy* 16, no. 6 (2024): 063306.

29. E. N. Lorenz, “Deterministic Nonperiodic Flow,” *Journal of the Atmospheric Sciences* 20, no. 2 (1963): 130–148.

30. P. C. Chu, “Two Kinds of Predictability in the Lorenz System,” *Journal of the Atmospheric Sciences* 56, no. 10 (1999): 1427–1432.

31. S. Al-Yahyai, Y. Charabi, and A. Gastli, “Review of the use of Numerical Weather Prediction (NWP) Models for Wind Energy Assessment,” *Renewable and Sustainable Energy Reviews* 14, no. 9 (2010): 3192–3198.

32. B. Storm, J. Dudhia, S. Basu, A. Swift, and I. Giammanco, “Evaluation of the Weather Research and Forecasting Model on Forecasting low-Level Jets: Implications for Wind Energy,” *Wind Energy* 12, no. 1 (2009): 81–90.

33. R. Floors, C. L. Vincent, S. E. Gryning, A. Pea, and E. Batchvarova, “The Wind Profile in the Coastal Boundary Layer: Wind Lidar Measurements and Numerical Modelling,” *Boundary-Layer Meteorology* 147, no. 3 (2013): 469–491.

34. S. E. Gryning, E. Batchvarova, R. Floors, et al., “Long-Term Profiles of Wind and Weibull Distribution Parameters up to 600 m in a Rural Coastal and an Inland Suburban Area,” *Boundary-Layer Meteorology* 150, no. 2 (2014): 167–184.

35. A. Gevorgyan, “A Case Study of low-Level Jets in Yerevan Simulated by the WRF Model,” *Journal of Geophysical Research* 123, no. 1 (2018): 300–314.

36. P. C. Kalverla, J. B. Duncan, G. J. Steeneveld, and A. A. M. Holtslag, “Low-Level Jets Over the North sea Based on ERA5 and Observations: Together They Do Better,” *Wind Energy Science* 4, no. 2 (2019): 193–209.

37. S. Brune, J. D. Keller, and S. Wahl, “Evaluation of Wind Speed Estimates in Reanalyses for Wind Energy Applications,” *Advances in Science and Research* 18 (2021): 115–126.

38. V. Pronk, N. Bodini, M. Optis, et al., “Can Reanalysis Products Outperform Mesoscale Numerical Weather Prediction Models in Modeling the Wind Resource in Simple Terrain?,” *Wind Energy Science* 7, no. 2 (2022): 487–504.

39. J. Sward, T. Ault, and K. Zhang, “Spatial Biases Revealed by LiDAR in a Multiphysics WRF Ensemble Designed for Offshore Wind,” *Energy* 262 (2023): 125346.

40. P. Pentikinen, E. J. O'Connor, and P. Ortiz-Amezcua, “Evaluating Wind Profiles in a Numerical Weather Prediction Model With Doppler Lidar,” *Geoscientific Model Development* 16, no. 8 (2023): 2077–2094.

41. E. Cheynet, J. M. Diezel, H. Haakenstad, Ø. Breivik, A. Peña, and J. Reuder, “Tall Wind Profile Validation Using Lidar Observations and Hindcast Data,” *Wind Energy Science Discussions* 2024 (2024): 1–29.

42. B. Storm and S. Basu, “The WRF Model Forecast-Derived low-Levelwind Shear Climatology Over the United States Great Plains,” *Energies* 3, no. 2 (2010): 258–276.

43. O. Krogster and J. Reuder, “Validation of Boundary Layer Parameterization Schemes in the Weather Research and Forecasting Model Under the Aspect of Offshore Wind Energy Applications Part I: Average Wind Speed and Wind Shear: WRF Validation for Offshore Wind Energy Applications-Part I,” *Wind Energy* 18, no. 5 (2015): 769–782.

44. M. C. Holtslag, W. A. A. M. Bierbooms, and G. J. W. Van Bussel, “Extending the Diabatic Surface Layer Wind Shear Profile for Offshore Wind Energy,” *Renewable Energy* 101 (2017): 96–110.

45. S. A. Mata, J. J. Pena Martínez, J. Bas Quesada, et al., “Modeling the Effect of Wind Speed and Direction Shear on Utility-Scale Wind Turbine Power Production,” *Wind Energy* 27, no. 9 (2024): 873–899.

46. P. Durn, S. Basu, C. Meiner, and M. S. Adaramola, “Automated Classification of Simulated Wind Field Patterns From Multiphysics Ensemble Forecasts,” *Wind Energy* 23, no. 4 (2020): 898–914.

47. M. J. Otte and J. C. Wyngaard, “A General Framework for an Unmixed Layer PBL Model,” *Journal of Atmospheric Sciences* 53, no. 18 (1996): 2652–2670.

48. M. Sommerfeld, M. Dörenkämper, J. De Schutter, and C. Crawford, “Offshore and Onshore Ground-Generation Airborne Wind Energy Power Curve Characterization,” *Wind Energy Science Discussions* 2020 (2020): 1–39.

49. E. Malz, V. Verendel, and S. Gros, “Computing the Power Profiles for an Airborne Wind Energy System Based on Large-Scale Wind Data,” *Renewable Energy* 162 (2020): 766–778.

50. R. K. Newsom and R. M. Banta, "Shear-Flow Instability in the Stable Nocturnal Boundary Layer as Observed by Doppler Lidar During CASES-99," *Journal of the Atmospheric Sciences* 60, no. 1 (2003): 16–33.

51. B. J. Jonkman. TurbSim user's guide. tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States): 2006.

52. J. S. Na, E. Koo, E. K. Jin, et al., "Large-Eddy Simulations of Wind-Farm Wake Characteristics Associated With a low-Level jet," *Wind Energy* 21, no. 3 (2018): 163–173.

53. Y. Chen, Z. Wang, Y. Gao, Z. Li, and B. Liu, "Application of Chebyshev Approaching Function in a Direct-Detection Wind Doppler Lidar Data Processing," *Optik* 179 (2019): 1049–1056.

54. T. J. Rivlin, *Chebyshev Polynomials* (Courier Dover Publications, 2020).

55. P. Mildner, "ber die Reibung in einer speziellen Luftmasse in den untersten Schichten der Atmosphre," *Beitr Phys Freien Atmos* 19 (1932): 151–158.

56. R. H. Clarke, "The Wangara Experiment: Boundary Layer Data," *Division of Meteorological Physics Technical Paper* 19 (1971): 13–14.

57. A. Peña, R. Floors, and S. E. Gryning, "The Høvsøre Tall Wind-Profile Experiment: A Description of Wind Profile Observations in the Atmospheric Boundary Layer," *Boundary-Layer Meteorology* 150 (2014): 69–89.

58. E. J. McCabe and J. M. Freedman, "Development of an Objective Methodology for Identifying the sea-Breeze Circulation and Associated low-Level jet in the new York Bight," *Weather and Forecasting* 38, no. 4 (2023): 571–589.

59. N. Bodini, M. Optis, S. Redfern, et al., "The 2023 National Offshore Wind Data set (NOW-23)," *Earth System Science Data Discussions* 2023 (2023): 1–57.

60. J. A. Brotzge, J. Wang, C. Thorncroft, et al., "A Technical Overview of the new York State Mesonet Standard Network," *Journal of Atmospheric and Oceanic Technology* 37, no. 10 (2020): 1827–1845.

61. J. Gottschall, M. Courtney, R. Wagner, H. E. Jørgensen, and I. Antoniou, "Lidar Profilers in the Context of Wind Energy–a Verification Procedure for Traceable Measurements," *Wind Energy* 15, no. 1 (2012): 147–159.

62. T. Klaas-Witt and S. Emeis, "The Five Main Influencing Factors for Lidar Errors in Complex Terrain," *Wind Energy Science* 7, no. 1 (2022): 413–431.

63. J. A. Carta, S. Velázquez, and P. Cabrera, "A Review of Measure-Correlate-Predict (MCP) Methods Used to Estimate Long-Term Wind Characteristics at a Target Site," *Renewable and Sustainable Energy Reviews* 27 (2013): 362–400.

64. S. Weekes, A. Tomlin, S. Vosper, A. Skea, M. Gallani, and J. Standen, "Long-Term Wind Resource Assessment for Small and Medium-Scale Turbines Using Operational Forecast Data and Measure–Correlate–Predict," *Renewable Energy* 81 (2015): 760–769.

65. S. Weekes and A. Tomlin, "Comparison Between the Bivariate Weibull Probability Approach and Linear Regression for Assessment of the Long-Term Wind Energy Resource Using MCP," *Renewable Energy* 68 (2014): 529–539.

66. C. L. Archer, B. A. Colle, D. L. Veron, F. Veron, and M. J. Sienkiewicz, "On the Predominance of Unstable Atmospheric Conditions in the Marine Boundary Layer Offshore of the US Northeastern Coast," *Journal of Geophysical Research: Atmospheres* 121, no. 15 (2016): 8869–8885.

67. M. Golbazi and C. L. Archer, "Methods to Estimate Surface Roughness Length for Offshore Wind Energy," *Advances in Meteorology* 2019, no. 1 (2019): 5695481.

68. F. Kristianti, J. Dujardin, F. Gerber, H. Huwald, S. W. Hoch, and M. Lehning, "Combining Weather Station Data and Short-Term Lidar Deployment to Estimate Wind Energy Potential With Machine Learning: A Case Study From the Swiss Alps," *Boundary-Layer Meteorology* 188, no. 1 (2023): 185–208.

69. D. Carvalho, A. Rocha, M. Gómez-Gesteira, and C. S. Santos, "WRF Wind Simulation and Wind Energy Production Estimates Forced by Different Reanalyses: Comparison With Observed Data for Portugal," *Applied Energy* 117 (2014): 116–126.

70. H. Baki and S. Basu, "Estimating High-Resolution Profiles of Wind Speeds From a Global Reanalysis Dataset Using TabNet," *Environmental Data Science* 3 (2024): e32.

71. B. Liu, X. Ma, J. Guo, et al., "Extending the Wind Profile Beyond the Surface Layer by Combining Physical and Machine Learning Approaches," *Atmospheric Chemistry and Physics* 24, no. 7 (2024): 4047–4063.

72. J. Park, S. Basu, and L. Manuel, "Large-Eddy Simulation of Stable Boundary Layer Turbulence and Estimation of Associated Wind Turbine Loads," *Wind Energy* 17, no. 3 (2014): 359–384.

73. Z. Liu, P. Jiang, L. Zhang, and X. Niu, "A Combined Forecasting Model for Time Series: Application to Short-Term Wind Speed Forecasting," *Applied Energy* 259 (2020): 114137.

74. V. Cherkassky and X. Shao, "Signal Estimation and Denoising Using VC-Theory," *Neural Networks* 14, no. 1 (2001): 37–52.

Appendix A

Identifying Optimal Polynomial Order

To identify the optimal polynomial order for the Chebyshev approximation, we adopted the Vapnik-Chervonenkis (VC) theory as a model selection strategy [74]. Traditional model selection methods typically assume infinite sample size and estimate the prediction risk as the multiplication of the empirical risk (i.e., training error) and a penalty term based on model complexity. Wherease, the VC framework is applicable for finite sample size but introduces a data-dependent penalization term known as Vapnik's measure (VM), given by the following empirical expression (Equation 10 in [74]):

$$r(p,n) = \left(1 - \sqrt{p - p\ln p + \frac{\ln n}{2n}}\right)^{-1}, \tag{A1}$$

where $p = \frac{m+1}{n}$, m is the polynomial order, and n is the number of height levels (or data points) in the profile.

We used root mean squared error (RMSE) as a measure of empirical risk, defined as:

$$R_{\text{emp}}(m) = \frac{1}{n}\sum_{z=1}^{n}\left(U_z - f(m,z)\right)^2, \tag{A2}$$

where $f(m,z)$ represents the Chebyshev polynomial approximation of order m at height level z.

By multiplying the empirical risk $R_{\text{emp}}(m)$ with the penalty factor $r(p,n)$, we computed the estimated prediction risk for each polynomial order. The risk was evaluated for wind speed profiles at all stations, using only those profiles with complete height levels (i.e., $n = 18$) for consistency. The resulting mean prediction risk as a function of polynomial order is shown in Figure A1.

From Figure A1, it is evident that polynomial order 5 yields the lowest prediction risk, suggesting it as the optimal order. However, upon inspecting the Chebyshev coefficients, we found that C_4 and C_5 contribute marginally to the overall approximation. Moreover, the difference in prediction risk between orders 4 and 5 is quite small. Therefore, to balance model accuracy with simplicity and interpretability, we opted for polynomial order 4 in this study.

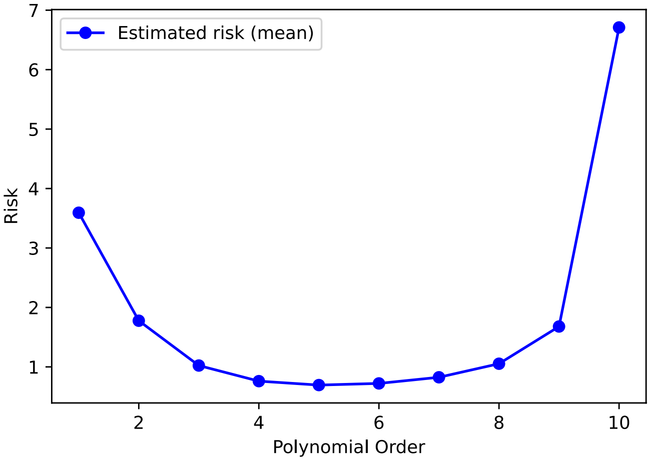


FIGURE A1 | Estimated prediction risk as a function of polynomial order.

Appendix B

Selecting Optimum Height Levels for Data Processing

Lidar wind profiles often include missing data at various height levels. For Chebyshev approximation, we selected fourth-order polynomials, which require at least five valid height levels. However, five points are insufficient to capture complex features like low-level jets (see Figure 2d). To ensure reliable approximation, we defined the following criteria for selecting valid profiles: (i) valid wind speed at 10 m, (ii) at least two valid points between 100–200 m, (iii) at least four between 225–375 m, and (iv) at least two between 400–500 m. This typically yields nine valid levels per profile.

These conditions are physically motivated. Chebyshev approximation benefits from valid values near boundaries (e.g., 10 m and 500 m). While 10 m is usually available, 500 m is often missing due to lidar limitations (e.g., aerosol scarcity, precipitation). Furthermore, the NOW23 dataset includes only 400 and 500 m levels at the upper end; hence, our top-level criterion ensures consistency across sources. The 100-200 m layer often shows a steady increase in wind speed, which can be captured with two values, whereas the 225-375 m layer typically includes inflection points and thus needs more levels.

To validate this setup, we used the Wantagh-1 low-level jet profile and generated 3500 combinations by selectively sampling valid height levels ($1 \times \binom{5}{2} \times \binom{7}{4} \times \binom{5}{2} = 3500$). As shown in Figure B1, the profiles fitted with only nine levels (gray) closely follow the full-level fit (black). The best match (red dashed) nearly overlaps with the full-profile fit, confirming that our conditional selection reliably preserves profile structure.

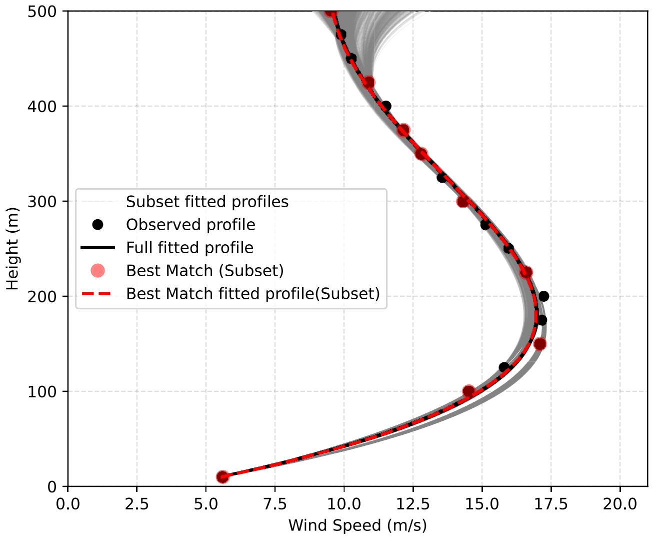


FIGURE B1 | Illustration of the observed Wantagh-1 low-level jet profile (red solid circles), all fitted profiles satisfying the adopted height-level conditions by selectively removing levels (gray lines), the reference profile fitted using all available height levels (black line), and the best-matching subset profile that closely replicates the full-level fit (red dashed line).

Appendix C

Limitation of 4th Order Polynomial in Capturing Non-Monotonic Wind Profiles

Figure C1 shows examples of non-monotonic wind speed profiles from lidar observations across multiple stations. These cases were flagged by fitting 4th-order Chebyshev polynomials and applying coefficient-based criteria. The red lines are the 4th-order fits, while the green lines are the 10th-order fits. As expected, 4th-order polynomials can only capture profiles with at most two inflection points (or two local minima/maxima), so they often fall short when the profile has more complex structure. Higher order polynomials, like the 10th order used here, have a much better chance of following these complicated patterns and, in this case, manage to match 6 out of the 8 examples quite well. That said, profiles with strong fluctuations, such as panels (d) and (h), are still difficult to capture even with the 10th-order fit. Overall, the 4th-order fits do a decent job of following the general shape, even though they miss some of the finer local fluctuations.

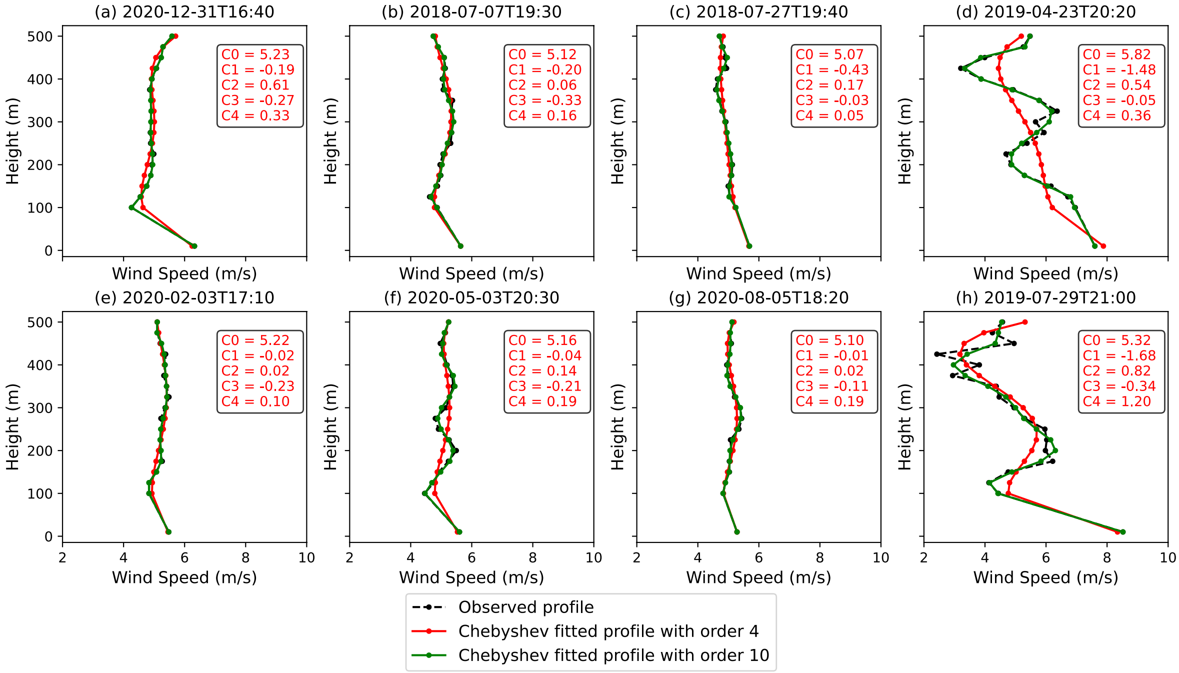


FIGURE C1 | Illustration of non-monotonic wind profiles from lidar observations across all the stations. These profiles were identified by fitting 4th order Chebyshev polynomials, then applying a conditional search by setting combinations of C1 < 0, C2 > 0, C3 < 0, and C4 > 0. The Chebyshev coefficients or 4th order are shown in text box of each subplot. The approximations of these identified profiles with 4th order (red line) and 10th order (green line) are illustrated in each subplot.