Supplement for Machine Learning Methods to Improve Spatial Predictions of Coastal Wind Speed Profiles and Low-Level Jets using Single-Level ERA5 Data

Table S1: ERA5 variables chosen as possible predictors for the two ML models and two prediction tasks.

ERA5 variable	Abbreviation	Unit	Importance for wind profile (physical motivation)
10 m wind speed	ws10	m s^{-1}	Wind speed conditions at lower levels are frequently coupled to upper wind speed conditions and are also associated with atmospheric stability regimes, synoptic scale weather events and sea state
10 m wind direction	wdir10	deg	wdir10 is associated with inherited properties of the air (onshore or offshore fetch), synoptic weather regimes, as well as meso-scale weather events
Sea surface temperature	SST	K	Temperature at the sea surface, describes seasonality and diurnal cycles and influences the atmospheric stability
Mean sea level pressure	MSLP	Pa	MSLP is linked to the synoptic regime
Total precipitation	precip.	m h^{-1}	Depth of all precipitation that falls to the surface, includes information about stability, synoptic scale weather events, seasonal variability and humidity
Convective available potential energy	CAPE	J K^{-1}	A measure of potential instability aloft derived for a lifted parcel, as a description of convection and general stability of the troposphere
Surface sensible heat flux	SHF	J m^{-2}	SHF describes the turbulent transport of heat between the surface and the atmosphere, as a description of stability and the diurnal and seasonal variability
Surface net radiation	Rn	J m^{-2}	Total sum of all shortwave and longwave radiation, both incoming and outgoing, as a description of seasonal and diurnal variability of surface energy availability
Low cloud cover	LCC	–	Fraction of the grid box covered by low clouds, as a description of seasonal and diurnal variability, mesoscale phenomena and synoptic scale weather conditions
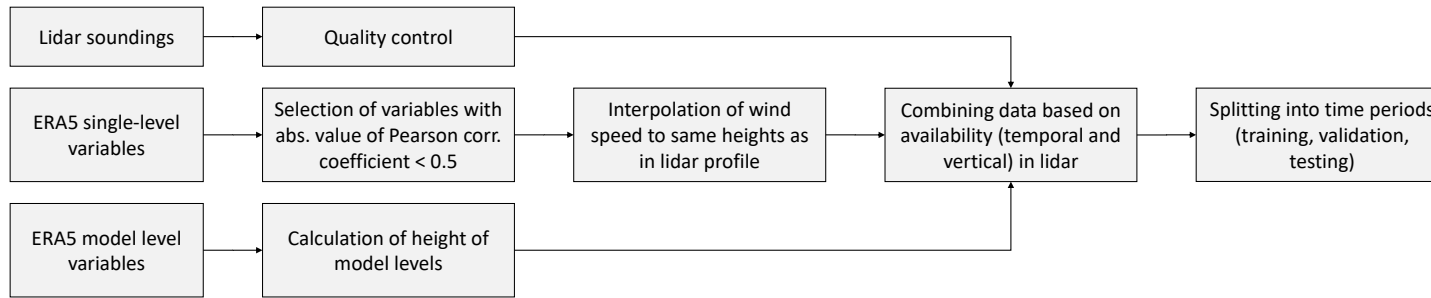
Figure S1: Flow chart of pre-processing of data from lidar profiles, ERA5 single-level variables and data on ERA5 model levels, before the ML methods were applied.
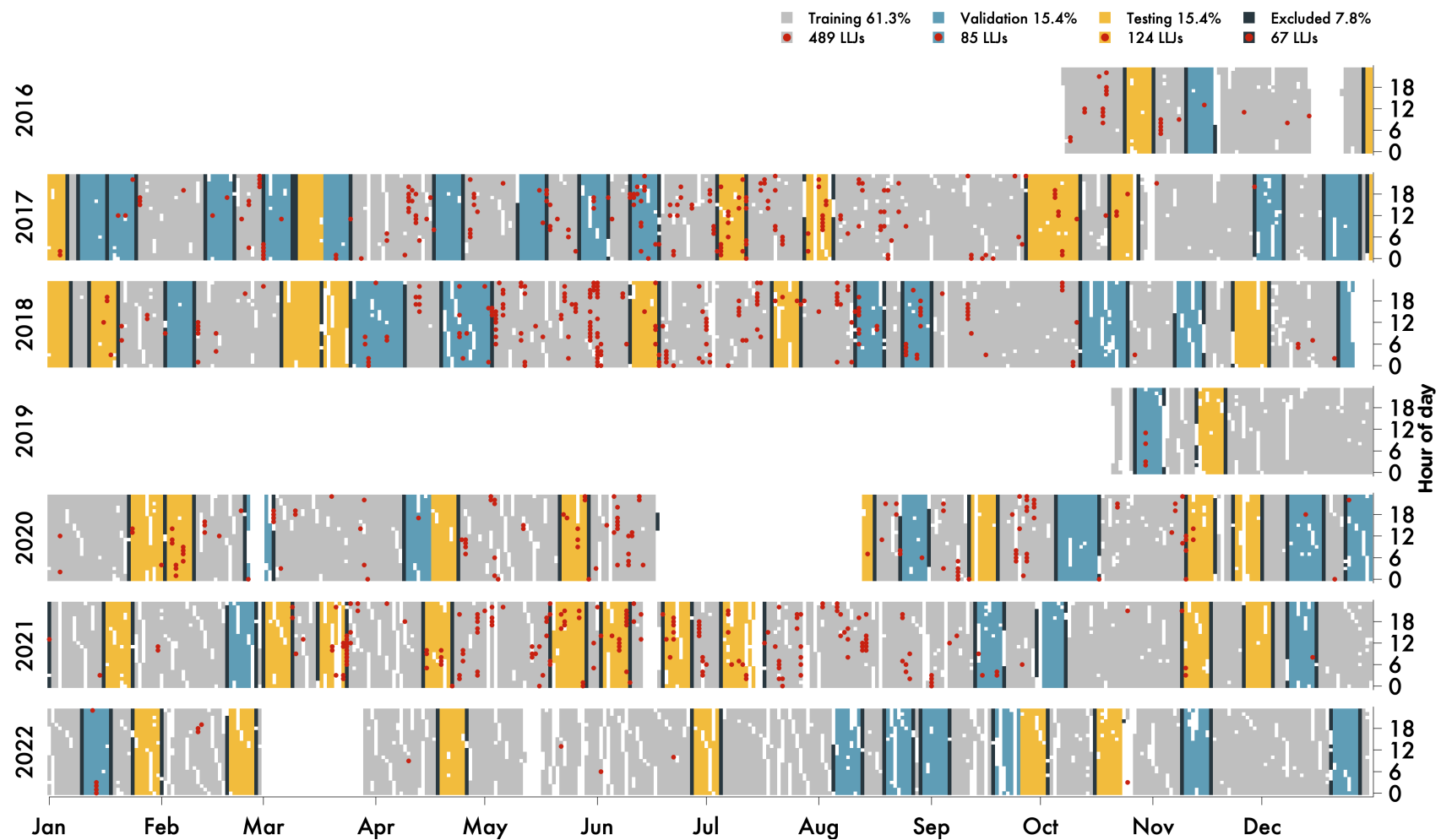
Figure S2: Overview of the data split for the ASIT time series (Oct. 2016 – Dec. 2022). The data is split into training (light gray), validation (blue), and testing (yellow) subsets and to minimize temporal auto-correlation one day of data before and after blocks of validation and testing are excluded (dark gray). Hours where there is an LLJ in the lidar profile are marked with red circles.
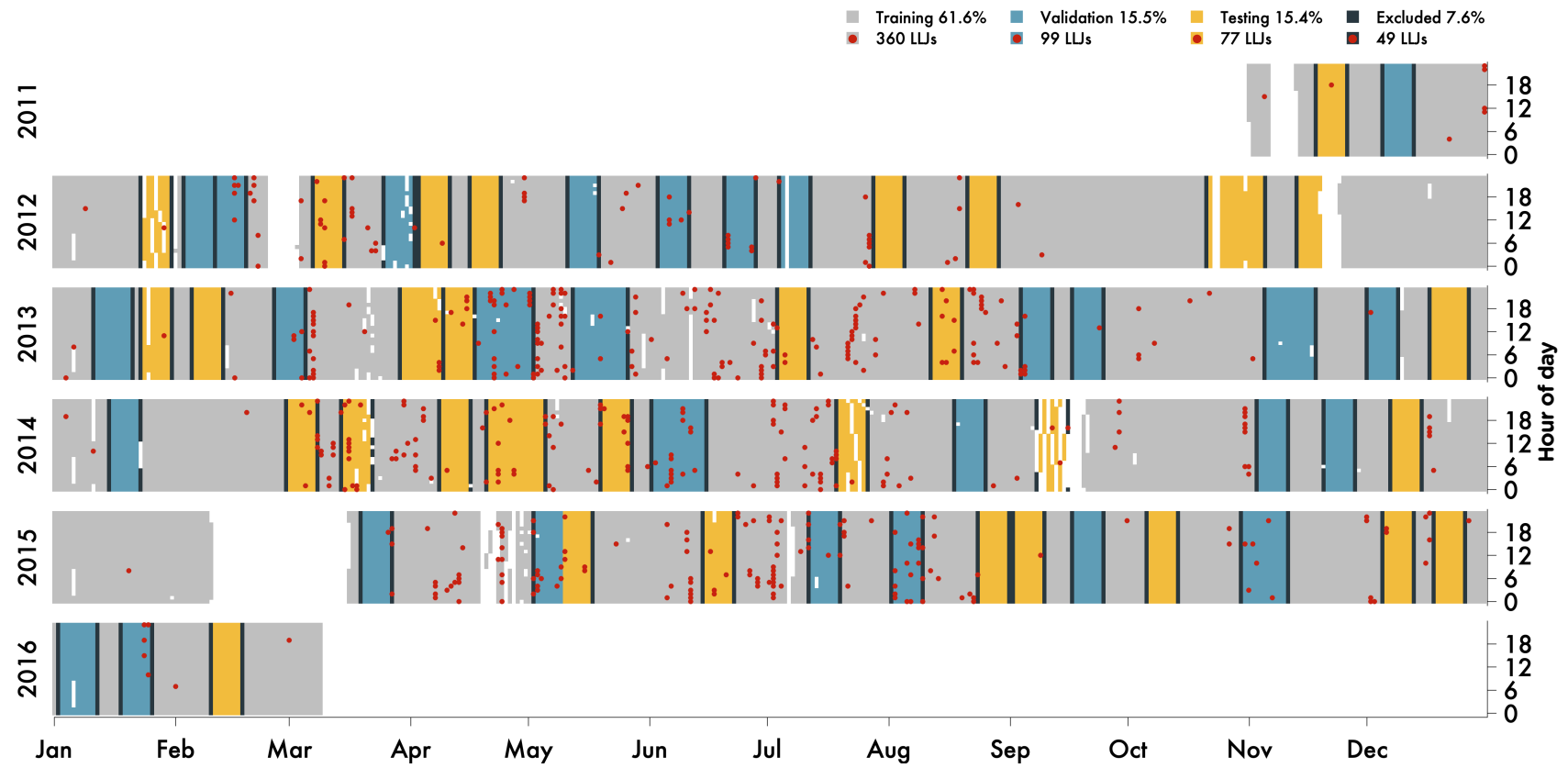
Figure S3: Overview of the data split for the MMIJ time series (Nov. 2011 – Mar. 2016). The data is split into training (light gray), validation (blue), and testing (yellow) subsets and to minimize temporal auto-correlation one day of data before and after blocks of validation and testing are excluded (dark gray). Hours where there is an LLJ in the lidar profile are marked with red circles.