Performance comparison of NWP model products for wind energy assessment in complex terrain

By Sean Butters
University of Colorado Boulder

A thesis submitted to the University of Colorado Boulder
In partial fulfillment of the requirements to receive
Honors designation in Environmental Studies

Defense date: April 6, 2022

Thesis committee
Julie Lundquist, Department of Atmospheric and Oceanic Sciences, Primary Advisor
Andrew Winters, Department of Atmospheric and Oceanic Sciences, Outside Reader
Colleen Scanlan Lyons, Department of Environmental Studies, Honors Council Representative

Abstract

Wind energy has seen impressive growth in recent decades and will fulfill a greater portion of the world's energy mix as decarbonization efforts ramp up. Developers must accurately assess the wind resource at potential project sites. Minimizing errors in preliminary predictions is critical for keeping wind energy viable, especially with expansion into complex terrain. Complex terrain introduces a number of complexities that might eliminate the possibility of an on-site assessment, meaning that numerical weather prediction (NWP) models, such as reanalysis products (ERA-5 or MERRA-2) or high-resolution simulations (WTK-LED/WRF) may be used. WTK is generally expected to perform better than ERA-5 or MERRA-2 because of the additional computing power required to extract datasets, and comparing these models to three observational sites at the WFIP2 project in the Columbia River Gorge confirms WTK's strength in complex terrain. None of these products were strong performers in complex terrain when compared to lidar observations (CC max ~ 0.7) and WTK, ERA-5, and MERRA-2 all have negative biases when compared to observations. ERA-5 has the largest average bias of ~ -3 m/s. WTK had the smallest average bias, ~ -1 m/s. WTK had specific strengths when compared to ERA-5 and MERRA-2, such as its low bias in summer months. WTK also forecasted the diurnal wind cycles best, despite exaggerating the cycle magnitude.

1 Introduction

The necessity of access to clean and reliable energy stands at an all-time high, with drivers such as fossil fuel market volatility, supply vulnerability, and concerns over climate change (Asif & Muneer, 2007). The Klass model projects that known oil and gas reserves will be extinguished within the next 50 years, so there is an acute sense of urgency about meeting humanity's energy needs (Shafiee & Topal, 2009).

The development of renewable energy has been a marker for global transformation, and wind power is an important piece of the puzzle. Global investment in wind power now exceeds $109 billion annually and continuously growing, solidifying this industry as one of the most rapidly-expanding industries worldwide (*Wind in Numbers | GWEC*, 2022). The potential to supply the electrical grid with renewables is increasing with the emergence of new platforms, such as the offshore wind industry, which has had a steady and unprecedented 36% annual growth since 2001 (Rodrigues et al., 2015). Investment in renewables has also been helped by federal and state government energy incentives, which have primarily benefited wind (Wiser & Bolinger, 2019).

This swift progression in the wind energy industry requires that developers accurately predict wind resources at new project locations (Brower, 2012). Developers habitually overpredict the wind resource at potential sites, and minimizing errors in such estimations remains one of the largest issues with wind energy development (J. C. Y. Lee & Fields, 2021). The expansion of wind energy into offshore and complex terrain locations even further necessitates a precise and accurate estimate of wind speeds. The preferred way to conduct a wind resource assessment campaign is to measure turbine hub-height wind speeds with meteorological towers or remote-sensing devices, such as lidars or sodars, that are placed on-site. However, the cost and person-hours required for conducting field campaigns to assemble meteorological masts or install remote sensing devices consistently exceeds projections (Sheppard, 2009). Remote sensing instruments for wind energy are also extremely expensive and are typically owned and deployed by only a few specific institutions (Goit et al., 2019). On top of this, developers must operate with constraints on time, electrical power, and physical topography (Pronk et al., 2021). Alternatives to an in-situ assessment campaign are NWP products such as reanalysis (ERA-5 or MERRA-2) or finer resolution mesoscale simulations.

Global reanalysis products serve as important tools that have numerous meteorological and industry applications and are especially useful for wind energy purposes. For over 30 years, organizations have collectively worked to create climate and weather reanalyses that use historical data (Compo et al., 2011). Two reanalysis products that are commonly used right now are ERA-5 and MERRA-2. ERA-5 is provided by the European Center for Medium-range Weather Forecasts (ECMWF). The spatial resolution of reanalysis products is greatly improving, as demonstrated by ERA-5, whose ~1 degree latitude/31 km horizontal resolution and 137 vertical pressure levels provide finer horizontal and vertical resolution than its predecessor, ERA-interim, which provided horizontal resolution of 80 km and only 37 pressure levels (Hersbach et al., 2020). MERRA-2 comes from the Global Modeling and Assimilation Office (GMAO) at the National Aeronautics and Space Administration (NASA) and operates in between ERA-5 and ERA-interim with a resolution of ~2 degrees latitude/60 km and 42 pressure levels (Bosilovich et al., 2016; Jourdier, 2020). Despite these improvements, ERA-5 and MERRA-2's spatial and temporal resolution can still fail to capture certain critical meteorological phenomena, which could result in an inaccurate wind speed prediction on the wind farm scale. Some advantages, though, of these reanalysis products, is that they can be easily queried by anybody and accessed via a simple download.

Use of reanalysis for wind energy purposes has been considered for more than 20 years (Schwartz et al., 1999). Some developers utilize reanalysis as a preliminary assessment for certain sites before conducting a full-scale assessment campaign (Ayik et al., 2021; Samal, 2021). Developers speculate about whether reanalysis products are reliable enough to be the basis for an entire wind resource assessment campaign, and expansion of the wind energy industry into offshore locations has only increased debates about reanalysis suitability (Ahmad et al., 2022; Gualtieri, 2021). Typically, ERA-5 has outperformed MERRA-2 for wind energy applications in simple terrain (Fan et al., 2021; Gruber et al., 2022; Ramon et al., 2019). Few studies have assessed these products in complex terrain, and a deeper hub-height validation study of reanalysis, especially in complex terrain locations, where wind energy has great potential to expand, could boost trust in their use.

Advanced, high-resolution NWP modeling can provide accurate wind speed data, but running the simulations is quite costly and resource-intensive. Their main benefit comes from their ability to utilize a reanalysis product as a substratum for boundary conditions and refine this with additional computing power for a higher resolution. They also resolve a range of scales of motion at a particular site for an overall rigorous prediction involving a variety of scales of motion. More refined mesoscale modeling has been considered for wind energy resource assessment since before 2010, and improvements in its resolution have yielded even more analyses of its performance (Al-Yahyai et al., 2010; Castorrini et al., 2021; Fernández-González et al., 2018). There have also been efforts to apply NWP models across the United States and Europe and create a wind speed reference dataset (Dörenkämper et al., 2020; Draxl et al., 2015; Hahmann et al., 2020). Specific wind resource assessment campaigns that primarily used NWP have occurred at many sites globally, including Malaysia (Nor et al., 2014), Fiji (Dayal et al., 2021), Greece (Giannaros et al., 2017), Thailand (Chancham et al., 2017), Alaska (J. A. Lee et al., 2019), and Portugal (Salvação & Guedes Soares, 2018). NWP has further been used in complex terrain and offshore campaigns (Indasi et al., 2017; Rybchuk, Optis, et al., 2021).

As a whole, as wind energy continues to expand, investigation about placing turbines into complex terrain will continue and new resource assessment campaigns will be necessary. The aforementioned issues with completing an in-situ assessment with meteorological instruments are exacerbated when topography, such as mountains or coastlines, is introduced as an additional factor. NWP and reanalysis are alternatives for wind predictions, and it is important to

know whether easily-obtainable reanalysis products, such as ERA-5 and MERRA-2, are capable of predicting wind speeds as precisely as NWP. Preliminary research shown in Pronk, et al., 2021 demonstrated similar performance between ERA-5 and the Weather Research and Forecasting (WRF) NWP model in simple terrain. This superior performance by the less expensive tool is highly surprising, given the magnitude of computational power required for WRF. To expand this research and test reanalysis and NWP in complex terrain, I performed a model validation at three complex terrain sites and analyzed the performance of ERA-5, MERRA-2, and more advanced simulations in a complex terrain scenario by comparing these models to available surface observations.

2 Data and Methods

I utilized locations with observational data from the Second Wind Forecast Improvement Project (WFIP2), a four year study initiated by the U.S. Department of Energy (DOE) to expand industry understanding of meteorological phenomena and model performance in complex terrain. The WFIP2 campaign took place in the Columbia River Gorge, bordering Washington and Oregon, with the overall goal of improving wind energy forecasting (Shaw et al., 2019). This experiment took place as part of the Atmosphere to Electrons (A2e) initiative, which partnered the DOE with several national laboratories and federal agencies to optimize predictions of energy production from wind farms.

2.1 Observations

The WFIP2 campaign had 27 total observational sites with varying combinations of surface instruments at each (Wilczak et al., 2019). Three observational sites containing lidars are utilized for this analysis (Figure 1). Observational data from heights below 60 meters and higher than 140 meters above ground level (AGL) will not be considered in this analysis because of the limitations of available heights from model products and tailoring this comparison for wind energy applications (Table 1). Turbine hub heights in the region are typically 80 meters AGL, meaning the rotor disc spans 40 meters to 120 meters AGL. Additionally, as shown in Table 2, observational data that is missing, unavailable, or not a number (NaN) are linearly interpolated if there is surrounding data within 30 minutes (Figure 2). Then, observations taken over time intervals shorter than 1 hour are resampled using an hourly average for comparison to hourly model data.

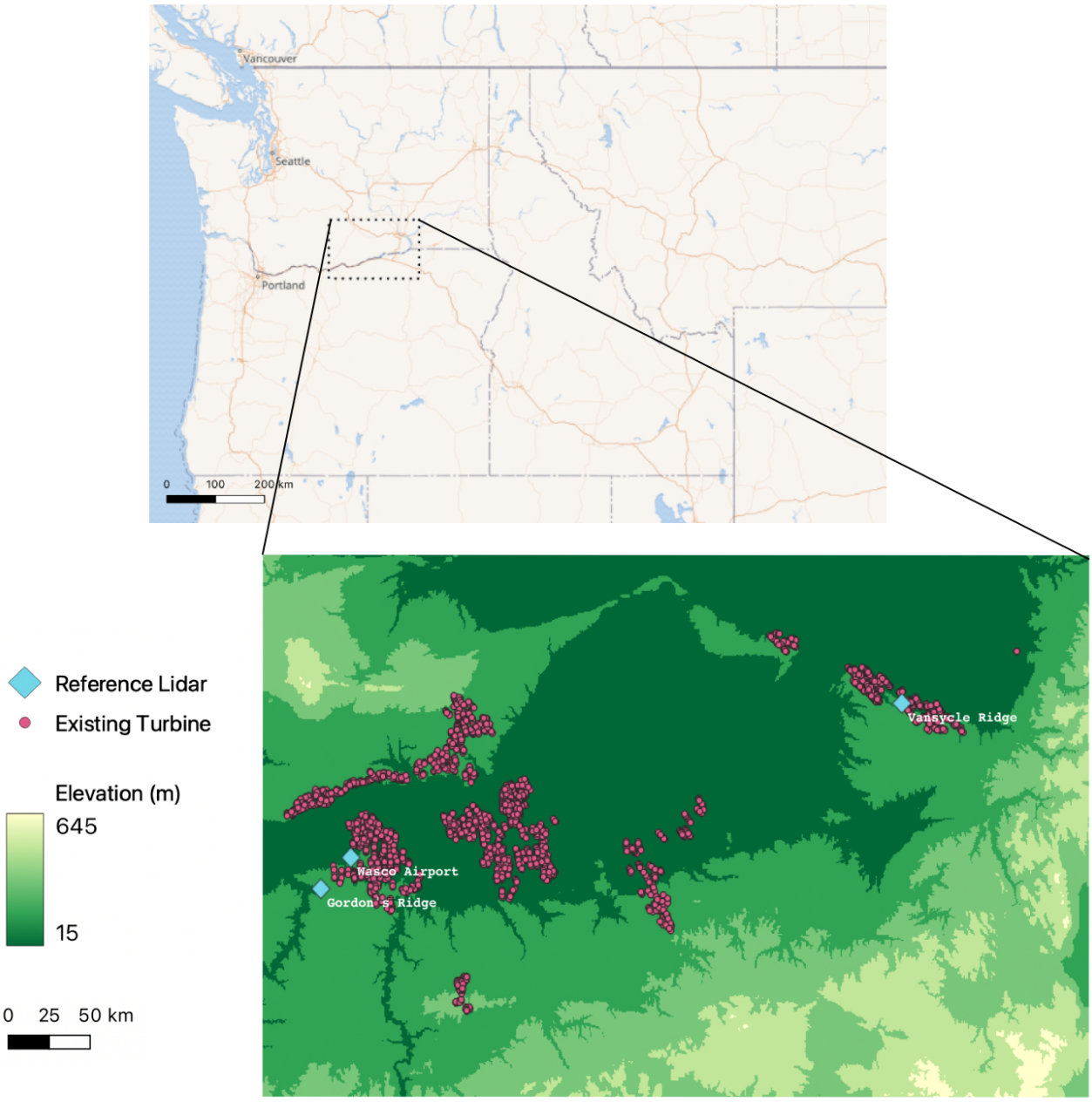

Figure 1: Map of the WFIP2 project location and observational sites

I consider observations from Windcube V1 and V2 scanning lidars. These scanning devices capture wind speeds using conical scans and measuring backscatter off of aerosols, a method that manufacturers report is accurate to 0.1 m/s (Aitken et al., 2012). The WindCube V2 gauges

Table 1: Heights and time intervals available for observations and models used in comparison

	Heights available (meters AGL)	Time interval
Gordon's Ridge (GOR) Lidar	40, 60, 80, 100, 120, 140, 160, 180, 200, 220, 240, 260	2 minutes (resampled hourly)
Wasco Airport (WAS) Lidar	40, 60, 80, 100, 120, 140, 160, 180, 200	10 minutes (resampled hourly)
Vansycle Ridge (VAN) Lidar	40, 60, 80, 100, 120, 140, 160, 180, 200	2 minutes (resampled hourly)
WRF: WTK-LED	60, 80, 100, 120, 140, 160, 180, 200	hourly
ERA-5	10, 31, 54, 66.5*, 79, 106, 121.5* 137, 170, 205, 245, 288, 334 *= interpolated from surrounding 2 heights	hourly
MERRA-2	40, 60, 80, 100, 120, 140	hourly
Used in comparison	**60, 80, 100, 120, 140**	**hourly**

line-of-sight velocity with a zenith angle of 28° across 4 directions at 1 Hz, which is improved upon slightly in the WindCube V2, which captures 5 directions (Bodini et al., 2019).

Table 3 summarizes the dates that observations were taken at these three sites. The first reference site is Gordon's Ridge, Washington (GOR), at an elevation of 723 meters above sea level (ASL). This site has available hub-height wind speed data from the University of Colorado's Windcube V2 that was active from November 16, 2015 to March 9, 2017. The second reference site is Wasco State Airport, Oregon (WAS) at an elevation of 460 meters ASL. This site has hub-height wind speed data from multiple lidars: one WindCube 200s scanning lidar from the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) and another Windcube V1 Profiler from the University of Colorado. This analysis will consider reviewed data from the Windcube V1 that was on-site from February 23, 2016 to January 27, 2017. The third site is Vansycle Ridge, Oregon (VAN) at an elevation of 532 meters ASL. Relevant wind speed observations at this location come from a WindCube V2 provided by the Lawrence Livermore National Laboratory (LLNL) that was active from March 9, 2016 to April 1, 2017.

2.2 ERA-5 Reanalysis

ERA-5 is one reanalysis product used for comparison against WTK-LED. ERA-5 data was retrieved from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu). ERA-5 features a horizontal resolution of ~1 degree latitude (31 km) with 137 vertical pressure levels.

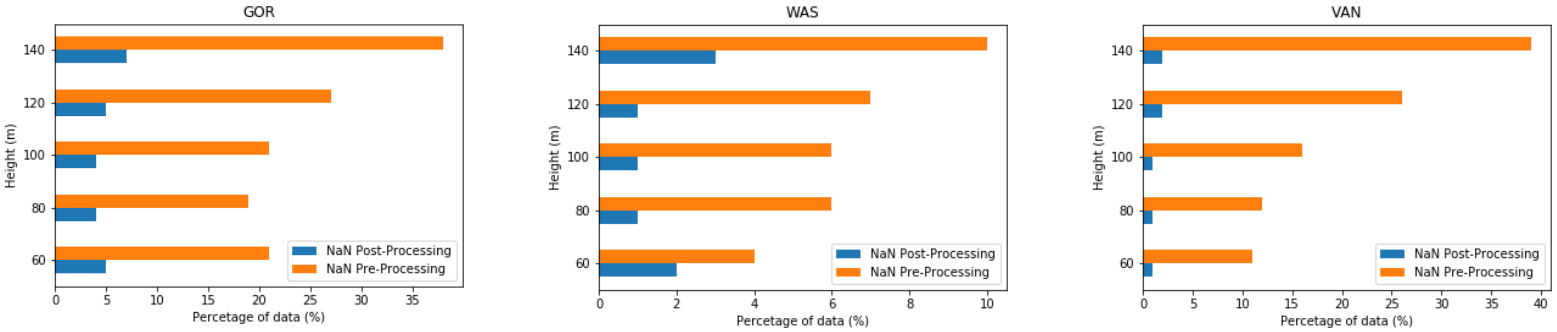

Figure 2: Data availability for each observational site before and after processing NaNs using a linear interpolation from usable points within 30 minutes

For this analysis, I used the closest ERA-5 latitude-longitude grid point to the exact location of each corresponding observational site, which has been proven to result in the best reanalysis performance (Sheridan et al., 2020). The grid points and respective distances from sites are summarized further in Table 4. Some ERA-5 data between two heights were interpolated to closely match heights from other model products, as shown in Table 1.

2.3 MERRA-2 Reanalysis

The second reanalysis product used here is MERRA-2. MERRA-2 data can be run and extracted from Renewables.Ninja (https://www.renewables.ninja/). MERRA-2 has a resolution of ~2 degrees of latitude (60 km) and 42 pressure levels. Like ERA-5, for this project, I used data from the closest longitude-latitude grid point to the 3 surface observation sites.

2.4 Numerical Weather Prediction: WRF/WTK-LED

The WTK-LED data for the three sites at WFIP2 came from the Wind Integration National Dataset (WIND) Toolkit Long-term Ensembel Dataset (WTK-LED), provided by the National Renewable Energy Laboratory (NREL). The data were extracted by NREL's Nicola Bodini, who personally performed the WTK-LED simulations for the purpose of this research. The attributes of the WRF runs are summarized in the table below and have the same characteristics as those mentioned in Pronk, et al., 2021. The model output was initialized monthly, 2 days prior to and up to 1 day after the end of every month. The first day is for the model to spin-up and the second and final days are for combining the months. The closest 2-km grid point is used for WRF.

Table 3: Time periods that each observational site was active

	Data time period:
Gordon's Ridge (GOR)	2015-11-16 to 2017-03-09
Wasco Airport (WAS)	2016-02-23 to 2017-01-27
Vansycle Ridge (VAN)	2016-03-09 to 2017-04-01

Table 4: Locations of each observational site and corresponding grid points used for analysis

	Site Exact Location (°)	Coordinates used for ERA-5 and MERRA-2 (°)	Reanalysis point distance from observational site (km)
Gordon's Ridge (GOR)	(45.51581,-120.78040)	(46, -121)	56.53
Wasco Airport (WAS)	(45.59011,-120.67193)	(46, -121)	52.25
Vansycle Ridge (VAN)	(45.95509,-118.68763)	(46, -119)	24.68

Table 5: WTK-LED WRF simulation characteristics

WRF Version	4.2.1
Spatial resolution	2 km
Temporal resolution	5 minutes (resampled hourly)
Heights (meters AGL)	12, 34, 52, 69, 86, 107, 134, 165, 200
Atmospheric forcing	ERA-5
Atmospheric nudging	Spectral nudging
Planetary boundary layer (PBL) scheme	Mellor-Yamada-Nakanishi-Niino Level 2.5 (Nakanishi & Niino, 2009)
Microphysics	Morrison double-moment (Morrison et al., 2009)
Longwave & shortwave radiation	Rapid radiative transfer model (Iacono et al., 2008)
Topography database	Global multi resolution terrain elevation data from the United States Geological Survey (USGS) and the National Geospatial-Intelligence Agency (NGIA)
Land-use data	Moderate Resolution Imaging Spectroradiometer 30 s

2.5 Error Metrics

In order to quantify the performance of WTK-LED against reanalysis products, four key metrics were utilized as recommended in (Optis et al., 2020). These metrics are bias, centered root-mean-squared-error (cRMSE), R-squared, also called correlation coefficient (CC), and Earth-mover's distance (EMD).

In order to decompose model error into bias and random error, I first calculated bias as:

$$Bias = \overline{p} - \overline{o}$$

$\overline{p}$ = mean of model

$\overline{o}$ = mean of the observations from lidars

cRMSE, which is part of the decomposed RMSE, where c refers to “centered” or “unbiased”. If models offered perfect predictions of wind, cRMSE would be 0. This is calculated as:

$$cRMSE = \sqrt{\left[\frac{1}{N}\sum_{n=1}^{N}\left[(p_n - \overline{p}) - (o_n - \overline{o})\right]^2\right]}$$

N = number of data points in the series

p_n = time series values of modeled wind

o_n = time series values of observed wind

The third metric of comparison between the models and observations is a simple R-squared correlation coefficient (CC). R-squared is used for showing how much statistical variance exists between two datasets. Identical data, or a perfect wind speed prediction by models when compared to observations, would have a CC of 1. The correlation coefficient is calculated as:

$$r^2 = \left(\frac{\frac{1}{N}\sum_{N=1}^{N}(p_n - \overline{p}) - (o_n - \overline{o})}{\sigma_p \sigma_o}\right)^2$$

σ_p = standard deviation of modeled data

σ_o = standard deviation of observed data

Finally, the EMD, also known as the Wasserstein distance, measures the difference between the two distributions as an area between two cumulative distribution functions. EMD captures a wider scope by integrating over the entire dataset. A perfect distribution between two wind speed datasets would have an EMD of 0.

$$W_p = \left(\int_0^1 \left|F_o^{-1}(z) - F_p^{-1}(z)\right|^p dz\right)^{1/p}$$

$F_o^{-1}(z)$= inverse probability distribution function of observed wind

$F_p^{-1}(z)$= inverse probability distribution function of modeled wind

3 Results

The results shown here give a comprehensive overview of the performance of each NWP

product versus observations at each site. The sections below analyze the overall mean wind speed profiles and discuss the four chosen error metrics, mean wind speed profiles by season, error metrics by season, overall diurnal cycles, and diurnal cycles by season. Figures 3-11 are referenced in this results section and appear below in Appendix A.

3.1 Overall annual mean wind speed prediction

The annual average shows that WTK, ERA-5, and MERRA-2 all under-predict observed wind speeds at all three sites. Figure 3 shows the overall mean wind speed profiles for each wind speed dataset at each site, which shows that WTK is the closest to accurately predicting wind speeds, followed by MERRA-2, then ERA-5. However, further analysis is required to truly understand the differences between these methods.

3.2 Overall annual wind speed error

Isolating the bias metric confirms that all model products demonstrated a negative bias at the three sites, as shown in Figure 4 a-c. WTK showed the smallest magnitude mean bias at all sites. At GOR, WTK showed an overall annual bias of ~ -1 m/s, MERRA-2 showed ~ -1.5 m/s, and ERA-5 had a large negative bias of ~ -4 m/s. At WAS, WTK did not have a significant bias, and showed only a minute positive bias at 60 meters AGL up to a slight (~ -0.5 m/s) negative bias at 140 meters AGL. MERRA-2 and ERA-5 also had smaller values of bias at WAS, showing ~ -1 m/s and ~-3.5 m/s, respectively. VAN demonstrated similar bias in the model products as GOR: ~ -1 m/s for WTK, ~ -2 m/s for MERRA-2, and ~ -3.5 m/s for ERA-5. VAN exhibited a decreased bias at higher levels, contrasting GOR and WAS.

Expanding bias to cRMSE (Figure 4 d-f) shows that at each site had a different product which performed best. At GOR, MERRA-2 performs best, followed by ERA-5, then WTK. At WAS, ERA-5 performs the best, followed by WTK, then MERRA-2. WTK is the best performer at VAN, followed by ERA-5 then MERRA-2. A possible reason for a different product having a cRMSE advantage at each site is the model grid point's proximity to the observational site.

No site had a strong correlation coefficient, but WTK has the best of the three models at every site. The third metric, R-squared, (Figure 4 g-i) at GOR was nearly identical for each method (~0.5). The correlation at WAS and VAN is strongest for WTK and ERA-5 is considerably close (within 0.1). The correlation coefficients are relatively low for every site (no better than 0.7), highlighting the challenges that exist with any model product when complex terrain is

introduced.

As shown in Figure 4 j-l, WTK had the best EMD for all sites (<1.5 m), which was subsequently followed by MERRA-2. ERA-5 has quite a high EMD for all three sites (3-4 m). The mean wind profiles and error metrics for the annual data suggest that the WTK-WRF produces more accurate predictions than ERA-5 or MERRA-2. This signal can be supported by looking at the seasonal and diurnal variability of these methods.

3.3 Seasonal mean wind speed prediction

The seasonal mean wind profiles for each site generally support the above findings from the overall profiles, but seasonal variability introduces complexity. There are some specific seasonal flows in the Columbia River Gorge region which certainly can influence model performance. Gap flows, or strong gorge winds, occur easterly in the wintertime and westerly in the summertime, promoted by temperature shifts, rapid amplification of a 500-mb ridge, and an offshore pressure gradient in the Pacific Northwest (Sharp & Mass, 2004). Moreover, the hills and valleys in the gorge create intense westerly flow in the cold and transition seasons that can last for hundreds of kilometers (Wilczak et al., 2019).

Each product performed differently compared to observations at the sites, with no clear pattern in the seasonal separation. At GOR, despite slight seasonal variability in observed wind speeds, all models under-predict in each season. MERRA-2 outperformed WTK in winter and fall, and the two had a nearly identical result in spring. WTK was the best performer at GOR during the summer, with MERRA-2 underestimating. ERA-5 followed with a large underestimation for every season.

WTK is the strongest seasonal performer at WAS. Winter at WAS (Figure 5 b) shows a spot-on WTK prediction, followed with a slight overestimation by MERRA-2. WTK again performs best at WAS in spring and summer with a slight underestimation in spring months and slight overestimation in summer months. MERRA-2 is second-best, showing better results in spring than summer (Figure 5 e, h). In the fall, MERRA-2 is the best performer at WAS, closely edging out WTK by being more precise at greater heights (Figure 5 k). ERA-5 once again underestimates speeds for all seasons at WAS.

Finally, at VAN, all methods under-predict in the winter, but MERRA-2 is the top performer.

Spring and summer performances of all three datasets are nearly identical at VAN, with a near-perfect WTK estimation, then MERRA-2 and ERA-5. MERRA-2 and WTK are relatively equivalent in the fall, but ERA-5 underestimates speeds in all seasons (Figure 5 c, f, i, l).

3.4 Seasonal wind speed error

ERA-5 had the largest wintertime bias at all three sites of around ~ -2 to -4.5 m/s. MERRA-2 and WTK differed a bit between the three sites during winter, though. At GOR, MERRA-2 and WTK performed similarly with an average bias of ~ 1 m/s, but WTK had a higher bias at higher heights, while MERRA-2 had a higher bias at lower heights. WAS also featured similar wintertime performance between MERRA-2 and WTK, with biases of ~ 0.5 m/s and -0.5 m/s, respectively. At VAN, MERRA-2 had the smallest wintertime bias of ~ -1 m/s, followed by WTK with ~ -1.5 m/s (Figure 6 a-c). The springtime bias again revealed ERA-5 with a stark under-prediction of ~ -3 to -4 m/s at each site. WTK was the best performer in spring, with a bias of ~ -1 m/s at GOR, an accurate prediction at WAS, and a bias of ~ -0.5 m/s at VAN. MERRA-2 was within 0.5 m/s of WTK at GOR, but was over 1 m/s behind at WAS and VAN. Summer showed ERA-5's largest negative bias of any sites of ~ -4 to -5 m/s (Figure 6 d-f). Summer is also where WTK performed best, never showing a bias of more than 1 m/s in magnitude. WTK predicted excellently at VAN in summer. MERRA-2 showed a consistent ~-2 m/s bias in summertime months at all three sites (Figure 6 g-i). Finally, during the fall, like all other seasons, ERA-5 has a bias of ~ -3 to -4 m/s of bias across all sites. MERRA-2 and WTK each have biases of ~ -1 m/s in the fall. WTK has a slightly better performance than MERRA-2 at lower levels at VAN (Figure 6 j-l).

cRMSE is generally lower during summer months compared to other seasons. GOR has very similar cRMSE performances for WTK, MERRA-2, and ERA-5 during every season, all within 1 m/s of each other. At WAS, ERA-5 is the consistent best performer, which comes as a surprise after the product's large biases (Figure 6). WTK is second-best in winter and summer, followed by MERRA-2. In spring, WTK and MERRA-2 performed almost identically (Figure 7 d-f). VAN has WTK with the lowest cRMSE for all four seasons (Figure 7 c, f, i, l). ERA-5 and MERRA-2 are similar for winter, spring, and summer with a cRMSE of ~1 m/s greater than that of WTK, but ERA-5 actually is quite similar to WTK in fall.

The correlations between WTK, MERRA-2, and ERA-5 and observations varied greatly with

seasons. At GOR, all methods had about the same R-squared during summer of ~ 0.6 (Figure 8 g). In winter, MERRA-2 had the highest CC, followed by WTK, then ERA-5 (Figure 8 a). In spring at GOR, WTK had the highest CC, then ERA-5, then MERRA-2 (Figure 8 d). In fall at GOR, MERRA-2 again has the strongest CC, this time followed by ERA-5, then WTK (Figure 8 j). At WAS, winter and spring demonstrate WTK as having the strongest correlation, followed by ERA-5 (Figure 8 b, e). Summer and fall at WAS have ERA-5 and WTK with similar correlations (Figure 8 h, k). At VAN, WTK has the best CC for winter, spring, and summer, followed by ERA-5 (Figure 8 c, f, i). In summer at VAN, ERA-5 has the strongest CC, the only site where the product has the strongest correlation to observations (Figure 8 l). MERRA-2, despite having little bias and closely matching WTK in that metric, has the weakest R-squared for all seasons at WAS and VAN. As a whole, the correlations change seasonally, proving lower in winter and increasing in summer at GOR and WAS.

EMD also changed a fair bit when isolating seasonal periods for the three sites. Across the board, ERA-5 had the highest EMD, typically ~ 3 m or more. WTK consistently had the smallest EMD, barely exceeding 1 m. MERRA-2 was the most variable throughout the seasons. MERRA-2 did well in winter and fall, even exceeding WTK for higher levels at GOR (Figure 9 a, j). EMD for MERRA-2 increased during spring and summer months, though, deviating starkly from WTK.

3.5 Diurnal mean wind speed prediction

The relative strengths of these reanalysis products can also be considered by how accurately they predict diurnal cycles. A general weakness at predicting these day-night shifts is expected, given that the coarse spatial and temporal resolution of NWP products could miss expected phenomena in the boundary layer at night, like quiet near-surface winds and upper-level acceleration with low-level jet development and turbulence collapse. It is also challenging for NWP products to consider other meteorological details like surface variability and soil moisture.

The overall annual average diurnal cycles are shown in Figure 8. At all three sites, there is a clear absence of any diurnal wind speed cycle in ERA-5 and the same negative bias explained above is apparent in these plots. MERRA-2 shows a very slight diurnal variation, but drastically underestimates the cycle. On the other hand, WTK tends to over-exaggerate the diurnal cycle with higher day-to-night deviations than those coming from surface observations.

3.6 Seasonal diurnal wind speed

It is beneficial to break down the diurnal cycle by season, like what was done with the mean wind plots and performance metrics in sections 3.3 and 3.4. Neither ERA-5 nor WTK show much diurnal cycle in the wintertime.. WTK appears to match the shape of the cycle, but is timed poorly, as shown in Figure 11 a, b. During springtime, WTK exaggerates the diurnal cycle at all three sites. MERRA-2 and ERA-5 show no clear cycle at GOR or WAS (Figure 11 d, e). All methods surprisingly show a cycle at VAN, with WTK being the most accurate. MERRA-2 seems to match the profile best though, with a low-bias (Figure 11 f). In summertime, MERRA-2 and ERA-5 both match the diurnal profile excellently at GOR and WAS, but have low biases. WTK again has an over-emphasized cycle in summer for the first 2 sites. WTK predicts relatively well at VAN, where ERA-5 and MERRA-2 actually understate the day-night wind speed differences. In fall, WTK has the best diurnal cycle shape for replicating observations at GOR and WAS, but MERRA-2 appears to have the closest speed prediction. Consistent with the other sites and seasons, WTK has a cycle that is more prominent than real speeds. At VAN, MERRA-2 is close to WTK in speeds, but more accurate in profile.

4 Conclusion

Wind energy is taking off as a reliable, sustainable way that people will be able to obtain electricity. Wind is an extremely variable phenomena, though, and as wind energy starts to enjoy an increased share of the worldwide electricity market, the importance of accurately forecasting and assessing wind resources at potential sites will escalate (Marquis et al., 2011). When on-site measurements of wind speeds are not possible to obtain, NWP and reanalysis products can be used in their place. Performing a model validation using three sites from the WFIP2 project, located in mountainous terrain, helps to understand how WTK-WRF, ERA-5, and MERRA-2 work in such conditions.

As a whole, none of these products excelled in complex terrain. When looking specifically at correlation coefficient, every site and every method had extremely low correlation to actual lidar observations. This highlights that there are some significant challenges with using any of these products in complex terrain. The mean wind profiles and error metrics for the annual data give a primary indication that WTK-WRF runs produce more accurate predictions than ERA-5 or MERRA-2, and this is supported by separating out data into a seasonal comparison.

The seasonal mean wind plots again indicate that WTK has the most accurate mean wind

speed prediction. MERRA-2 is a close competitor, though, especially in fall and winter. ERA-5 vastly underestimated speeds at all three locations for all four seasons. WTK's greatest strength was a very low-magnitude bias in summer months. WTK was the overall best product based on the chosen metrics (bias, cRMSE, CC, EMD).

MERRA-2 and ERA-5 would need vast improvements to be reliable for proper wind resource assessment. The performance metrics further highlighted ERA-5's poor forecast accuracy, but surprisingly, the profiles of ERA-5 matched observations very well, indicating that it could be a stronger product to use if a bias-correction was applied to the reanalysis data. Use of bias-corrected reanalysis is already considered for some projects (Staffell & Pfenninger, 2016). One potential reason for ERA-5's overall slow-bias could be that the ERA-5 centroids are quite far from the locations of the lidars (Table 4). The accuracy advantages of the WTK are clear here, but if a fairly accurate prediction could be achieved with and cost-effective reanalysis, developers are sure to consider these simpler models.

This research about WTK, MERRA-2, and ERA-5 should be expanded by applying similar model validations at additional complex terrain sites that have available observational data. There is plenty of research that recognizes that the physical complex terrain space, such as that in the Swiss Alps, is advantageous for wind energy development (Clifton et al., 2014). The terrain is different there than in the Columbia River Gorge, and using these models in a variety of terrain types will further demonstrate their strengths and necessary improvements. As understanding of and wind behavior in complex terrain increases, so too will the importance of this research. Mountain flows, mountain wakes, and recirculation zones are just a few of the meteorological phenomena that occur in complex terrain (Lange et al., 2017; Menke et al., 2019). Continued experimentation will allow for more scientific knowledge about the weather and atmosphere in complex terrain and thus, the best way to utilize NWP products for wind energy development.

Improvements to the model products themselves could also be relevant for future research. Using an updated Planetary Boundary Layer (PBL) scheme, such as the 3DPBL, in models could significantly increase their performance (Arthur et al., 2022). This 3DPBL can be also be coupled with a wind farm parameterization for improved forecasting (Rybchuk, Juliano, et al., 2021). Combining wind resource assessment methods is proving to be a strong way to formulate an accurate prediction. Some European laboratories are utilizing satellite-borne lidars

to measure wind resources in the boundary layer, a method that could be especially strong when viewed alongside reanalysis products (Lux et al., 2022; Witschas et al., 2020).

Appendix A

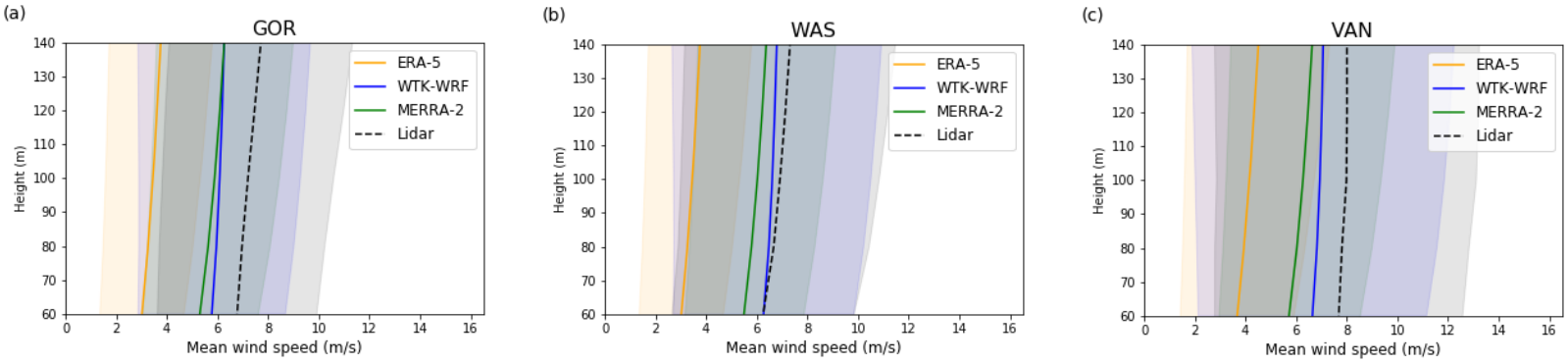

Figure 3: Plots of overall annual wind speed profiles for GOR, WAS, and VAN Shaded areas show 1 standard deviation of the mean of each dataset

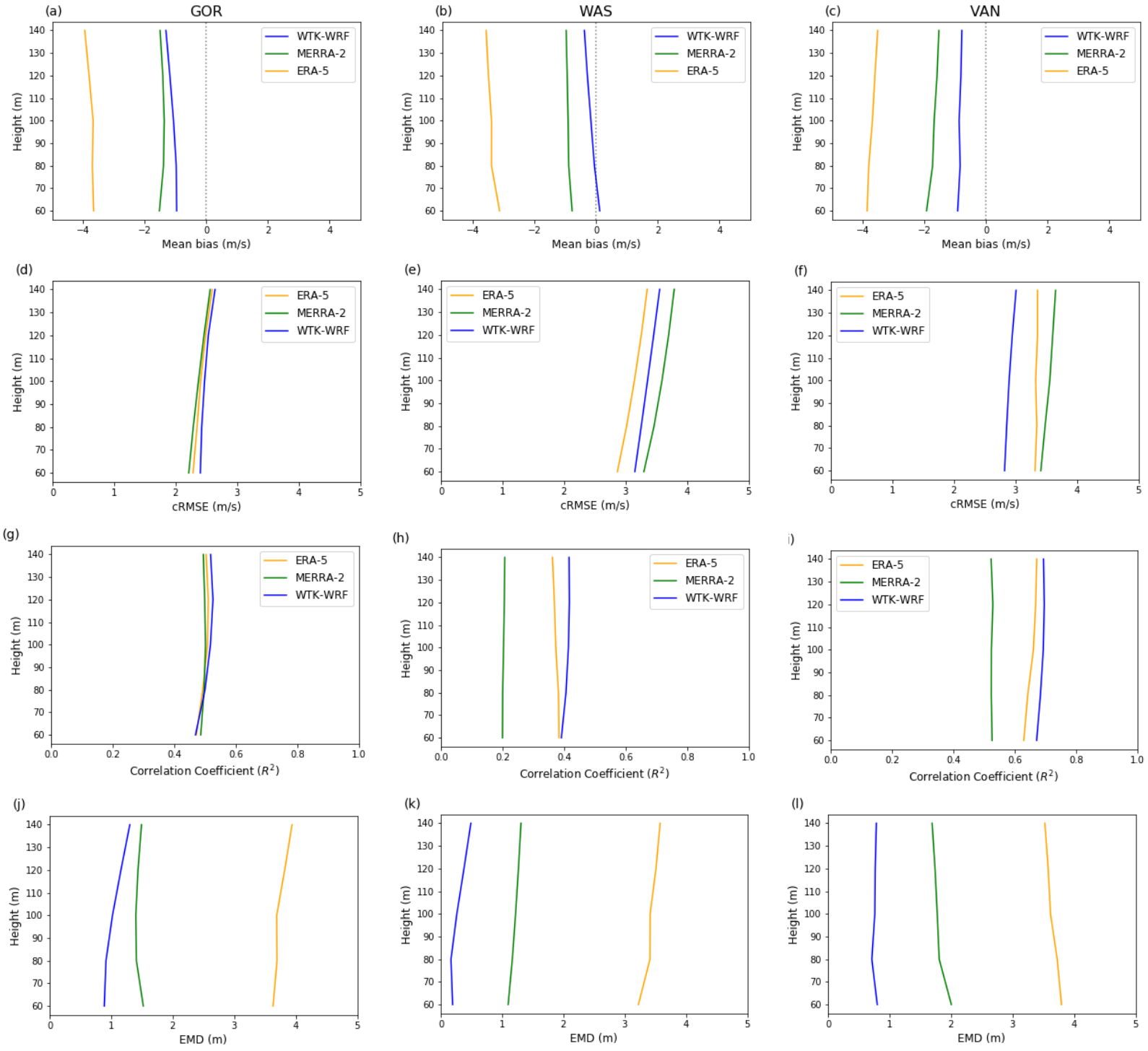

Figure 4: Plots of overall annual mean bias (m/s), cRMSE, CC, and EMD for GOR, WAS, and VAN.

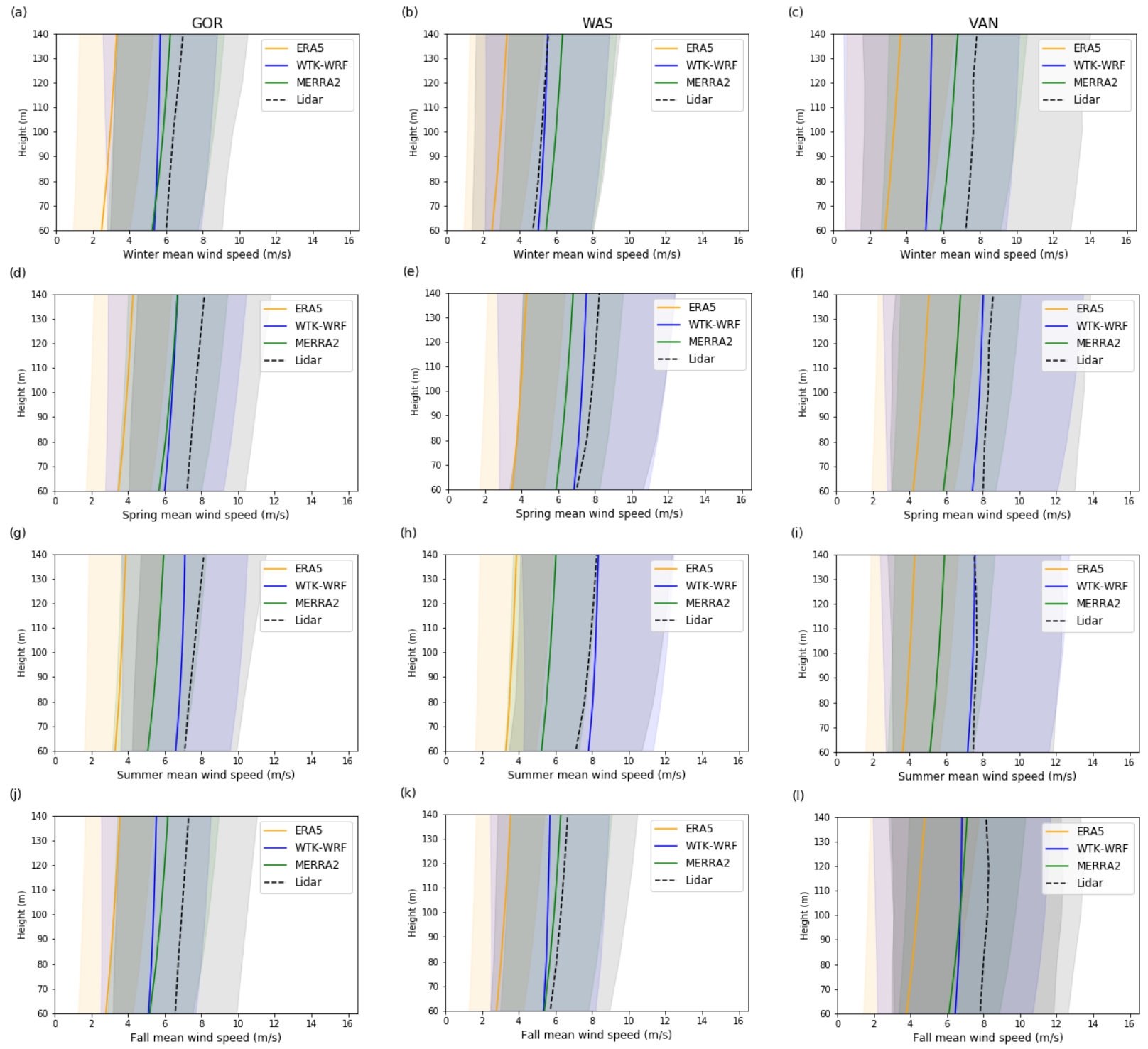

Figure 5: Plots of wind speed profiles for each season at GOR, WAS, and VAN.

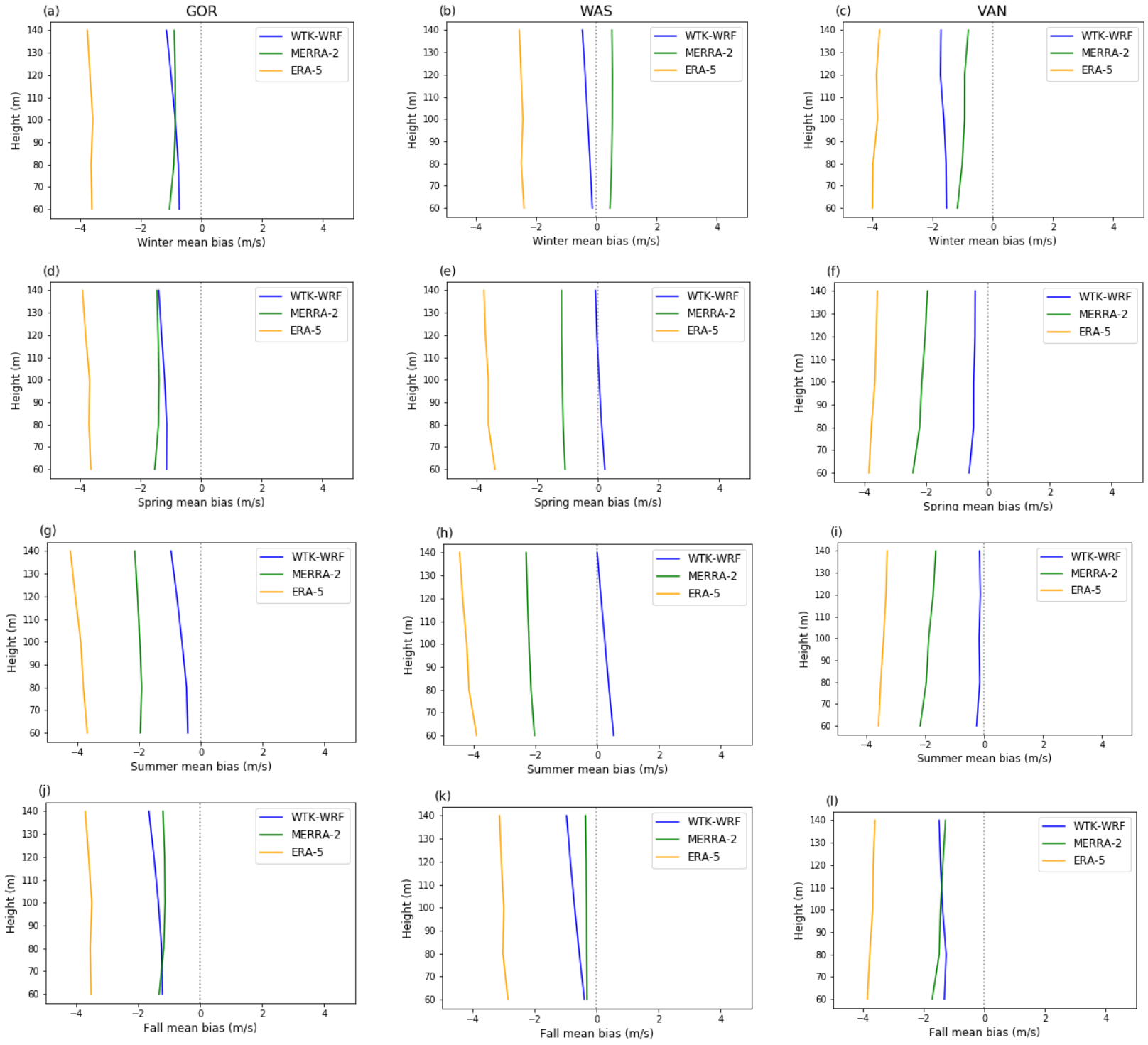

Figure 6: Plots of mean bias (m/s) at GOR, WAS, and VAN by season

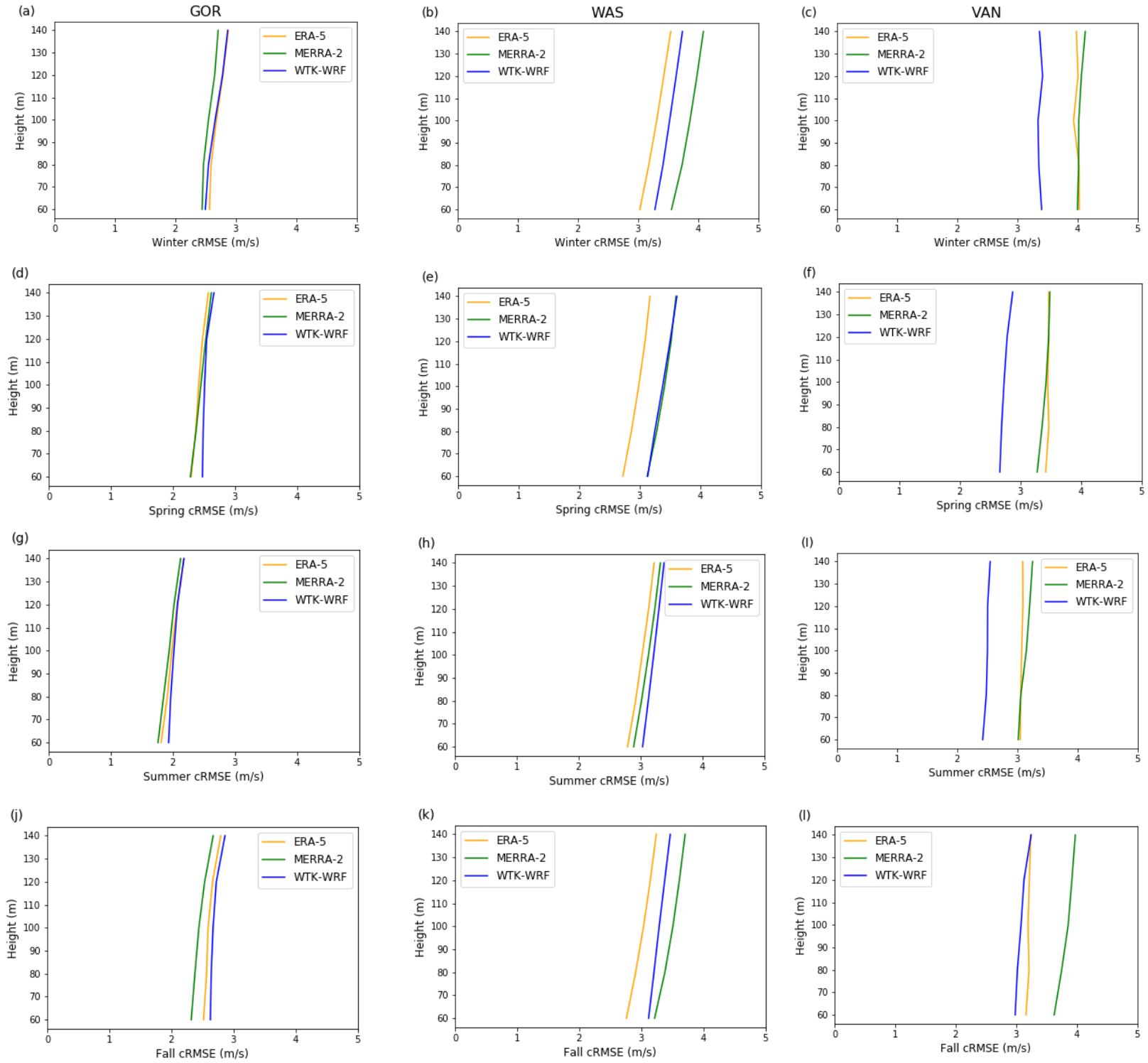

Figure 7: Plots of cRMSE (m/s) at GOR, WAS, and VAN by season

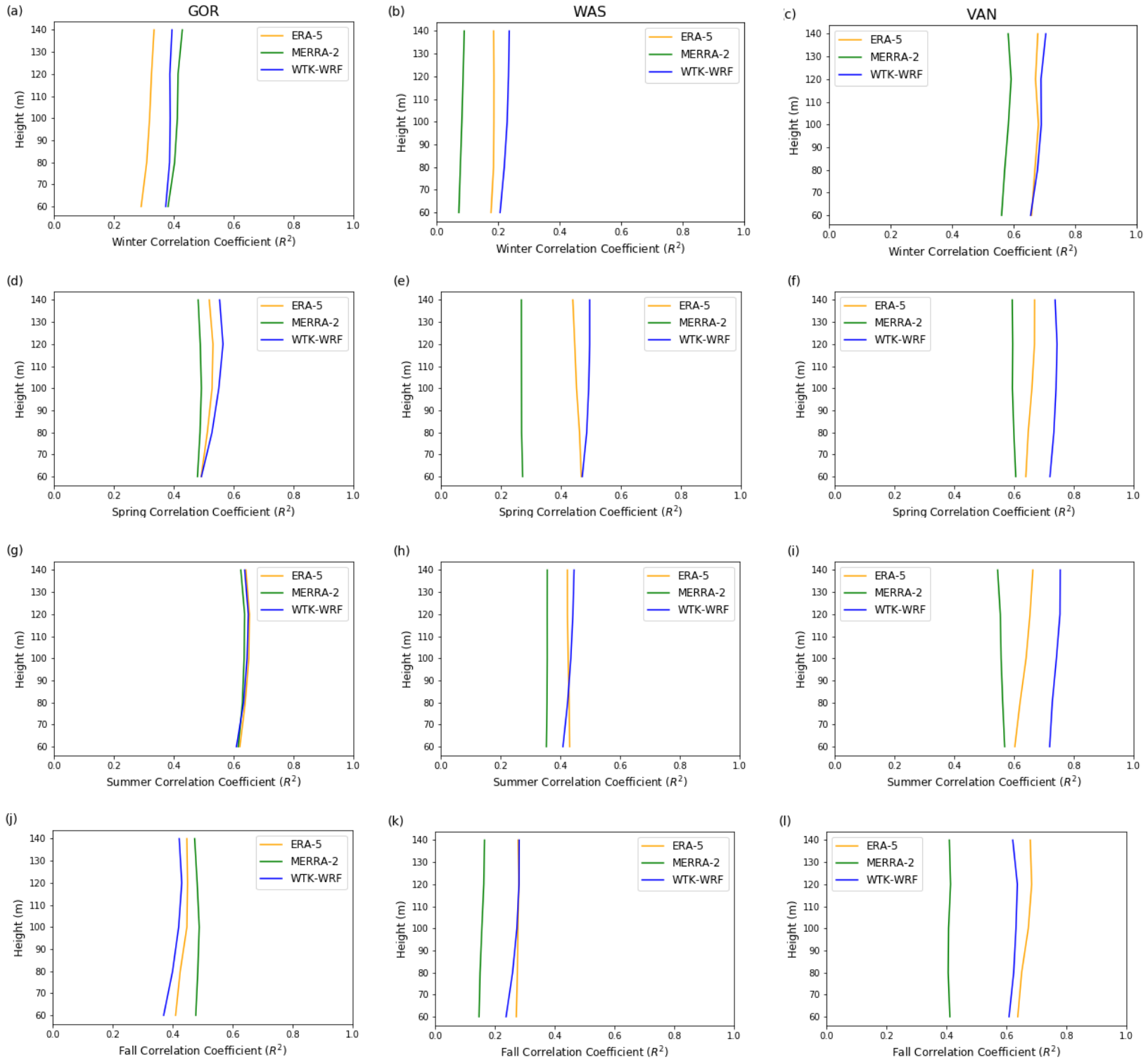

Figure 8: Plots of Correlation Coefficient at GOR, WAS, and VAN by season

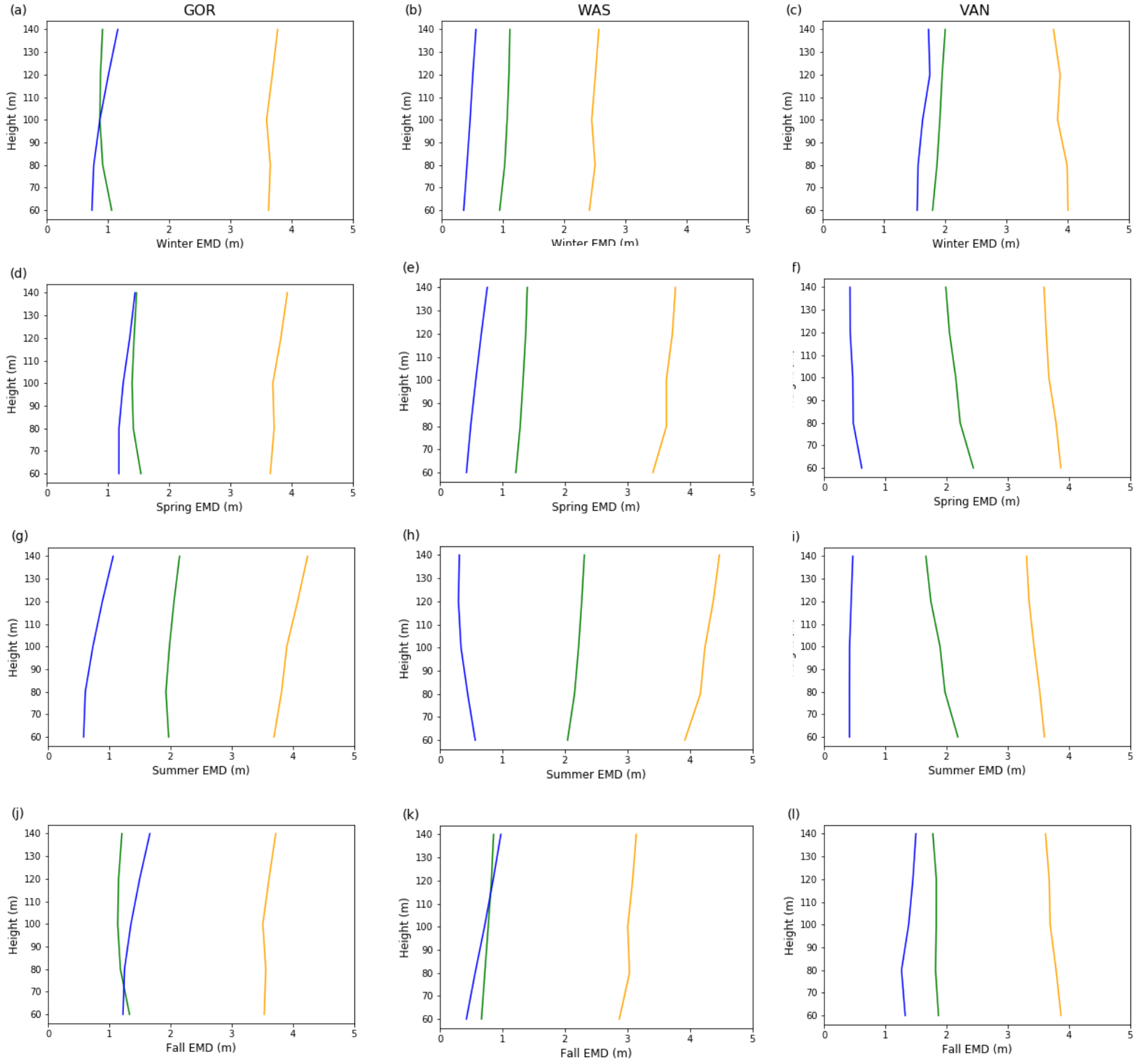

Figure 9: Plots of EMD (m) at GOR, WAS, and VAN by season

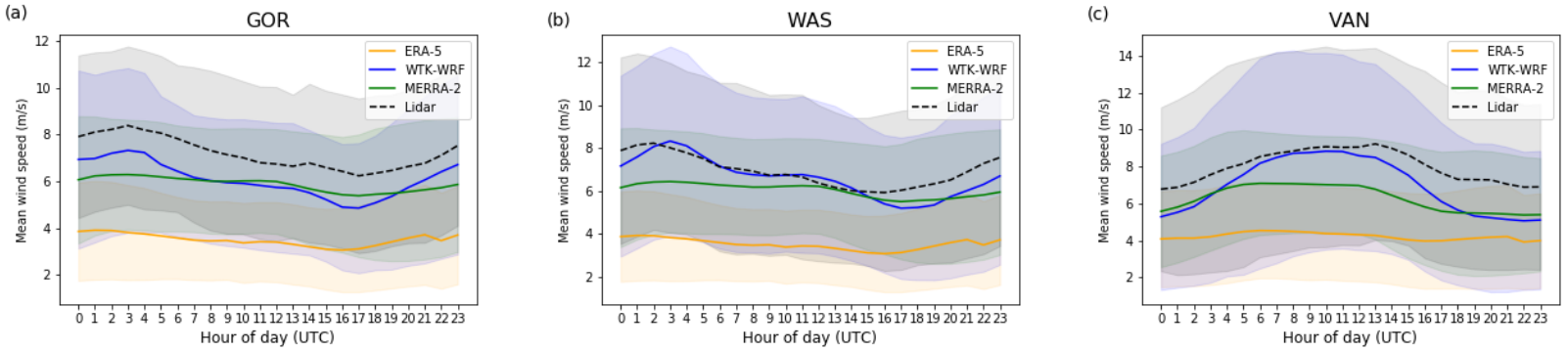

Figure 10: Plots of overall annual diurnal cycle profiles for GOR, WAS, and VAN. Shaded areas show 1 standard deviation of the mean of each dataset.

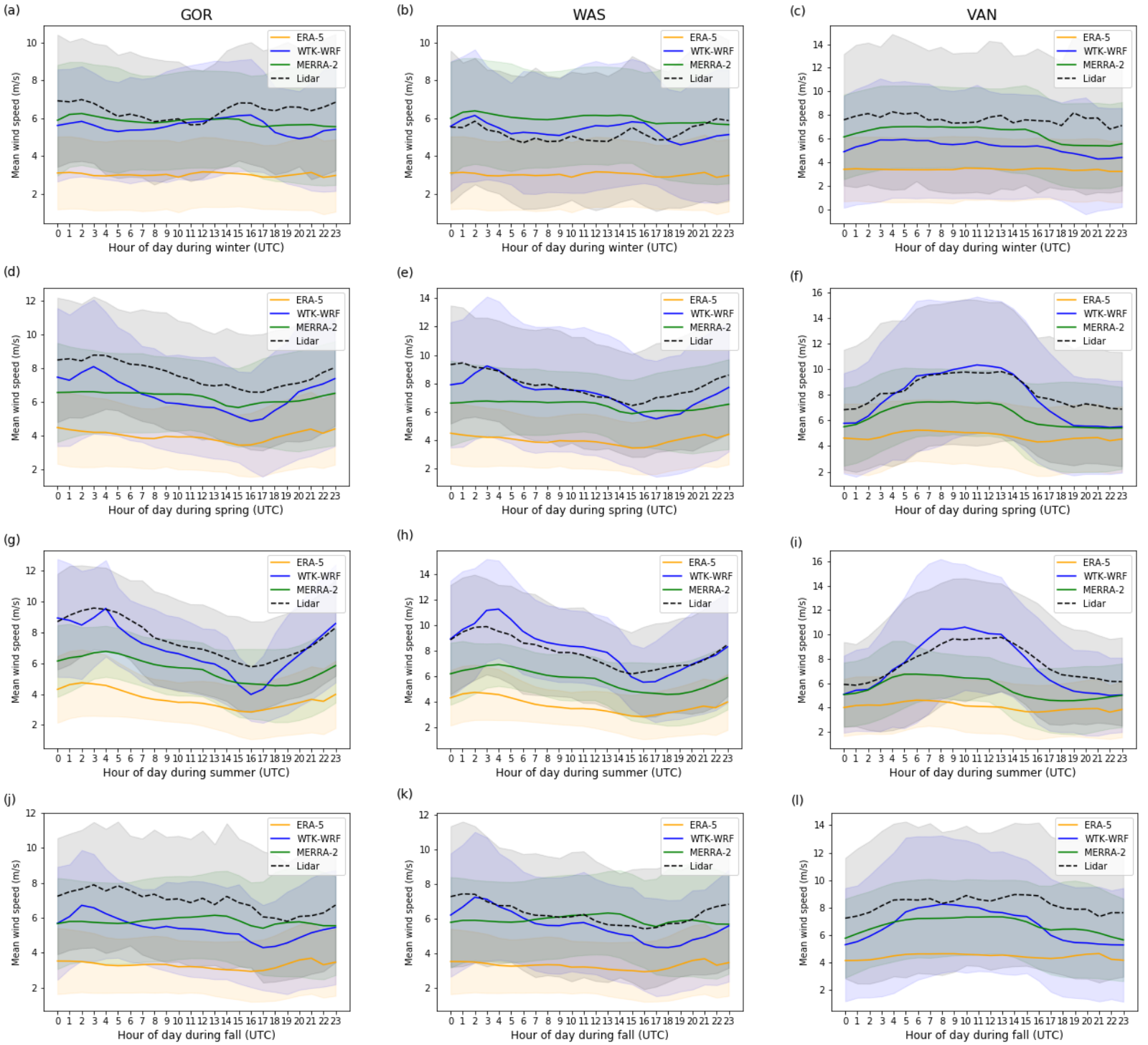

Figure 11: Plots of diurnal cycle profiles for each season at GOR, WAS, and VAN.

References

Ahmad, S., Abdullah, M., Kanwal, A., Tahir, Z. ul R., Saeed, U. B., Manzoor, F., Atif, M., & Abbas, S. (2022). Offshore wind resource assessment using reanalysis data. *Wind Engineering*, 0309524X211069384. https://doi.org/10.1177/0309524X211069384

Aitken, M. L., Rhodes, M. E., & Lundquist, J. K. (2012). Performance of a Wind-Profiling Lidar in the Region of Wind Turbine Rotor Disks. *Journal of Atmospheric and Oceanic Technology*, *29*(3), 347–355. https://doi.org/10.1175/JTECH-D-11-00033.1

Al-Yahyai, S., Charabi, Y., & Gastli, A. (2010). Review of the use of Numerical Weather Prediction (NWP) Models for wind energy assessment. *Renewable and Sustainable Energy Reviews*, *14*(9), 3192–3198. https://doi.org/10.1016/j.rser.2010.07.001

Arthur, R. S., Juliano, T. W., Adler, B., Krishnamurthy, R., Lundquist, J. K., Kosović, B., & Jiménez, P. A. (2022). Improved representation of horizontal variability and turbulence in mesoscale simulations of an extended cold-air pool event. *Journal of Applied Meteorology and Climatology*, *1*(aop). https://doi.org/10.1175/JAMC-D-21-0138.1

Asif, M., & Muneer, T. (2007). Energy supply, its demand and security issues for developed and emerging economies. *Renewable and Sustainable Energy Reviews*, *11*(7), 1388–1413. https://doi.org/10.1016/j.rser.2005.12.004

Ayik, A., Ijumba, N., Kabiri, C., & Goffin, P. (2021). Preliminary wind resource assessment in South Sudan using reanalysis data and statistical methods. *Renewable and Sustainable Energy Reviews*, *138*, 110621. https://doi.org/10.1016/j.rser.2020.110621

Bodini, N., Lundquist, J. K., Krishnamurthy, R., Pekour, M., Berg, L. K., & Choukulkar, A. (2019). Spatial and temporal variability of turbulence dissipation rate in complex terrain. *Atmospheric Chemistry and Physics*, *19*(7), 4367–4382. https://doi.org/10.5194/acp-19-4367-2019

Bosilovich, M., Lucchesi, R., & Suarez, M. (2016). MERRA-2: File Specification. *NASA Global*

Modeling and Assimilation Office, *9*(1.1), 0–73.

Brower, M. (2012). *Wind Resource Assessment: A Practical Guide to Developing a Wind Project*. John Wiley & Sons.

Castorrini, A., Gentile, S., Geraldi, E., & Bonfiglioli, A. (2021). Increasing spatial resolution of wind resource prediction using NWP and RANS simulation. *Journal of Wind Engineering and Industrial Aerodynamics*, *210*, 104499. https://doi.org/10.1016/j.jweia.2020.104499

Chancham, C., Waewsak, J., & Gagnon, Y. (2017). Offshore wind resource assessment and wind power plant optimization in the Gulf of Thailand. *Energy*, *139*, 706–731. https://doi.org/10.1016/j.energy.2017.08.026

Clifton, A., Daniels, M. H., & Lehning, M. (2014). Effect of winds in a mountain pass on turbine performance. *Wind Energy*, 1543–1562. https://doi.org/10.1002/we.1650

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., … Worley, S. J. (2011). The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 1–28. https://doi.org/10.1002/qj.776

Dayal, K. K., Bellon, G., Cater, J. E., Kingan, M. J., & Sharma, R. N. (2021). High-resolution mesoscale wind-resource assessment of Fiji using the Weather Research and Forecasting (WRF) model. *Energy*, *232*, 121047. https://doi.org/10.1016/j.energy.2021.121047

Dörenkämper, M., Olsen, B. T., Witha, B., Hahmann, A. N., Davis, N. N., Barcons, J., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Sastre-Marugán, M., Sīle, T., Trei, W., Žagar, M., Badger, J., Gottschall, J., Sanz Rodrigo, J., & Mann, J. (2020). The Making of the New European Wind Atlas – Part 2: Production and evaluation. *Geoscientific Model Development*, *13*(10), 5079–5102. https://doi.org/10.5194/gmd-13-5079-2020

Draxl, C., Clifton, A., Hodge, B.-M., & McCaa, J. (2015). The Wind Integration National Dataset (WIND) Toolkit. *Applied Energy*, *151*, 355–366. https://doi.org/10.1016/j.apenergy.2015.03.121

Fan, W., Liu, Y., Chappell, A., Dong, L., Xu, R., Ekström, M., Fu, T.-M., & Zeng, Z. (2021). Evaluation of Global Reanalysis Land Surface Wind Speed Trends to Support Wind Energy Development Using In Situ Observations. *Journal of Applied Meteorology and Climatology*, *60*(1), 33–50. https://doi.org/10.1175/JAMC-D-20-0037.1

Fernández-González, S., Martín, M. L., García-Ortega, E., Merino, A., Lorenzana, J., Sánchez, J. L., Valero, F., & Rodrigo, J. S. (2018). Sensitivity Analysis of the WRF Model: Wind-Resource Assessment for Complex Terrain. *Journal of Applied Meteorology and Climatology*, *57*(3), 733–753. https://doi.org/10.1175/JAMC-D-17-0121.1

Giannaros, T. M., Melas, D., & Ziomas, I. (2017). Performance evaluation of the Weather Research and Forecasting (WRF) model for assessing wind resource in Greece. *Renewable Energy*, *102*, 190–198. https://doi.org/10.1016/j.renene.2016.10.033

Goit, J. P., Shimada, S., & Kogaki, T. (2019). Can LiDARs Replace Meteorological Masts in Wind Energy? *Energies*, *12*(19), 3680. https://doi.org/10.3390/en12193680

Gruber, K., Regner, P., Wehrle, S., Zeyringer, M., & Schmidt, J. (2022). Towards global validation of wind power simulations: A multi-country assessment of wind power simulation from MERRA-2 and ERA-5 reanalyses bias-corrected with the global wind atlas. *Energy*, *238*, 121520. https://doi.org/10.1016/j.energy.2021.121520

Gualtieri, G. (2021). Reliability of ERA5 Reanalysis Data for Wind Resource Assessment: A Comparison against Tall Towers. *Energies*, *14*(14), 4169. https://doi.org/10.3390/en14144169

Hahmann, A. N., Sīle, T., Witha, B., Davis, N. N., Dörenkämper, M., Ezber, Y., García-Bustamante, E., González-Rouco, J. F., Navarro, J., Olsen, B. T., & Söderberg, S. (2020). The making of the New European Wind Atlas – Part 1: Model sensitivity.

Geoscientific Model Development, *13*(10), 5053–5078. https://doi.org/10.5194/gmd-13-5053-2020

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research: Atmospheres*, *113*(D13). https://doi.org/10.1029/2008JD009944

Indasi, V. S., Lynch, M., McGann, B., & Sutton, J. (2017). WIND RESOURCE ASSESSMENT USING WRF MODEL IN COMPLEX TERRAIN. *International Journal of Latest Engineering Research and Applications*, *02*, 2455–7137.

Jourdier, B. (2020). Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France. *Advances in Science and Research*, *17*, 63–77. https://doi.org/10.5194/asr-17-63-2020

Lange, J., Mann, J., Berg, J., Parvu, D., Kilpatrick, R., Costache, A., Chowdhury, J., Siddiqui, K., & Hangan, H. (2017). For wind turbines in complex terrain, the devil is in the detail. *Environmental Research Letters*, *12*(9), 094020. https://doi.org/10.1088/1748-9326/aa81db

Lee, J. A., Doubrawa, P., Xue, L., Newman, A. J., Draxl, C., & Scott, G. (2019). Wind Resource Assessment for Alaska's Offshore Regions: Validation of a 14-Year High-Resolution WRF Data Set. *Energies*, *12*(14), 2780. https://doi.org/10.3390/en12142780

Lee, J. C. Y., & Fields, M. J. (2021). An overview of wind-energy-production prediction bias, losses, and uncertainties. *Wind Energy Science*, *6*(2), 311–365.

https://doi.org/10.5194/wes-6-311-2021

Lux, O., Lemmerz, C., Weiler, F., Marksteiner, U., Witschas, B., Rahm, S., Geiß, A., Schäfler, A., & Reitebuch, O. (2022). Retrieval improvements for the ALADIN Airborne Demonstrator in support of the Aeolus wind product validation. *Atmospheric Measurement Techniques*, *15*(5), 1303–1331. https://doi.org/10.5194/amt-15-1303-2022

Marquis, M., Wilczak, J., Ahlstrom, M., Sharp, J., Stern, A., Smith, J. C., & Calvert, S. (2011). Forecasting the Wind to Reach Significant Penetration Levels of Wind Energy. *Bulletin of the American Meteorological Society*, *92*(9), 1159–1171. https://doi.org/10.1175/2011BAMS3033.1

Menke, R., Vasiljević, N., Mann, J., & Lundquist, J. K. (2019). Characterization of flow recirculation zones at the Perdigão site using multi-lidar measurements. *Atmospheric Chemistry and Physics*, *19*(4), 2713–2723. https://doi.org/10.5194/acp-19-2713-2019

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. *Monthly Weather Review*, *137*(3), 991–1007. https://doi.org/10.1175/2008MWR2556.1

Nakanishi, M., & Niino, H. (2009). Development of an Improved Turbulence Closure Model for the Atmospheric Boundary Layer. *Journal of the Meteorological Society of Japan. Ser. II*, *87*(5), 895–912. https://doi.org/10.2151/jmsj.87.895

Nor, K. M., Shaaban, M., & Abdul Rahman, H. (2014). Feasibility assessment of wind energy resources in Malaysia based on NWP models. *Renewable Energy*, *62*, 147–154. https://doi.org/10.1016/j.renene.2013.07.001

Optis, M., Bodini, N., Debnath, M., & Doubrawa, P. (2020). *Best Practices for the Validation of U.S. Offshore Wind Resource Models* (NREL/TP-5000-78375). National Renewable Energy Lab. (NREL), Golden, CO (United States). https://doi.org/10.2172/1755697

Pronk, V., Bodini, N., Optis, M., Lundquist, J. K., Moriarty, P., Draxl, C., Purkayastha, A., &

Young, E. (2021). *Can Reanalysis Products Outperform Mesoscale Numerical Weather Prediction Models in Modeling the Wind Resource in Simple Terrain?* [Preprint]. Wind and turbulence. https://doi.org/10.5194/wes-2021-97

Ramon, J., Lledó, L., Torralba, V., Soret, A., & Doblas-Reyes, F. J. (2019). What global reanalysis best represents near-surface winds? *Quarterly Journal of the Royal Meteorological Society*, *145*(724), 3236–3251. https://doi.org/10.1002/qj.3616

Rodrigues, S., Restrepo, C., Kontos, E., Teixeira Pinto, R., & Bauer, P. (2015). Trends of offshore wind projects. *Renewable and Sustainable Energy Reviews*, *49*, 1114–1135. https://doi.org/10.1016/j.rser.2015.04.092

Rybchuk, A., Juliano, T. W., Lundquist, J. K., Rosencrans, D., Bodini, N., & Optis, M. (2021). The Sensitivity of the Fitch Wind Farm Parameterization to a Three-Dimensional Planetary Boundary Layer Scheme. *Wind Energy Science Discussions*, 1–39. https://doi.org/10.5194/wes-2021-127

Rybchuk, A., Optis, M., Lundquist, J. K., Rossol, M., & Musial, W. (2021). A Twenty-Year Analysis of Winds in California for Offshore Wind Energy Production Using WRF v4.1.2. *Geoscientific Model Development Discussions*, 1–41. https://doi.org/10.5194/gmd-2021-50

Salvação, N., & Guedes Soares, C. (2018). Wind resource assessment offshore the Atlantic Iberian coast with the WRF model. *Energy*, *145*, 276–287. https://doi.org/10.1016/j.energy.2017.12.101

Samal, R. K. (2021). Assessment of wind energy potential using reanalysis data: A comparison with mast measurements. *Journal of Cleaner Production*, *313*, 127933. https://doi.org/10.1016/j.jclepro.2021.127933

Schwartz, M., George, R., & Elliott, D. (1999). *The Use of Reanalysis Data for Wind Resource Assessment at the National Renewable Energy Laboratory* (NREL/CP-500-26152). National Renewable Energy Lab. (NREL), Golden, CO (United States).

https://www.osti.gov/biblio/7074

Shafiee, S., & Topal, E. (2009). When will fossil fuel reserves be diminished? *Energy Policy*, *37*(1), 181–189. https://doi.org/10.1016/j.enpol.2008.08.016

Sharp, J., & Mass, C. F. (2004). Columbia Gorge Gap Winds: Their Climatological Influence and Synoptic Evolution. *Weather and Forecasting*, *19*(6), 970–992. https://doi.org/10.1175/826.1

Shaw, W. J., Berg, L. K., Cline, J., Draxl, C., Djalalova, I., Grimit, E. P., Lundquist, J. K., Marquis, M., McCaa, J., Olson, J. B., Sivaraman, C., Sharp, J., & Wilczak, J. M. (2019). The Second Wind Forecast Improvement Project (WFIP2): General Overview. *Bulletin of the American Meteorological Society*, *100*(9), 1687–1699. https://doi.org/10.1175/BAMS-D-18-0036.1

Sheppard, C. (2009). Analysis of the measure-correlate-predict methodology for wind resource assessment. *Undefined*. https://www.semanticscholar.org/paper/Analysis-of-the-measure-correlate-predict-for-wind-Sheppard/2a5a2aa29365ea6f7fb0009e2bd6d07be583aa17

Sheridan, L. M., Krishnamurthy, R., Gorton, A. M., Shaw, W. J., & Newsom, R. K. (2020). Validation of Reanalysis-Based Offshore Wind Resource Characterization Using Lidar Buoy Observations. *Marine Technology Society Journal*, *54*(6), 44–61. https://doi.org/10.4031/MTSJ.54.6.13

Staffell, I., & Pfenninger, S. (2016). Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, *114*, 1224–1239. https://doi.org/10.1016/j.energy.2016.08.068

Wilczak, J. M., Stoelinga, M., Berg, L. K., Sharp, J., Draxl, C., McCaffrey, K., Banta, R. M., Bianco, L., Djalalova, I., Lundquist, J. K., Muradyan, P., Choukulkar, A., Leo, L., Bonin, T., Pichugina, Y., Eckman, R., Long, C. N., Lantz, K., Worsnop, R. P., … White, A. B. (2019). The Second Wind Forecast Improvement Project (WFIP2): Observational Field

Campaign. *Bulletin of the American Meteorological Society*, *100*(9), 1701–1723. https://doi.org/10.1175/BAMS-D-18-0035.1

Wind in Numbers | *GWEC*. (2022). Retrieved February 17, 2022, from http://www.gwec.net/global-figures/wind-in-numbers/

Wiser, R., & Bolinger, M. (2019). *2018 Wind Technologies Market Report*.

Witschas, B., Lemmerz, C., Geiß, A., Lux, O., Marksteiner, U., Rahm, S., Reitebuch, O., & Weiler, F. (2020). First validation of Aeolus wind observations by airborne Doppler wind lidar measurements. *Atmospheric Measurement Techniques*, *13*(5), 2381–2396. https://doi.org/10.5194/amt-13-2381-2020