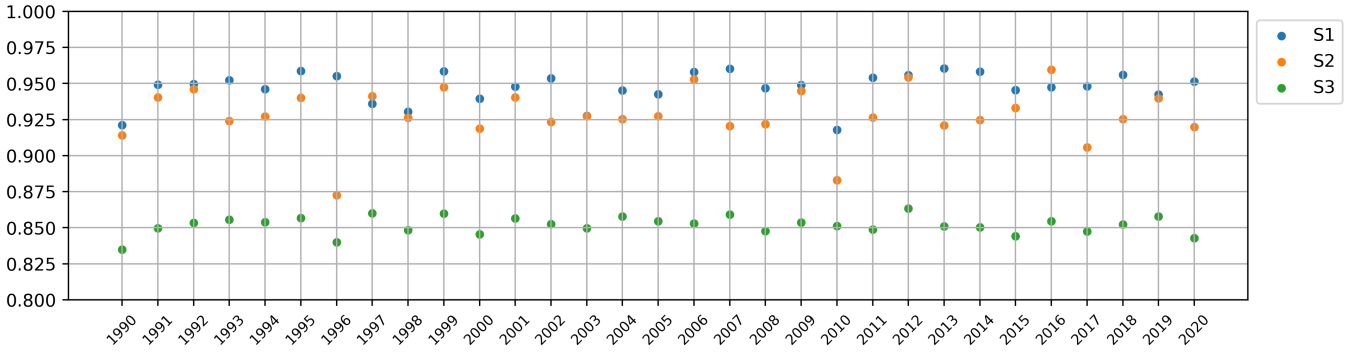

Figure S1: Median S1, S2 and S3 scores over the considered North Sea grid cells for year (1990-2020). Higher values correspond to a larger overlap between the single-year and multi-year histograms.

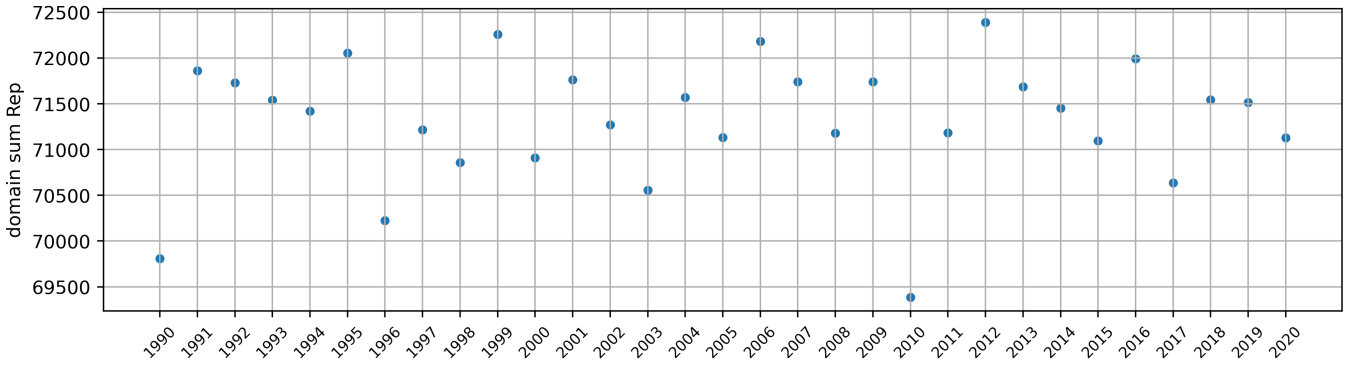

Figure S2: Final representativeness value for each year (1990-2020).

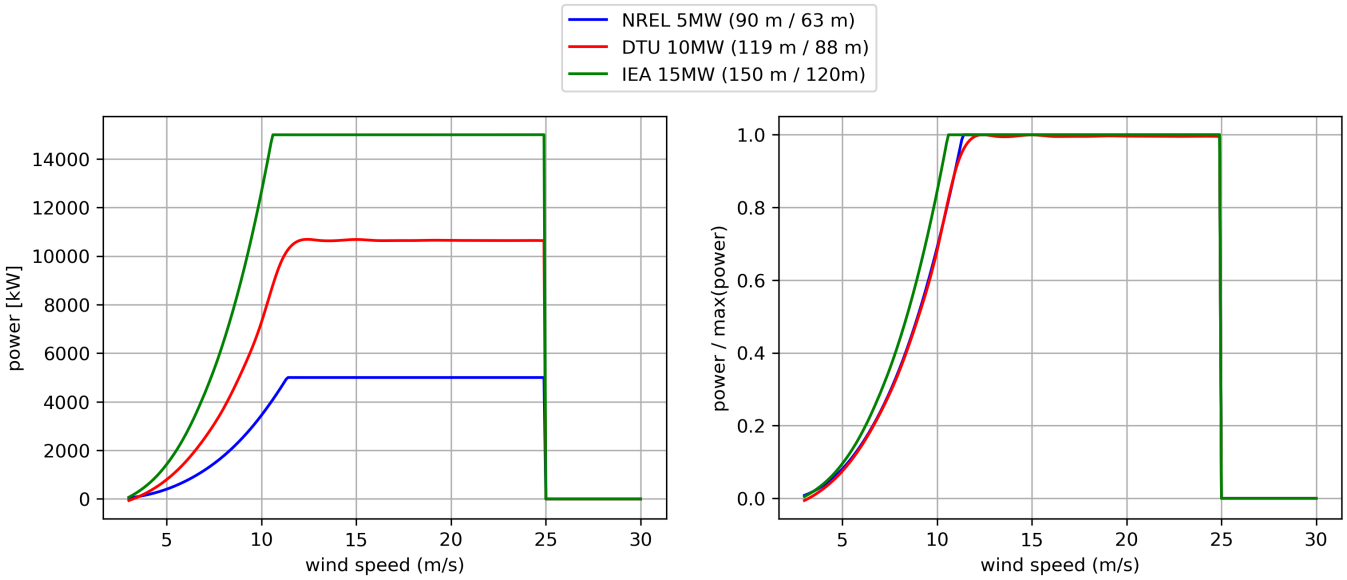

Figure S3: Left: power curves of the three reference turbines employed in this study. Right: same as left, but the power curves are normalized by the maximum power output.

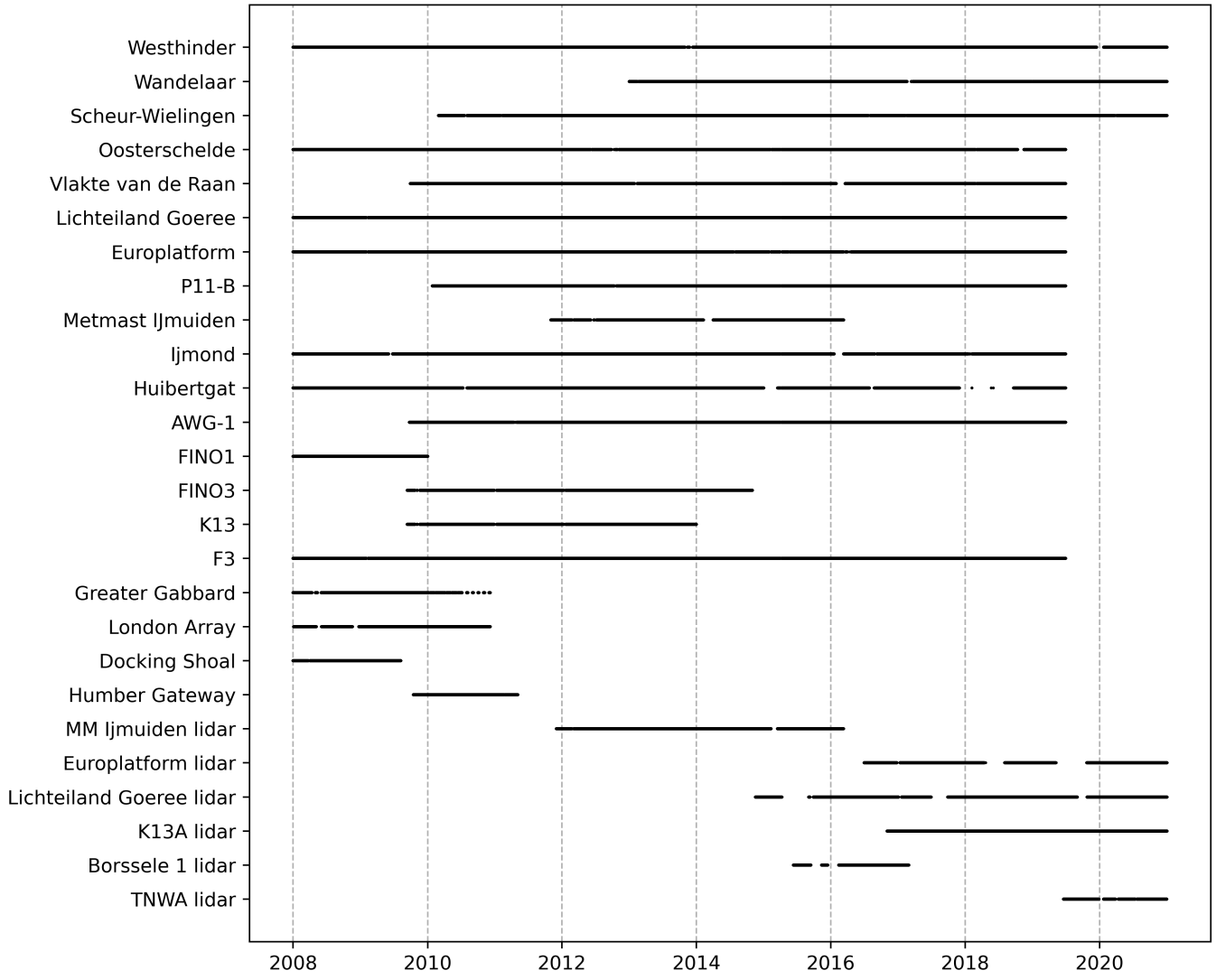

Figure S4: Temporal coverage of the different measurement datasets.

Station	**wind farms**	**filtering**
FINO1	Alpha Ventus	omission after July 2009
FINO3	Dantysk, Sandbank	omission after November 2014
WH	Belgian offshore cluster	removed [27°, 77°]
WA	Belgian offshore cluster	removed [325°, 15°]
SW	Belgian offshore cluster	removed [287°, 341°]
BO	Belgian offshore cluster	removed [150°, 300°]
OS	Belgian offshore cluster	removed [243°, 255°]
EPL	Belgian offshore cluster	removed [202°, 228°]
VVDR	Belgian offshore cluster	removed [243°, 255°]
IJM	Eneco Luchterduinen, Prinses Amalia, Egmond aan Zee	removed [249°,261°] and [300°, 311°] and [330°, 348°]
AWG1	Riffgat	removed [55°, 59°]
HGT	Riffgat	removed [0°, 36°]
TNW	Gemini, Deutsche bucht, Veja Mate, Bard Offshore 1	removed [27°, 93°]
LA	Gunfleet Sands, Thanet wind farm	removed [130°,145°] and [308°,340°]

Table S1: Description of the measurement filtering for the presence of wind farms.

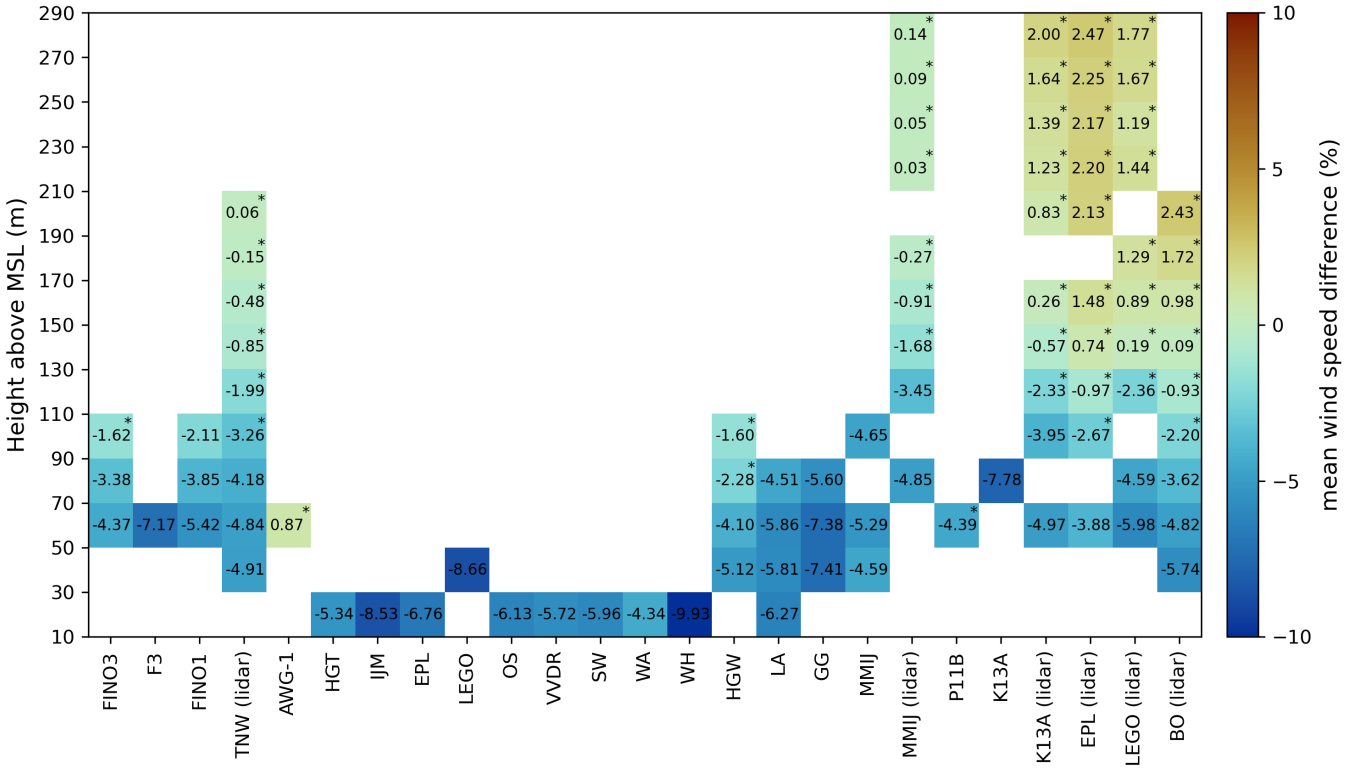

Figure S5: Wind speed bias (%) for the complete time period of the measurement data. This concerns measurements between 10 and 290 meters to MSL. The vertical range is subdivided into 20 meter intervals for readability. The presence of an asterisk indicates that the bias is within the measurement uncertainty. Stations are clustered per region.

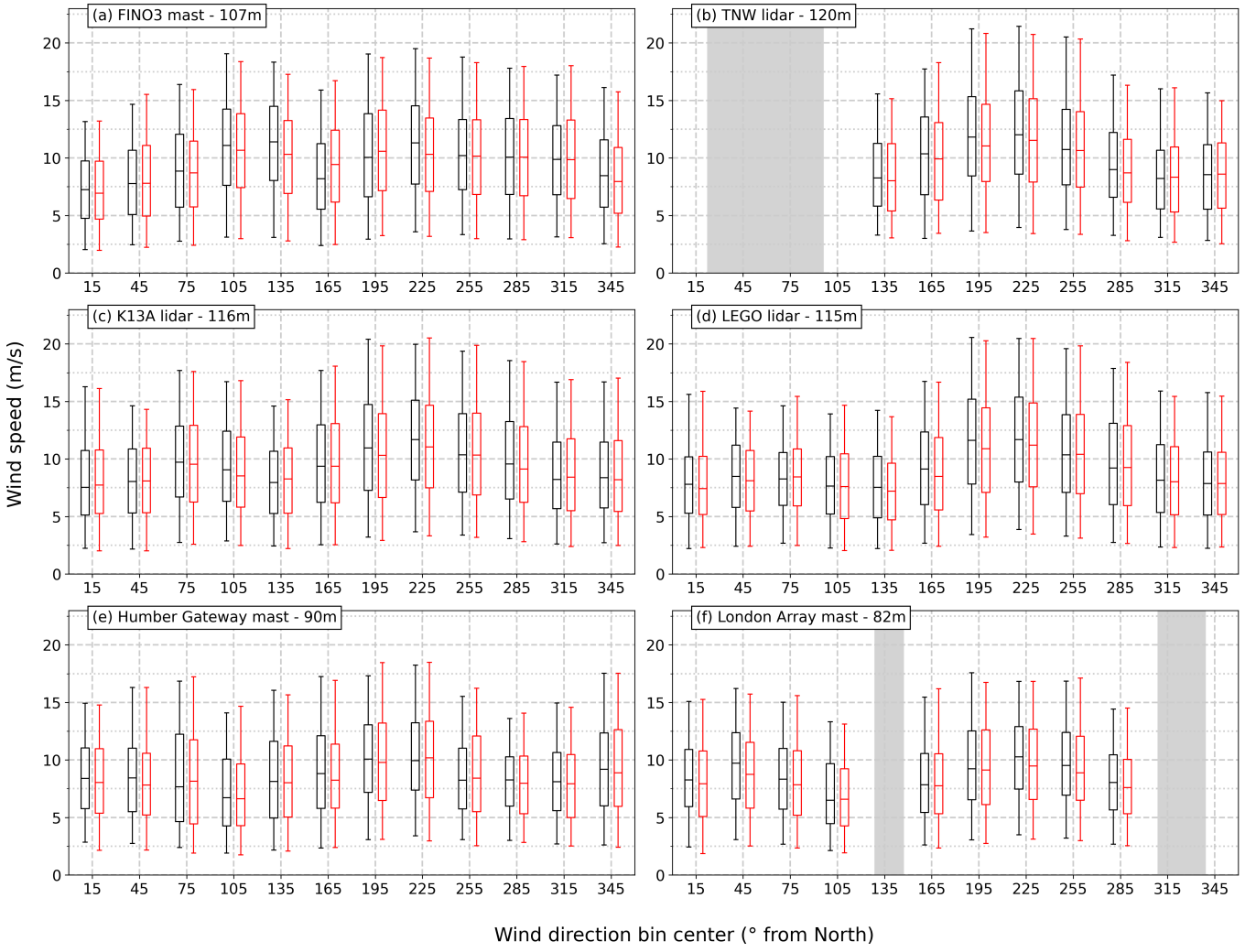

Figure S6: Boxplots representing the wind speed distribution per 30° wind direction bin for the observations (black) and the model (red). Shown for the three masts and three lidar stations at turbine hub height. The box corresponds to the Q25-Q50-Q75 wind speeds. The lower and upper whisker are the Q5 and the Q95 percentiles, respectively. Shaded regions correspond to sectors that were filtered out due to the presence of wind farms within a distance of 50km.

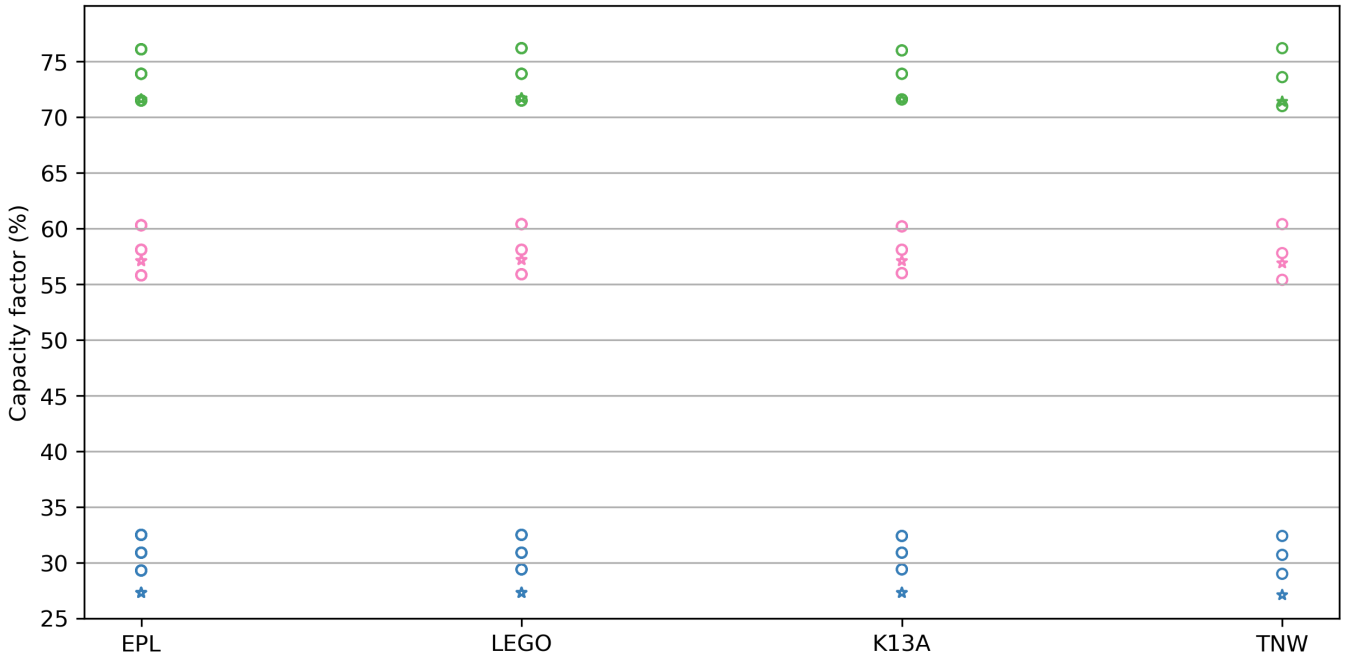

Figure S7: Capacity factor values (%) inferred from the wind speed distributions under different stability conditions for lidar observations (hollow circles) and the corresponding model grid cell (asterisk). The stability classes are unstable (pink), weakly stable (green) and stable (blue). For the observations, upper and lower uncertainty bounds on the capacity factor are provided based on the wind speed uncertainty.

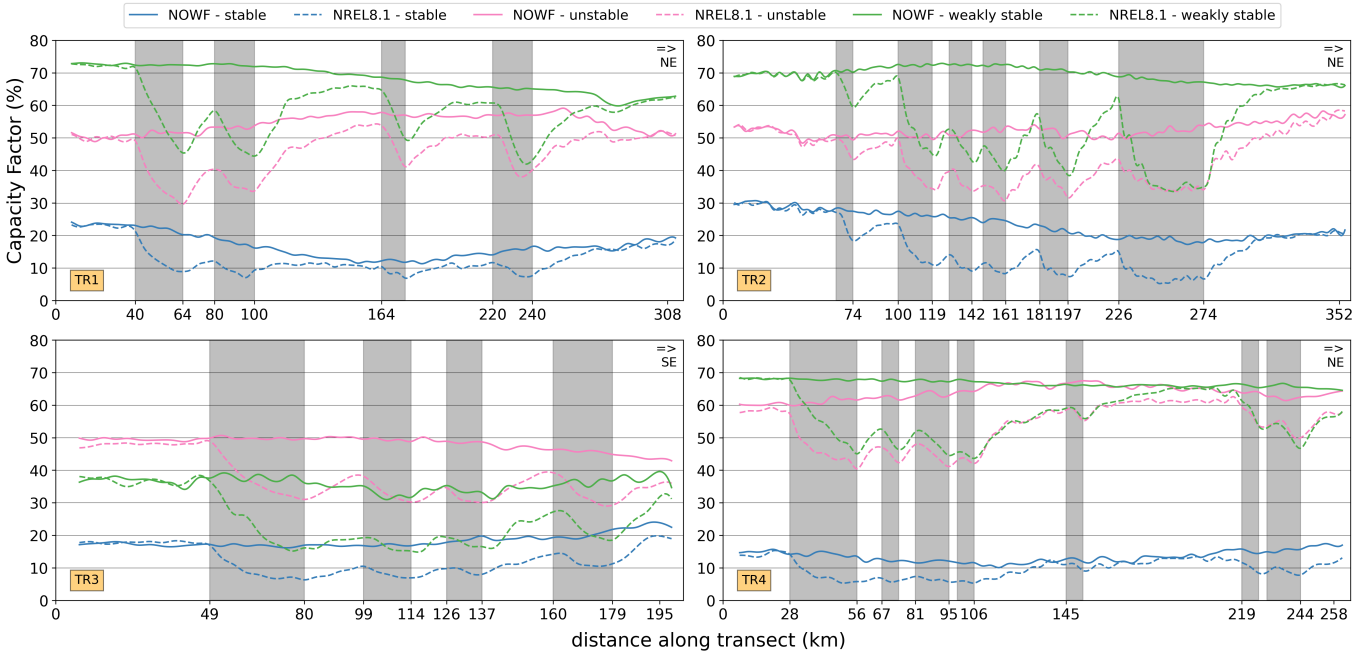

Figure S8: 2016 capacity factor transects for the four lines indicated on Fig. 1. for the NOWF (full line) and PD (dashed line) scenario, subdivided in the three dynamic stability classes: unstable (pink), weakly stable (green) and stable (blue). The X-axis is in km. Wind data is only considered when the wind direction deviates within +-15° from the transect orientation (W to E) at the middle grid cell of each transect. Grey shadings represent wind farms.

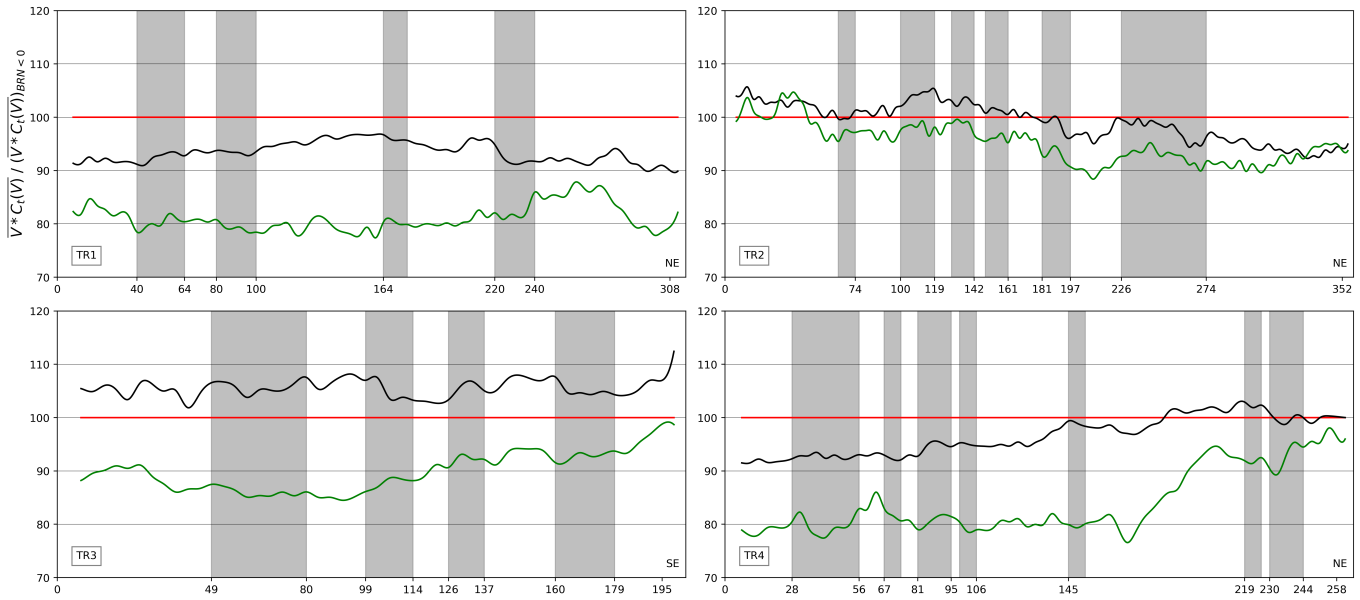

Figure S9: Along-transect mean profiles of $V * C_t$ for the different stability classes relative to the unstable case (NOWF scenario) for the four transects indicated on Fig. 1. This is based on the NREL 5MW turbine thrust curve. Values larger than 100% correspond to locations where the relative wind speed reduction due to the wind farm is larger than for the unstable case. Values below 100% correspond to locations where the relative wind speed reduction due to the wind farm is less than for the unstable case.%). The three stability classes are unstable (red), weakly stable (black) and stable (green). The X-axis is in km. Wind data is only considered when the wind direction deviates within +-15° from the transect orientation (W to E) at the middle grid cell of each transect. Grey shadings represent wind farms.