

Assessing Inter-Model Agreement in Convection-Permitting Simulations of Extreme Winds for Wind Energy Applications

Nathalia Correa-Sánchez¹, Xiaoli Guo Larsén², Eleonora Dallan¹, Marco Borga¹, and Francesco Marra³

¹Department of Land Environment Agriculture and Forestry, University of Padova, Padova, Italy.

²Department of Wind Energy, Technical University of Denmark, Roskilde, Denmark.

³Department of Geosciences, University of Padova, Padova, Italy.

Correspondence: Nathalia Correa-Sánchez (nathalia.correasanchez@phd.unipd.it)

Abstract.

Convection-permitting models (CPMs) have great potential for wind energy applications such as wind energy planning, turbine design loads, and operational safety in a climate change context. In fact, compared to coarse-resolution models, they have an improved representation of atmospheric processes and surface characteristics that directly influence winds at heights relevant for turbines. Evaluating how well CPMs perform at reproducing extreme wind events is crucial for wind energy applications. Inter-model comparisons provide insights into uncertainties and enhance the credibility of CPM-based applications. Here, we use a new framework to examine the agreement among three CPMs from the CORDEX Flagship Pilot Study in simulating extreme wind speeds in central Europe. We categorise locations by climate, roughness, and topography; then use Principal Component Analysis to quantify inter-model agreement and divergences; and apply the Simplified Metastatistical Extreme Value (SMEV) method to the 10-year simulations to estimate 50-year return levels (U_{50}) required for wind turbine design. Our results show large agreement between the models, with the first Principal Component explaining 74.2% of the total variance and indicating a strong consensus in extreme wind patterns, despite systematic differences in magnitudes. Stronger agreement emerges during the winter, when extreme winds are driven by synoptic conditions, and less concordance during summer, when localised convective events cause most extremes. Our research emphasises the value of using CPM ensembles over single-model assessments of extreme winds, and establishes a methodological framework that provides the wind energy community with baseline information on the CPM capabilities and limitations in estimating wind speed extremes.

Copyright statement. TEXT

1 Introduction

Wind energy is a key part of Europe's shift to renewable energy, providing nearly 19% of the continent's electricity (European Commission, 2025). A successful wind energy development requires an accurate assessment of extreme wind speeds at turbine hub heights, typically around 100 m, as they are essential for project planning, turbine design loads, and operational safety considerations (Ma et al., 2022). Climate models have served as an effective tool for these applications by enabling wind

resource assessments across diverse geographical regions, providing spatially consistent datasets for extreme value analysis, and supporting long-term planning under changing climate scenarios (Prein et al., 2015; Archer and Jacobson, 2005). In climate studies, multi-model ensembles are the primary means of obtaining accurate estimates of model uncertainties (Prein et al., 2015; Fosser et al., 2024). Regional variability in extreme wind changes under future scenarios exhibits both intensification and attenuation depending on season and location (Pryor et al., 2020; Outten and Sobolowski, 2021; Ma et al., 2025), underscoring the need for high-resolution climate modelling in wind energy planning and assessment

In recent years, global and regional climate models (GCMs and RCMs) have served as essential tools for understanding large-scale atmospheric flows. However, while GCMs (~ 100 km horizontal grid spacing) capture synoptic scales and RCMs (~ 10 km horizontal grid spacing) extend into larger mesoscales, neither can explicitly resolve convective processes and represent fine-scale surface heterogeneity that affects local wind behaviour and extremes (Soares et al., 2012; Prein et al., 2015; Prein, 2023). For wind resource assessment applications specifically, recent operational datasets have adopted similar fine-scale resolutions (2–3 km grid spacing), including the Global Wind Atlas (Davis et al., 2023), the Wind Toolkit for the continental United States (Draxl et al., 2015), and the offshore-focused NOW-23 dataset (Bodini et al., 2024). In contrast, Convection-Permitting Models (CPMs) provide high-resolution simulations with ~~a grid spacing of less than~~ grid spacing ≤ 4 km ~~where, in which~~ deep convection is explicitly resolved ~~instead of being rather than~~ parameterised (Prein et al., 2015; Schär et al., 2020), thereby offering an alternative to the limitations of GCMs and RCMs. Originally developed to improve the representation of future extreme precipitation, CPMs also bring added potential benefits for other atmospheric variables, such as winds at 100 m. Although these CPM projections remain limited by aspects such as boundary layer turbulence parameterisations or sub-scale momentum transport schemes, they enable a more accurate representation of surface characteristics and their interactions with surface winds, and explicitly resolve processes of gust front propagation and downdrafts associated with convective storms (Manning et al., 2022; Prein, 2023), and also preserve the natural spectral characteristics of the wind in the high mesoscale frequency range (Correa-Sánchez et al., 2025b).

The CORDEX Flagship Pilot Study (CORDEX-FPS) framework on convective phenomena (Coppola et al., 2020) developed the first multi-model ensemble of convection-permitting simulations in the Euro-Mediterranean region. Its simulations are generated for historical conditions as well as near- and far-future projections, eventually offering opportunities for medium- and long-term planning. This framework provides the CPM ensemble analysed in the present study. Previous research that relied on the wind speed field from CPMs has focused mainly on average conditions (Ruiz et al., 2022; Ma et al., 2022), rather than on extreme wind speed. While Chen et al. (2024) showed significant improvements using CPMs compared to coarser resolution simulations for renewable energy assessment, methodological challenges emerge from other studies. For example, Frisius et al. (2024) recognises the biases that may be introduced by vertical extrapolation when working only with simulations of winds at 10 m, and focuses on the changes in the average wind energy metrics instead of extreme events. Ma et al. (2025), instead, used return level estimations based on threshold-exceedances to study destructive winds in Canada. However, this work is based on relatively few exceedance samples, which may introduce important uncertainties in the subsequent extrapolations. In part, the extreme value extrapolations with CPMs are limited by the short simulations available, typically on the order

of 10 years. This occurs because CPMs are computationally expensive simulations (Prein et al., 2015), and conducting the multi-decadal simulations needed for extreme value analysis is highly costly.

60 An effective use of CPMs for wind energy requires addressing key methodological challenges. First, the limited availability of long-term wind observations at hub heights restricts direct and absolute validation of CPM performance. We cover this by comparing models using Principal Component Analysis (PCA) to break down the variance structure into consensus signals and differences. This method provides a relative assessment that shows where CPMs agree, which builds confidence in climatological patterns, and where they differ, which highlights structural uncertainties.

Second, estimating extreme return levels like U_{50} from short CPM simulations, usually spanning 10-20 years, brings sta-
65 tistical challenges. Traditional extreme value methods, like Generalised Extreme Value (GEV used for block maxima) and Generalised Pareto Distribution (GPD, used for threshold exceedances), are well-documented in wind studies (Donat et al., 2010; Pfahl, 2014; Harris and Cook, 2014; Ma et al., 2025). However, these methods are sensitive to record length, making parameter estimation less reliable with shorter time series (Larsén et al., 2015; Coles et al., 2001). We address this issue using the Simplified Metastatistical Extreme Value (SMEV) approach (Marra et al., 2019). SMEV looks at the entire distribution of
70 independent events rather than just annual maxima or threshold exceedances, making it more reliable for limited time series (Dallan et al., 2024). SMEV has been successfully used for extreme precipitation from in-situ observations, remote data, and CPM simulations (Miniussi and Marra, 2021; Dallan et al., 2023; Correa-Sánchez et al., 2025a; Marra et al., 2022). Here, we apply SMEV for the first time to wind-speed analysis, expanding its use to wind energy applications.

Third, surface variability influences wind patterns at hub heights due to interactions between the surface and the atmosphere.
75 Climate types, surface roughness, and topography affect boundary-layer dynamics through differential heating, roughness effects, and turbulence generation (Courault et al., 2007; Avissar and Schmidt, 1998). These factors impact winds at a height of 100 m. Therefore, we create a spatial categorisation framework based on climate, roughness, and topography. This framework allows for a systematic assessment of CPM performance across different surface conditions relevant to wind energy.

Inter-model agreement assessment is especially important when examining CPMs, because there are several ways to imple-
80 ment the model schemes and resolve convection (Coppola et al., 2020; Abramowitz et al., 2019), which can result in differences in how they represent wind fields, providing an indicator of prediction uncertainty, and recognising that the consensus among different models reflects consistency rather than absolute accuracy (Bellucci et al., 2015). Notably, these schemes may lead to diverse representations of wind fields over surfaces with different characteristics. Considering these factors, ensemble approaches have proven effective in assessing model uncertainty, as they use a set of individual models to detect model-specific
85 biases and reduce overall prediction errors (Yan and Tang, 2013). Assessing the agreement of multi-model ensembles provides information about the ability to distinguish between robust signals (where models agree) and uncertain predictions (where models diverge), thereby offering insight into the reliability of the simulated patterns (Yan and Tang, 2013).

To advance in renewable infrastructure planning, not only point-based validation with observations for operational design is required, but also spatially complete extreme wind climatologies for regional planning and site prospecting across data-sparse
90 areas are needed. Our work addresses the second need by estimating extreme wind events at turbine height (100 m) using SMEV through an assessment of the relative agreement among three CPMs, establishing a methodological framework for wind energy

applications. To do so, we structured an approach with four key objectives. First, we conduct a domain-wide analysis examining inter-model differences in extreme wind representation across the full study area. Second, we establish spatial categories based on climate, surface roughness, and topographic features to account for heterogeneous surface-atmosphere interactions that modulate wind patterns. Using this categorisation framework, our third objective quantifies spatio-temporal agreement among the three CORDEX-FPS CPMs in simulating turbine-height wind speeds across different surface conditions. Fourth, we apply SMEV to estimate extreme wind speeds from short simulations for each spatial category. We then discuss how CPM ensemble approaches provide added value for wind extremes and may contribute to long-term wind energy planning.

2 Study area and data

The study domain extends from 0.5°E to 16.3°E in longitude and from 40.2°N to 49.7°N in latitude, covering approximately 250,000 km² in central Europe (Figure 1). The region encompasses considerable topographical diversity, ranging from the Mediterranean sea, the flat plains of the Po Valley and northern France, to elevations exceeding 4,000 m a.s.l. in the Alpine region. It spans a range of climatic conditions, from Mediterranean influences in the south to continental regimes in the north. There are also complex land-sea interactions along the Mediterranean and Adriatic coastlines within this domain, providing varied surface characteristics that modulate wind patterns at turbine-relevant heights, creating a suitable scenario for testing CPM performance across different wind generation mechanisms.

We focus on wind speed at 100 m height, as it represents the international standard for wind turbines (Davis et al., 2023; National Renewable Energy Laboratory, 2023; Petersen, 1997). We analyse data from three CPM members of the CORDEX-FPS (Coppola et al., 2020) for which such a variable was made available at the time of the study (Table 1). Specifically, wind speeds were derived directly from the zonal (u 100 m) and meridional (v 100 m) wind components, available at hourly frequency, for the reanalysis period of 2000-2009. All three CPM models provide consistent hourly outputs representing instantaneous model states at each hourly output time (i.e., snapshots at specific hours, not hourly averages or hourly maxima). In all models, ERA-Interim (Dee et al., 2011) provides consistent initial and lateral boundary conditions, ensuring that inter-model differences arise from model physics and stochastic variability only, with no contribution from the boundary forcing. Intermediate-resolution regional climate models (RCMs, with a 12 km horizontal resolution) are nested within ERA-Interim, with the CPMs subsequently nested within these RCMs. The models rely on two atmospheric cores: CMCC and ETH use the COSMO model, while CNRM uses AROME. For representing the planetary boundary-layer (PBL), CMCC and ETH use the TKE-based scheme (Mellor and Yamada, 1982; Raschendorfer, 2001), while CNRM employs the CBR turbulence scheme (Cuxart et al., 2000; Bougeault and Lacarrere, 1989) with PMMC09 for shallow convection (Pergaud et al., 2009). The models have different vertical levels (ETH: 60 levels, CNRM: 60 levels, CMCC: 50 levels) and horizontal diffusion methods: CMCC uses fourth-order Smagorinsky hyper-diffusion, ETH has no explicit horizontal diffusion, and CNRM uses semi-Lagrangian horizontal diffusion (Adinolfi et al., 2020; Leutwyler et al., 2016; Caillaud et al., 2021; Correa-Sánchez et al., 2025b). Even though CMCC and ETH share the COSMO framework and PBL scheme, they differ in computational implementation (ETH: GPU-accelerated COSMO; CMCC: standard COSMO-CLM), horizontal resolution, vertical levels, and horizontal diffusion

125 methods (Leutwyler et al., 2016; Adinolfi et al., 2020), which may affect how they represent convective processes. All models follow the CORDEX-FPS protocol with continuous free-running simulations. Table 1 summarises model cores, convection treatment, PBL schemes, vertical levels, and diffusion methods. To enable direct comparison, all CPM outputs were remapped onto a common 3 km regular grid covering the study domain using a bilinear interpolation. This method preserves the spatial continuity of the wind field while keeping the values within a plausible physical range.

Table 1. CPM members used for inter-model agreement assessment, showing reference names, original resolutions, coupled RCM configurations, and key technical specifications. All models were remapped to a common 3 km grid for comparison.

Institute	CPM	RCM	Atmospheric Core	PBL Scheme	Vertical Levels	Horizontal Diffusion
CMCC	CCLM	CCLM	COSMO	TKE-based	50	Fourth-order Smagorinsky hyper-diffusion
Euro-Mediterranean Center on Climate Change	3 km (Adinolfi et al., 2020; Rockel et al., 2008)	12 km (Adinolfi et al., 2020; Rockel et al., 2008)		(Mellor and Yamada, 1982; Raschendorfer, 2001)		
CNRM	CNRM- AROME41t1	CNRM- ALADIN63	AROME	CBR + PMMC09	60	Semi-Lagrangian (SLHD)
Centre National de Recherches Météorologiques	2.5 km (Caillaud et al., 2021)	12 km (Nabat et al., 2020)		(Cuxart et al., 2000; Bougeault and Lacarrere, 1989; Pergaud et al., 2009)		
ETH	COSMO- crCLIM	COSMO- crCLIM	COSMO	TKE-based	60	None (implicit)
Institute for Atmospheric and Climate Science	2.2 km (Leutwyler et al., 2016; Rockel et al., 2008)	12 km (Leutwyler et al., 2017; Rockel et al., 2008)	(GPU- accelerated)	(Mellor and Yamada, 1982; Raschendorfer, 2001)		

We describe here our analysis of the CPM ensemble, which combines a spatial stratification into climate-roughness-topography categories (Section 3.1), quantification of inter-model agreement and divergences using Principal Component Analysis (PCA) and seasonal correlations (Section 3.2), and estimation of extreme wind speeds using SMEV (Section 3.3).

3.1 Spatial categorisation

135 In order to understand how inter-model differences in wind extremes at 100 m relate to surface characteristics and provide a spatial comparison context for subsequent analyses, we develop a spatial categorisation based on climate patterns, roughness, and topography.

On the one hand, climate classification systems capture this heterogeneity, as different Köppen-Geiger climate zones exhibit different surface wind characteristics due to contrasting thermal regimes, land-ocean distributions, and seasonal circulation patterns that modify boundary-layer processes across spatial scales (Beck et al., 2023; Ulbrich et al., 2009). On the other hand, Petersen (1997) showed that wind power meteorology needs to consider other surface-atmosphere interactions that influence superficial wind flows, such as surface roughness effects and orographic acceleration. Orographic lifting and thermal circulation modify local wind patterns (Smith, 1979), while surface roughness controls boundary-layer dynamics and turbulent mixing (Kaimal and Finnigan, 1994), including the modulation of convective processes (Avissar and Schmidt, 1998; Kang and Bryan, 145 2011). An analogous categorisation was implemented in the European Wind Atlas methodology (Troen and Lundtang Petersen, 1989) and further developed with modern multivariate categorisation schemes in the New European Wind Atlas (Dörenkämper et al., 2020).

3.1.1 Climate

Climate is the primary classification category, because extreme winds strongly depend on large-scale atmospheric flows, circulation weather types, North Atlantic Oscillation (NAO) phases, cyclone tracks, mean sea level pressure patterns, and pressure gradients (Donat et al., 2010; Pfahl, 2014). We use the Köppen-Geiger climate classification by Beck et al. (2023), which represents the 1991-2020 period at high resolution (1 km). Köppen-Geiger climate zones reflect large-scale atmospheric circulation patterns that also influence regional wind characteristics. After remapping to a 3 km resolution to match the CPM common grid, we obtained 13 distinct climate types in our study domain. To ensure statistical robustness and a more straightforward interpretation of the results, we grouped these 13 climate types into four levels: Arid (Ar), Temperate (Tm), Cold (Co), and 155 Tundra (Td), as shown in Fig. 1a. This aggregation balances interpretability with preserving the major atmospheric circulation regimes that control extreme wind generation across Europe (Ulbrich et al., 2009; Pfahl, 2014).

3.1.2 Roughness

Surface roughness at a particular location is modulated by local terrain characteristics, with surface roughness influencing how 160 wind speed varies with height, as described by models such as the logarithmic wind profile (Stull, 1988; Petersen, 1997); greater

roughness causes larger wind retardation near the ground. Surface roughness categorisation was based on the roughness length (z_0) data that has been used as input to the COSMO model, and was derived from the methodology of Demuzere et al. (2008), which associates roughness length values with land cover types from the CORINE Land Cover Map (European Environment Agency, 2020). This approach aligns with the roughness classification framework established in the European Wind Atlas (Troen and Lundtang Petersen, 1989), ensuring consistency with established wind energy assessment practices. We defined five roughness levels ranging from R_1 (very smooth) to R_5 (very rough), representing the spectrum from water surfaces and open terrain to dense urban areas and forests (Fig. 1b). Detailed information of the roughness levels is in Table 2.

Table 2. Roughness classes based on European Wind Atlas with logarithmic scale of aerodynamic roughness length (z_0). Categories range from very smooth surfaces (R_1) to very rough terrain (R_5), following standard wind energy assessment protocols.

Roughness Level	Start z_0 [m]	End z_0 [m]	Centroid (approx.)	Description (Examples)
R_1 : Very smooth	0	0.0003	0.0001	Very smooth surfaces (water, ice, fresh snow)
R_2 : Smooth	0.0003	0.03	0.003	Smooth surfaces (short grass, compacted snow, ploughed fields)
R_3 : Moderate	0.03	0.3	0.1	Moderately rough surfaces (grasslands, crops, vineyards)
R_4 : Rough	0.3	0.7	0.5	Rough surfaces (low forests, urban areas with low scattered buildings)
R_5 : Very rough	0.7	∞	> 1	Very rough surfaces (dense forests, urban centres with tall buildings)

3.1.3 Topography

Topography has a significant influence on wind flow near the surface and up to heights relevant for wind turbines. Topographic patterns cause wind acceleration near ridge crests and deceleration in the valleys (Jackson and Hunt, 1975; Troen and Lundtang Petersen, 1989). The slope variance can be used as a metric for terrain complexity, able to capture the presence of significant elevation changes (Riley et al., 1999), indicating areas where these acceleration/deceleration effects are most likely and pronounced.

We use the ETOPO 2022 global relief dataset at 15 arc-second resolution (MacFerrin et al., 2025), which was remapped to 3 km resolution to match the CPM common grid. Using terrestrial elevation values (> 0 m), we calculated slope for each grid cell and subsequently computed slope variance by comparing each cell's slope against its eight neighbouring cells within a 3x3 moving window. To objectively define topographic complexity levels, we applied the Jenks natural breaks optimisation algorithm (Chen et al., 2013), which iteratively partitions the slope variance data into classes that minimise within-class variance while maximising between-class variance, thus identifying natural clustering boundaries in the data distribution. This

180 classification yielded four topographic levels: T_1 (Flat), T_2 (Gentle), T_3 (Moderate), and T_4 (Complex), representing increasing degrees of terrain complexity, as shown in Fig. 1c.

3.1.4 Combined Classification

The three categorical layers defined above (climate, roughness, topography) were combined pixel-by-pixel to obtain 52 possible total category combinations (Figure 1d). Each grid point receives an identifier composed of the level of each of the three classification layers, representing a unique environmental signature. This layer combination generates composite codes following the format [Climate level][Roughness level][Topography level], where each component retains its original categorical designation. For example, ArR_3T_4 represents a location characterised by arid climate (Ar), moderate surface roughness (R_3), and complex topographic conditions (T_4).

The spatial distribution of the combined categories reveals distinct clustering patterns that reflect the underlying surface heterogeneity in the domain. Inner-continental areas exhibit the greatest diversity of spatial categories, ranging from moderate-roughness and flat terrain in the northern plains to high-roughness mountainous regions in the Alps and other elevated areas. Coastal and marine environments are represented by specific categories that capture land-sea transitions and varying degrees of surface roughness associated with different coastal morphologies.

190

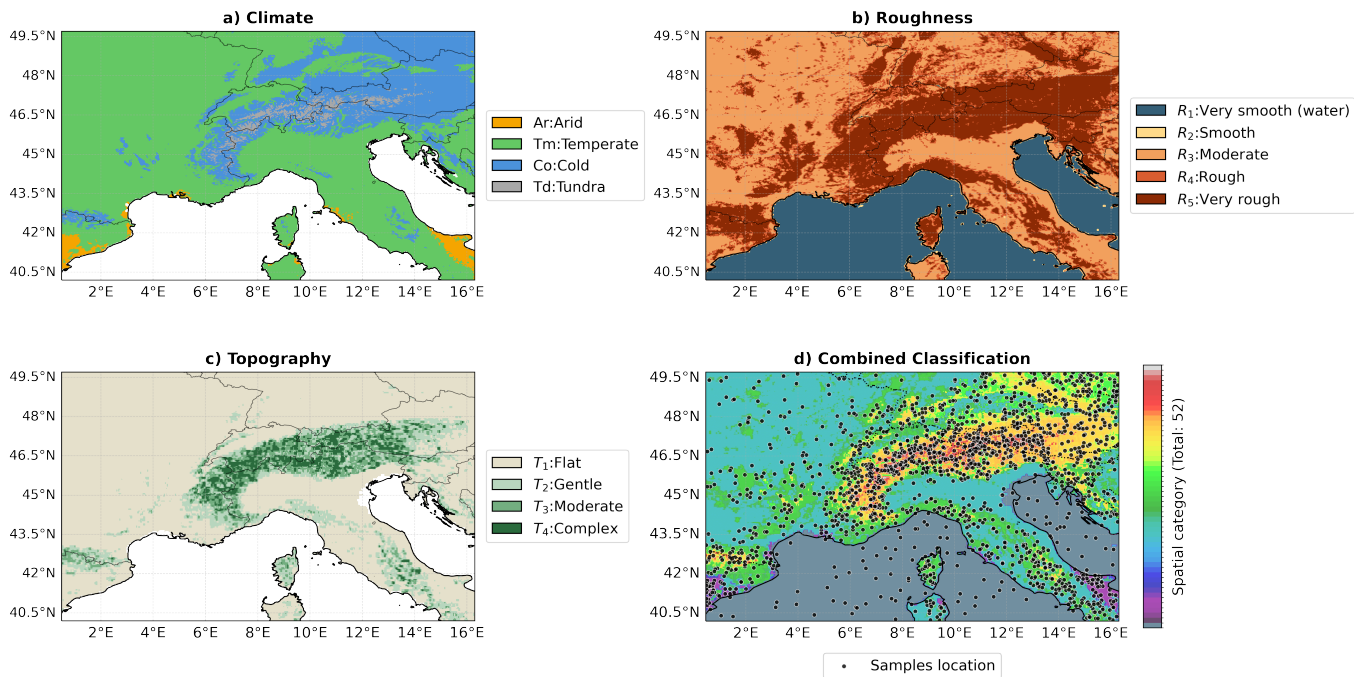


Figure 1. Spatial categories and their respective levels a) Climate, b) Roughness c) Topography. A combination of the 3 layers in panel d) with 52 total resulting categories within the domain.

3.1.5 Stratified random sampling

195 The stratified random sampling is a sampling technique that incorporates two key principles: firstly, stratified sampling ensures the representation of all spatial categories, with random selection that mitigates spatial autocorrelation and clustering effects; secondly, equal number selection was applied to ensure that each spatial category contributed an identical number of time series, thereby providing equal analytical weight regardless of the category's spatial extent within the domain.

Given the need to focus on predominant representative conditions, a frequency analysis was conducted to identify the most
200 prevalent spatial combinations. The categories were ranked by their relative frequency within the domain, and the 17 most spatially abundant categories were selected, collectively representing 97.7% of the domain's pixels (Figure 2). The 35 less frequent categories, which collectively contribute less than 2.3% of domain coverage, were excluded to focus on predominant conditions. This coverage-based selection maintains an adequate representation of the diverse geographic conditions of the region (Figure 2) while ensuring that rare or transitional combinations with limited spatial extent do not bias the extremes
205 CPMs assessment.

Spatial randomisation was implemented within each of the dominant categories to mitigate potential spatial autocorrelation effects and sampling bias that could arise from clustered or systematic point selection. Here, we selected 100 points per category to provide a statistically representative sample size that ensures statistical independence for a robust parameter estimation (Cochran, 1977) while maintaining computational feasibility for subsequent analyses. This stratified random sampling is stan-
210 dard in spatial statistics (Thompson, 2012; Webster and Oliver, 2007). This approach was employed to extract representative wind speed time series from each spatial category, ensuring statistical comparability across diverse geographical conditions. This allows the subsequent statistical analyses and performance metrics to be directly comparable across all spatial categories, enabling a robust assessment of CPM behaviour under varying surface-atmosphere interaction regimes.

The random selection yields a total of 5,100 complete time series covering the entire 10-year period (17 categories $\times$ 3
215 models $\times$ 100 points) and representing the various surface characteristics within the study domain (Figures 1 and 2). The selected locations are shown in Fig. 1d, while the statistical distribution of all spatial categories, and the most frequent ones, can be found in Fig. 2.

3.2 Inter-model agreement assessment

Principal Component Analysis (PCA) offers a framework for multi-model agreement evaluation, by decomposing the temporal
220 co-variability of multiple model time series into orthogonal components that distinguish common signals from model-specific divergences (Smith et al., 2012). Similarly, inter-model correlations of the seasonal variations provide insights into when and why CPM members differ throughout the year.

3.2.1 Principal Component Analysis for inter-model agreement

Principal Component Analysis (PCA) has been extensively employed in atmospheric sciences to identify coherent spatio-
225 temporal structures and modes of variability within Earth system datasets (Schmidt et al., 2019). This methodology focuses

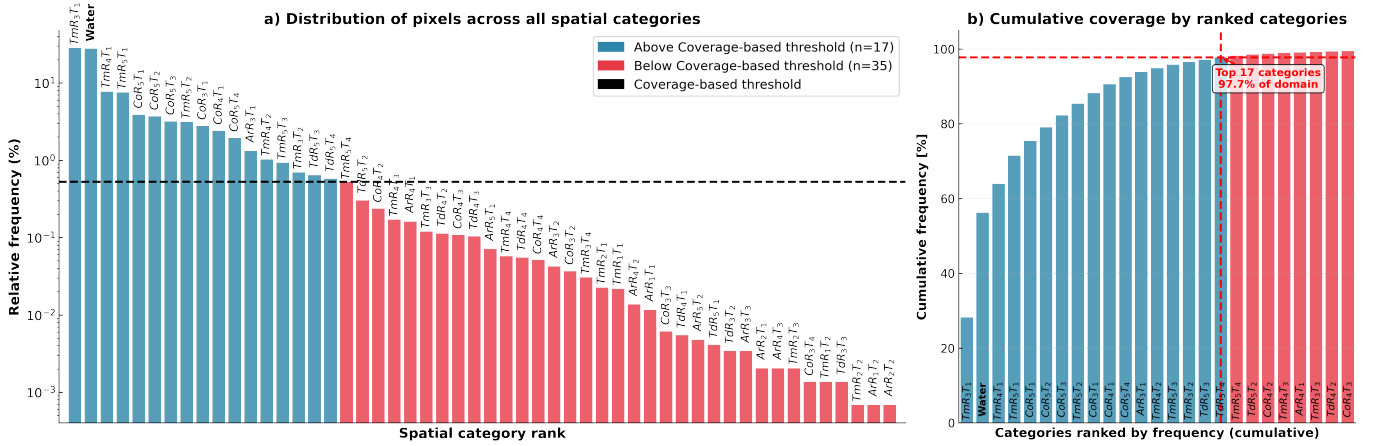


Figure 2. Spatial category distribution and filtering strategy. a) Distribution of pixel relative frequencies across all 52 spatial categories ranked by abundance. Blue bars indicate the 17 most spatially abundant categories retained for analysis, which collectively represent 97.7% of the domain. Red bars show the 35 less frequent categories (collectively 2.3% of the domain) excluded to focus on predominant conditions. The log scale allows visualisation of both dominant and rare categories spanning four orders of magnitude. b) Cumulative frequency distribution of spatial categories ranked by abundance. The top 17 categories (blue) capture 97.7% of the domain, validating the filtering approach for the following analyses.

on identifying consistent signals and systematic divergences among models rather than making an absolute validation against observations (Bellucci et al., 2015). This multivariate approach addresses limitations of correlation analyses, which assess only pairwise relationships rather than the complete inter-model interaction structure. High inter-model agreement, measured through PCA loadings, may serve as a proxy for inter-model consistency (Smith et al., 2012; Yan and Tang, 2013), as independent models with strong consensus indicate reproducibility across different model formulations. However, this consistency does not guarantee accuracy without observational validation.

Following the mathematical framework presented in Jolliffe (2002), we treat each model as a variable in the analysis, enabling simultaneous quantification of consensus signals and divergences across the ensemble. We transform the three-dimensional space defined by the CPM simulations (ETH, CNRM, CMCC) into three principal components that capture the complete variance structure of inter-model relationships. At each grid point, the three time series represent the same meteorological variable (wind speed at 100 m) but simulated by different models under identical atmospheric boundary forcing (ERA-Interim), enabling PCA to decompose the pure inter-model covariance structure that reveals systematic patterns in how models co-vary. The first principal component represents shared variance (consensus), while subsequent components capture structured disagreements between specific model combinations.

PCA was applied to daily maximum wind speed time series derived from hourly 100 m wind speed data (2000-2009) simulated by each of the three CPMs (Section 3.1.5). Daily maximum wind speeds are calculated as the maximum hourly wind speed for each day. To address the right-skewed distribution typical of wind data while ensuring mathematical validity for zero

wind speeds, daily maximum wind speeds are transformed using $WG_{\log} = \log(1 + WG)$, where WG represents the original daily maximum wind speed values in m s^{-1} . Each time series was then standardised (by subtracting the mean and dividing by the standard deviation) to ensure equal contribution to the variance structure. Here we use a daily aggregation since it serves two methodological purposes for PCA: 1) It substantially reduces temporal autocorrelation inherent in hourly data that would cause components to reflect temporal dependencies rather than inter-model covariance, and 2) It eliminates the diurnal cycle that would mix a common atmospheric signal with actual model differences, which allows PCA to decompose inter-model relationships in significant wind events (Jolliffe, 2002).

The loadings represent linear coefficients that quantify how each CPM contributes to each component, where $PC_1 = (\text{loading}_{\text{ETH}} \times \text{ETH}) + (\text{loading}_{\text{CNRM}} \times \text{CNRM}) + (\text{loading}_{\text{CMCC}} \times \text{CMCC})$ (Jolliffe, 2002). PC_1 loadings indicate each model's contribution to the consensus signal, with balanced positive loadings ($\approx \frac{1}{\sqrt{3}} \approx 0.577$) indicating equal model contribution to the shared variance. PC_2 and PC_3 loadings reveal which models drive primary and secondary divergence patterns, respectively, with loading magnitude and sign indicating the direction and strength of each model's departure from the consensus. This loading-based analysis, applied across the 100 sampled points within each spatial category, enables systematic characterisation of how inter-model agreement varies across different spatial categories.

PCA shows the temporal concordance among the model simulations at each spatial location. The use of this method, in conjunction with stratified sampling across spatial categories, can provide an identification of locations where models show strong consensus versus areas where their differences reflect fundamental uncertainties in representing wind speed patterns. This approach provides insights into inter-model consistency and reproducibility. While this does not guarantee accuracy without observational validation, it indicates that results are not arbitrary or dependent on implementation. High agreement among models based on different physics schemes suggests potentially higher confidence in their outputs. Nevertheless, observational validation is necessary to determine whether this consistency reflects reality or shared systematic bias.

3.2.2 Seasonal agreement

While PCA analyses general patterns in the time series, direct monthly correlations reveal seasonal changes in agreement or identify which specific model pairs differ during certain months due to different atmospheric processes, such as winter storms or summer convection.

Seasonal patterns are relevant for wind energy applications, and assessing the agreement of the models on a seasonal basis may reveal periods of higher model uncertainty that could affect the reliability of extreme wind projections, especially during months when convective processes dominate wind generation mechanisms. To evaluate seasonal patterns of inter-model agreement and identify temporal dependencies in CPM performance for extreme wind events, we conducted a monthly correlation analysis of wind speed 99th percentile across the 10-year study period. This approach aims to capture how models consistently reproduce monthly wind extremes across inter-annual variations, providing insights into the temporal alignment of CPM performance across diverse European geographical and climatic contexts. Moreover, this analysis complements the PCA approach by providing direct quantification of pairwise model agreement at a monthly resolution, enabling identification of seasonal periods when model consensus is strongest or weakest.

For each spatial category and calendar month, monthly maximum wind speeds were extracted from hourly time series and used to compute Pearson correlations between model pairs (ETH-CNRM, ETH-CMCC, CNRM-CMCC) throughout the decadal period. Correlations are calculated at fixed grid points and therefore reflect temporal co-occurrence at specific locations rather than spatial displacement patterns. Then, point-wise correlations were spatially averaged within each category to obtain representative monthly agreement profiles.

3.3 SMEV approach for wind speeds

The Simplified Metastatistical Extreme Value (SMEV) approach, proposed by Marra et al. (2019), relies on a different set of assumptions with respect to classical extreme value theory, which enables one to use the entire set of independent events, termed (temporally uncorrelated local maxima termed “ordinary events”, see Section 3.3.1), instead of annual maxima or peaks over threshold only.

Namely, SMEV assumes that the class of the parent distribution describing the ordinary events is known. It then estimates the parameters of this distribution from all the available independent realisations and explicitly accounts for their finite annual occurrence frequency, avoiding the asymptotic assumptions of classical extreme value theory (Marra et al., 2019). Therefore, SMEV accounts for both the distribution of wind-speed magnitudes (through shape and scale parameters) and the annual frequency of occurrence of the events, thereby providing a more direct physical interpretation of the atmospheric processes generating extreme winds (Marra et al., 2019). For the case of wind speed, the Weibull distribution represents a natural choice for the parent distribution, as supported by numerous previous studies (Cook et al., 2003; Cook, 2014; Harris and Cook, 2014; Tuller and Brett, 1984; Tye et al., 2014; Hennessey Jr, 1977; Palutikof et al., 1999) and has been shown to effectively approximate distributions with theoretical foundations in atmospheric physics (Harris and Cook, 2014).

3.3.1 Ordinary events identification for wind speeds

Ordinary events were identified through an iterative peak selection algorithm that identifies independent local maxima within the time series through a decorrelation-based separation. The largest value in the series is extracted as an ordinary event, and all the preceding and subsequent values within a temporal window defined by the temporal correlation function, including the identified maximum, are set to zero. The procedure is then iterated until the entire time series is explored.

The decorrelation time for the wind speed time series was calculated using the area-under-the-curve method (Kristensen et al., 2002; Larsén and Mann, 2006) applied to the autocorrelation function as in Eq. 1:

$$\tau_d = \int_0^{L_{\max}} R(\ell) d\ell \quad (1)$$

where $R(\ell)$ is the autocorrelation function at lag ℓ , and $L_{\max} = 200$ hours. The autocorrelation was computed at 1-hour intervals from 0 to 200 hours to accurately estimate the decorrelation time, which defines the temporal separation required for identifying independent events in the extreme value analysis. Figure C1 illustrates the patterns of latitudinal variation in decorrelation on a uniform grid of points for each model.

3.3.2 Weibull distribution to describe the tail of the ordinary events

Following the theoretical foundation established by Harris and Cook (2014) and Cook et al. (2003) regarding the appropriateness of Weibull distributions for parent wind speed populations, we use a two-parameter Weibull distribution to model the independent wind peaks. The Weibull cumulative distribution function $F(x) = 1 - \exp[-(x/\lambda)^\kappa]$, where x represents wind speed values, λ and κ represent scale and shape parameters, respectively, was fitted using least-squares linear regression in Weibull-transformed coordinates, following the methodology outlined by Marani and Ignaccolo (2015), whose codes are available from Marra et al. (2019). To address potential biases arising from low-magnitude events that may not represent the upper tail characteristics, we follow Marra et al. (2023) to implement a left-censoring in which the upper portion of the distribution of the ordinary events is treated as exactly known in the estimation, while the values below the threshold are treated as non-exceedances. It follows that left-censoring is not equivalent to threshold exceedance approaches such as the peak over threshold of extreme value theory (in which the values below the threshold are thrown away). Based on the Marra et al. (2023) approach, we selected the top 10% of ordinary events (90th percentile threshold), balancing the need for sufficient sample size while ensuring representative characterisation of the upper tail that governs extreme value behaviour. Moreover, this approach allows for assessing the adequacy of the Weibull distribution through Weibull probability plots, with coefficients of determination (R^2) exceeding 0.9 for all CPM ordinary events in different locations. To illustrate, Weibull probability plots from 12 randomly selected locations of each CPM are provided in the supplementary material (Figures D1, D2, D3; Table D1 shows their location and the spatial category to which they correspond). The linear relationship in ln-ln space validates the Weibull distribution assumption for the upper tail across diverse geographical and climatic conditions.

3.3.3 Estimation of extreme wind return levels

Once ordinary events were identified and Weibull parameters estimated, extreme wind speeds for specified return periods were calculated using the SMEV formulation. The SMEV approach estimates return levels through the relationship $\text{SMEV}(x) = [F(x; \theta)]^n$, where $F(x; \theta)$ represents the fitted Weibull distribution with parameters $\theta = (\lambda, \kappa)$ and n is the mean annual frequency of ordinary events (Marra et al., 2019). For each spatial category and model combination, return levels corresponding to 50-year return periods were computed using the quantile expression (Eq. 2):

$$U_T = \lambda \left[-\ln \left(1 - \left(1 - \frac{1}{T} \right)^{1/n} \right) \right]^{1/\kappa} \quad (2)$$

where U_T represents the extreme wind speed for return period T , and the inner term $(1 - 1/T)$ defines the annual non-exceedance probability. The mean annual frequency n was calculated as the average number of ordinary events per year across the 10-year simulation periods. It should be noted that SMEV return levels are more sensitive to the scale parameter λ (approximately linear relationship: $\delta q/q \approx \delta \lambda/\lambda$) than to the shape parameter k (logarithmic dependence through exponent $1/k$) (Marra et al., 2019; Siena et al., 2023). This differential sensitivity is critical for interpreting inter-category differences (see Fig. E1). The estimation procedure was applied independently to each of the three CPM datasets (ETH, CNRM, CMCC).

All computations were performed at the individual grid point level before aggregation to spatial category statistics, preserving the full spatial variability information within each category. The codes to test the SMEV assumptions and apply it are freely available from Marra (2020) and Marra (2022).

A bootstrap resampling was used to derive 95% confidence intervals for ensemble mean return levels (U_{50}) for each disaggregated spatial category level shown in Fig.6 (Tibshirani and Efron, 1993). Ensemble means were first calculated at each point by averaging the three CPM return level estimates. Then, for each spatial category level (e.g., Climate “Temperate” - Tm), all points from spatial categories containing that characteristic were aggregated, and the confidence intervals were computed by spatially resampling the ensemble means from all points within the category (with replacement, 1000 iterations), computing the mean of each bootstrap sample, and extracting the 2.5th and 97.5th percentiles. This quantifies spatial sampling uncertainty within categories.

4 Results

4.1 Inter-model differences in wind extremes

Figure A1 shows that CPMs with different bulk statistics (panel a) can produce similar extreme wind behaviour (panel b, winds >25 m/s), confirming the need for dedicated extreme-value analysis in wind engineering. Given this, we first quantify baseline differences in extreme wind representation among CPMs by analysing the spatial distribution of mean and coefficient of variation (CV, standard deviation divided by the mean) of annual maximum wind speed. These results are shown in Fig. 3, which reveals preliminary differences among the three CPMs. ETH exhibits systematically higher wind speeds across the inland domain, with average annual maxima ranging from 15-20 m s⁻¹ in low-lying areas to over 35 m s⁻¹ in mountainous regions and coastal zones. CNRM produces intermediate values with a more moderate spatial gradient, while CMCC consistently generates the lowest annual maxima, particularly in complex terrain where values remain below 20 m s⁻¹. All models seem to reproduce a clear orographic enhancement and land-sea contrast effects on the extreme wind speeds. On the other hand, the CV reveals generally low inter-annual variability across all models (CV < 0.45), with CNRM showing higher variability than ETH and CMCC, particularly in mountainous regions. These spatial patterns remain consistent when disaggregated by individual surface categories (Figure B1), persisting across all spatial category levels.

The differences in magnitude and patterns across models reveal the systematic inter-model uncertainty in extreme wind representation, particularly in regions with complex topography where surface-atmosphere interactions are most challenging to simulate accurately.

4.2 Inter-model agreement

4.2.1 Principal Component Analysis

The principal component analysis reveals a strong common signal explaining 74.2% of the total variance (PC1), with systematic and consistent loading patterns across all spatial categories (Figure 4). There is a clear hierarchical contribution of CPMs

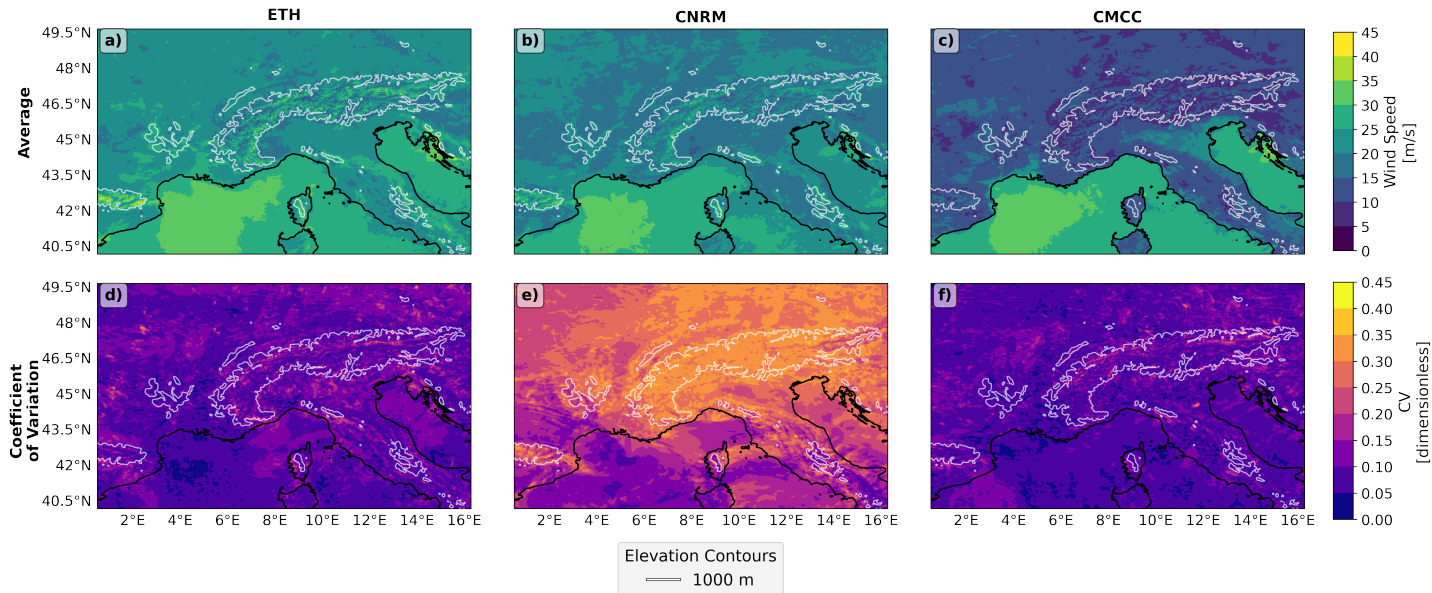


Figure 3. Annual maximum wind speeds: 10-year average (a-c) and coefficient of variation (d-f) for each pixel in ETH, CNRM, and CMCC models, respectively. The white outlines mark areas with elevations above 1,000 metres above sea level.

to this main signal. ETH contributes most strongly to the consensus signal by consistently exhibiting the highest positive loadings on PC1, exceeding the balanced loading threshold ($1/\sqrt{3} \approx 0.577$) across all categories, whereas CNRM demonstrates intermediate loadings with greater variability around the balanced threshold, and CMCC shows consistently lower loadings below the balanced line. Cold climates (*Co*) exhibit the greatest separation between models, whereas temperate climates (*Tm*) show the most consistent loadings across all CPMs.

The second principal component (PC2, 14.4% of variance) captures the primary divergence pattern in temporal covariance. ETH loadings are centred near zero, CNRM shows predominantly positive loadings, and CMCC negative loadings, indicating that CNRM and CMCC systematically diverge in opposite directions from the ensemble mean; i.e., when CNRM predicts higher wind speeds, CMCC predicts lower speeds, with ETH falling between them. These patterns of variability are consistent with the spatial patterns observed in Fig. 3 d-f.

The third principal component (PC3, accounting for 11.4% of the variance) reveals a secondary, and independent, divergence pattern, in which ETH exhibits negative loadings. At the same time, both CNRM and CMCC show positive loadings with greater variability across categories. This pattern extends almost uniformly across all spatial categories, with the most pronounced differences occurring in the arid climates category.

Overall, the loading patterns remain remarkably stable across different spatial categories for PC1, PC2, and PC3, suggesting that inter-model relationships are consistent regardless of surface characteristics.

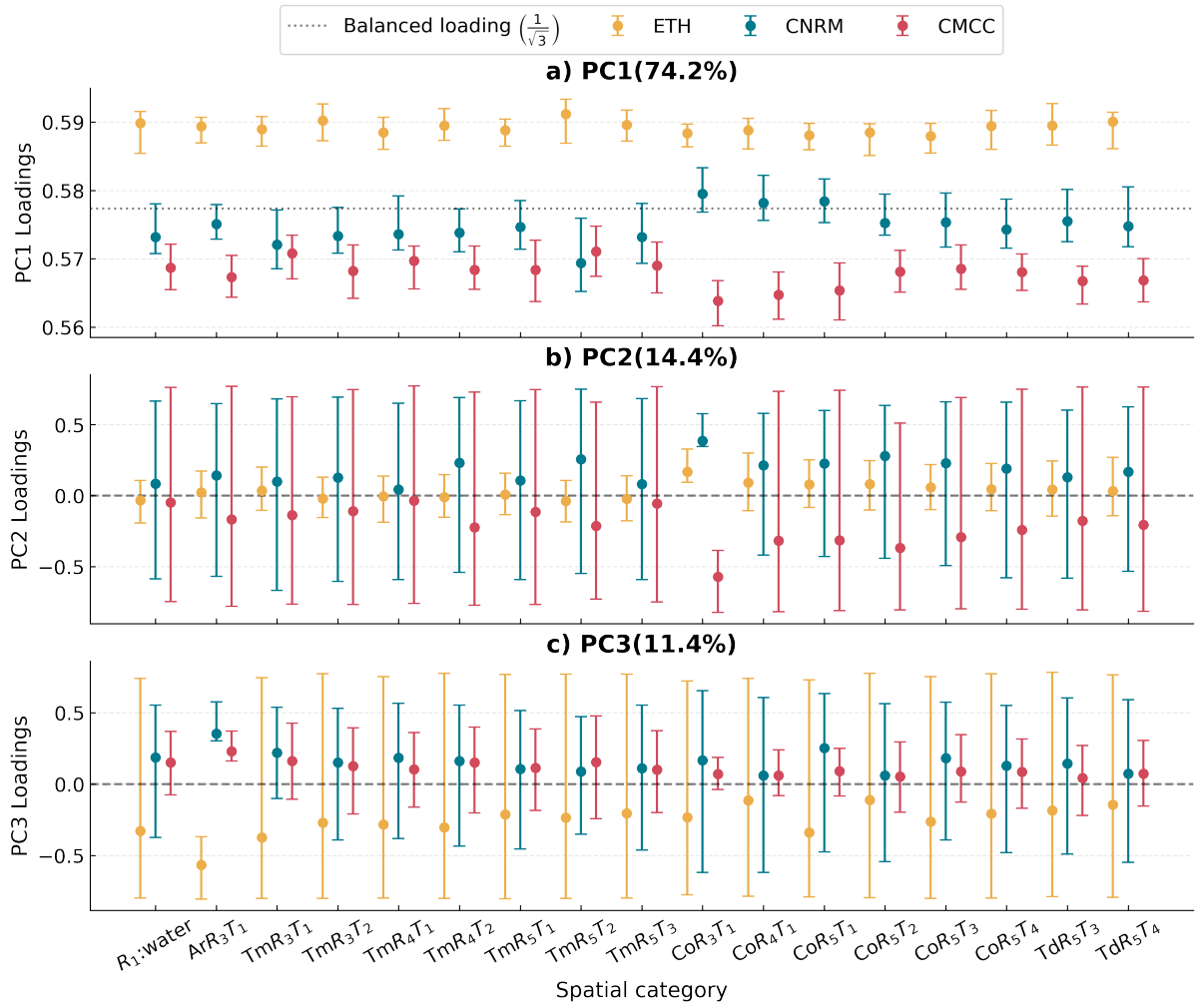


Figure 4. Principal components (PC) loadings distribution across spatial categories. a) PC 1, b) PC2, c) PC3. Points represent mean loadings, and error bars show the interquartile range (IQR) across the 100 sampled points within each spatial category.

4.2.2 Seasonal variability

Figure 5 reveals distinct seasonal patterns in the inter-model correlations of monthly 99th percentile wind speeds. The extended cold season, from October to March, shows the highest correlations and lowest inter-category variability, with values typically exceeding 0.7 across most categories, being lower in the cold and tundra climate categories during these months. The pair of
390 CNRM-CMCC models consistently exhibits the lowest correlations across all months, particularly during winter, consistent with the divergence patterns in Fig. 3. The summer months (June through August) exhibit moderately reduced correlations. May and September, as transitional months, show the widest inter-category correlation range (0.3-0.8) and greatest differentiation between model pairs, reflecting divergent representation of convective processes.

The seasonal correlation results (Figure 5) show that the CNRM-CMCC pair exhibits generally lower correlations than the
395 ETH-CNRM and ETH-CMCC pairs across most months and spatial categories, consistent with the PC2 divergence pattern, in which CNRM and CMCC systematically diverge while ETH remains near consensus across spatial categories. The spatial categories show modest influence on correlation magnitudes (typically ~ 0.1 -0.3 variation), with high-elevation complex terrain categories exhibiting slightly lower correlations during the winter and transitional periods. We also performed this same analysis with the monthly maximum, and the correlational patterns obtained were similar, yielding consistent model rankings
400 and seasonal patterns.

4.3 Extreme wind return levels

Figure 6 presents the distribution of 50-year wind return level (U_{50}) estimates. Figure 6a-c illustrates the inherent intra-model variability across disaggregated spatial categories. In d-f, the ensemble return levels (boxplot, average extreme wind estimates of each single CPM member) are compared with the sampling uncertainty (Bootstrap 95% confidence interval shown as grey
405 bands). Sample sizes vary substantially across categories (Fig. 6g-i), with temperate and cold climates providing the largest samples (700 and 600 points, respectively), while arid and polar climates contribute smaller samples (100 and 200 points). Roughness category R_5 dominates the domain with 900 points, while water surfaces (R_1) and R_2 provide limited and null representation. The topographic distribution is more balanced, though T_1 contributes the largest sample (700 points) compared to T_2 - T_4 (200-400 points each).

410 The differences among models remain uniform across spatial categories (Fig. 6a-c). ETH consistently produces the highest extreme winds across all spatial categories, showing the most consistent estimates within categories. Conversely, CMCC generates the lowest estimates, with differences of approximately 10 - 15 m s^{-1} in the central distribution among the models and has greater internal variability. CNRM has intermediate behaviour with moderate dispersion within categories, which is closer to ETH patterns.

415 The ensemble mean U_{50} return levels within their bootstrap confidence intervals (Fig. 6d-f) reveal differences in extreme wind speed across spatial categories. A clear progression can be noticed, with tundra (Td) and cold (Co) climates having higher ensemble means ($\sim 28 \text{ m s}^{-1}$) compared to arid (Ar) climates ($\sim 23 \text{ m s}^{-1}$). Water surfaces (R_1) show the highest inter-model consistency, while terrestrial roughness categories (R_3 - R_5) exhibit greater model disagreement despite similar

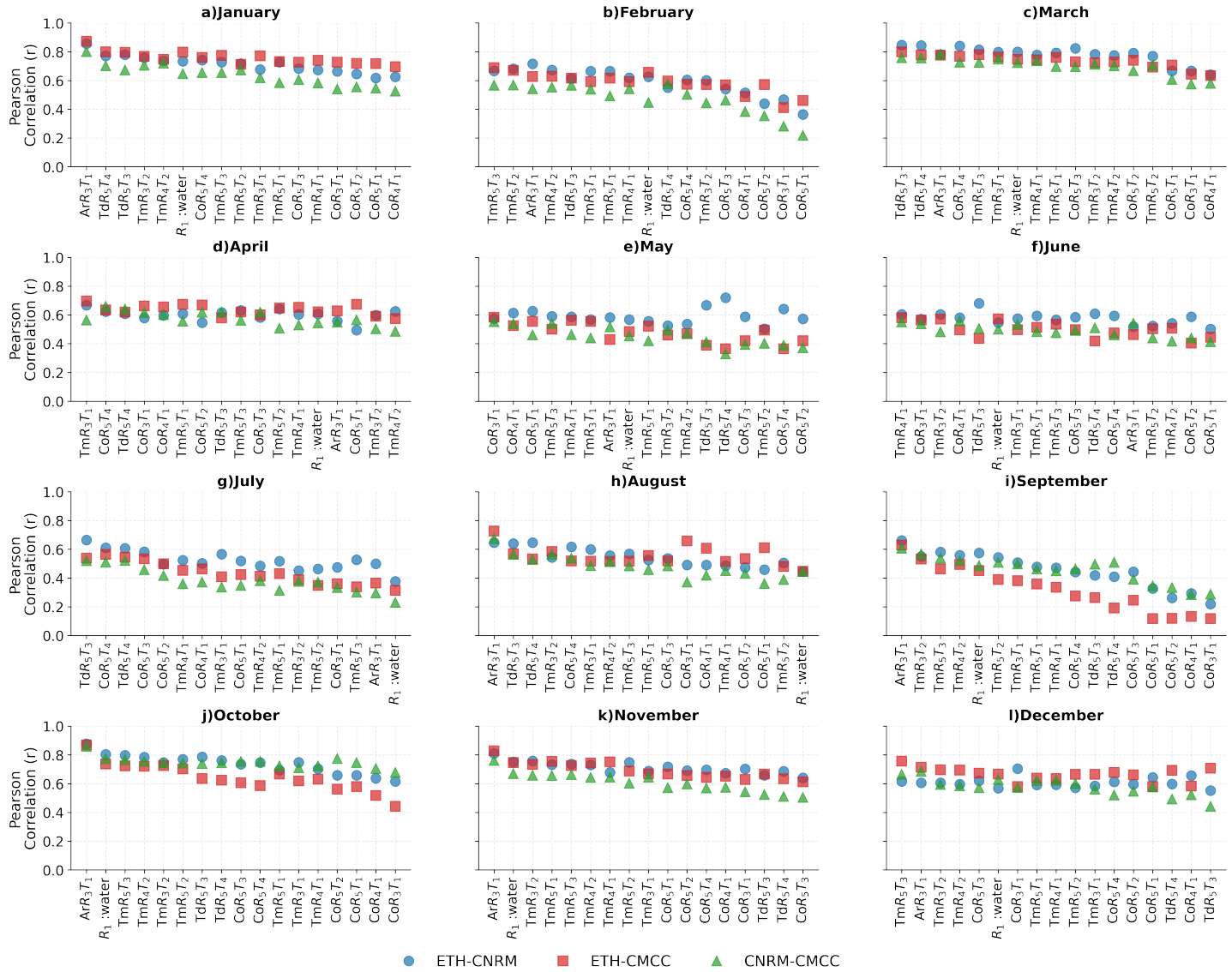


Figure 5. Monthly 99th percentile wind speed correlation coefficient per spatial category and CPMs couple.

ensemble means. Topographic categories demonstrate both similar ensemble means and comparable inter-model variability, with all terrestrial categories (T_1 - T_4) yielding consistent extreme wind estimates of 25- 27 m s^{-1} .

The ensemble approach captures the central tendency across the three CPMs, providing more balanced extreme wind estimates than using individual models. The bootstrap confidence intervals confirm the statistical robustness of the ensemble means across all spatial categories. Moreover, it shows consistent inter-model differences for all surface types. This indicates that the estimation uncertainty comes mainly from model formulation rather than surface-dependent processes.

425 To complement the U_{50} estimations in Fig. 6, we have analysed the distribution of the Weibull parameters across the disaggregated spatial categories (Fig. E1). The lower U_{50} estimates in R_1 primarily reflect an 18% lower scale parameter λ (median: 9.1 m/s vs 10.4–11.1 m/s in R_3 – R_5), while shape parameter k shows minimal variation (3–5%; Fig. E1).

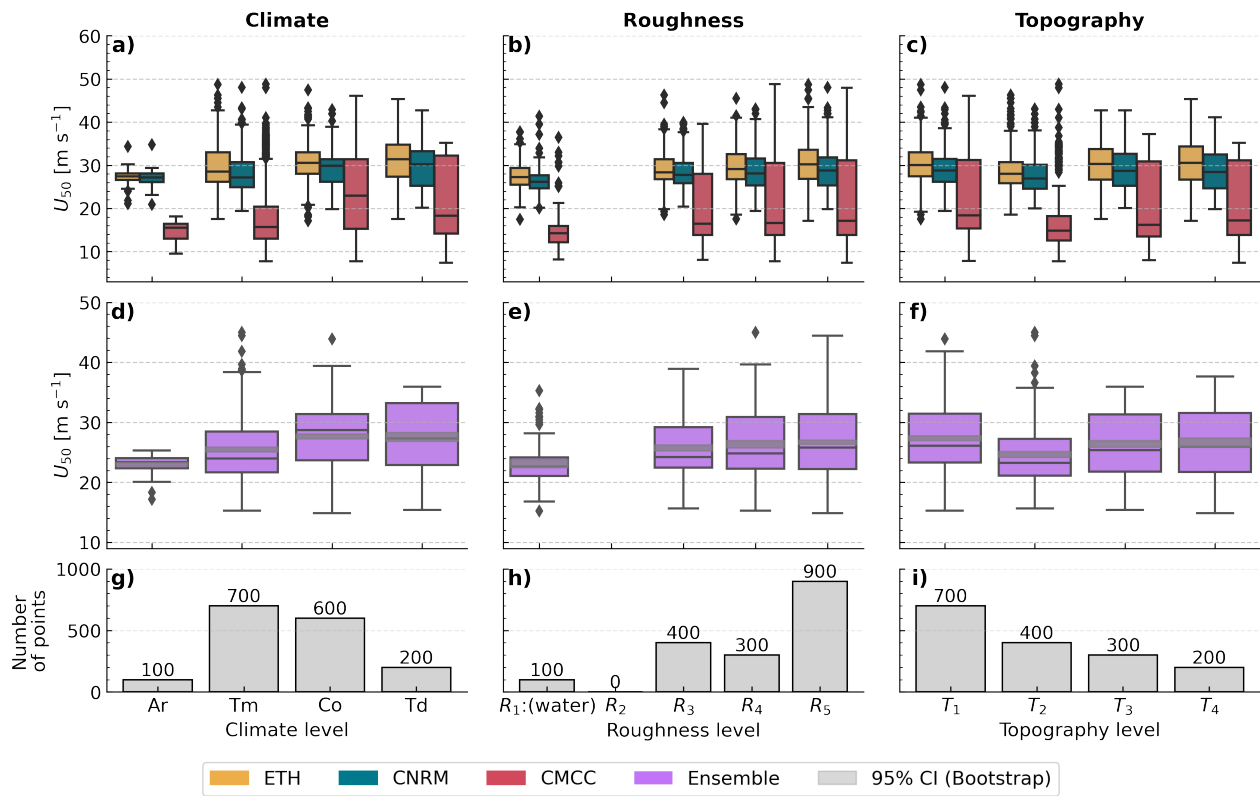


Figure 6. Wind return levels with 50 years return period estimated per model and spatial category (a-c). Distribution of the ensemble average 50 years of extreme wind speeds for each spatial category with 95% confidence intervals from bootstrapped samples (d-f). Number of points used per spatial category in the estimation of the extreme winds (g-i).

5 Discussion

5.1 Implications of inter-model differences

430 CPMs explicitly resolve convective processes, allowing each model to develop its own mesoscale meteorology, including convective systems, localised circulations, and fine-scale momentum transport processes (Coppola et al., 2020). One of the first findings is the systematic hierarchy in the annual maxima magnitudes observed in Figure 3, with $ETH > CNRM > CMCC$, suggesting that each model’s treatment of these explicitly resolved processes leads to systematically different extreme wind speeds at 100 m height.

435 Figure 3 provides information on the physical processes driving this model hierarchy. On one hand, continental areas, particularly those with mountainous terrain, exhibit the largest inter-model spreads in annual maxima, reflecting the complexity of representing orographic flows, channelisation effects, and topographically induced turbulence, where small-scale processes strongly influence wind field development. These differences in Fig. 3a-c can exceed $15\text{-}20\text{ m s}^{-1}$ over mountainous terrain, and are important because they could significantly alter turbine design loads and compromise structural safety analyses, making
440 the selection of a single CPM a critical source of uncertainty. On the other hand, marine environments show reduced inter-model variability, in line with more spatially uniform surface characteristics that reduce sensitivity to differences in surface-atmosphere coupling approaches. Moreover, the coefficient of variation patterns in Fig. 3d-f reveal that CNRM exhibits greater interannual variability in extreme wind speed, particularly in complex terrain, indicating higher sensitivity to meteorological forcing variations compared to ETH and CMCC. These findings challenge the reliability of isolated CPM applications for
445 extreme wind estimation and demonstrate that inter-model differences require an evaluation to distinguish robust physical signals from model-specific artefacts. Without an assessment of inter-model divergences, extreme wind estimates lack the statistical foundation necessary for engineering applications where structural integrity depends on accurate load projections.

5.2 Physical interpretation of model agreement patterns

The PCA (Fig. 4) reveals insights into the CPM representation of extreme wind processes across different surface characteristics, preserving the systematic hierarchy $\text{ETH} > \text{CNRM} > \text{CMCC}$ detected in the previous analysis. The dominant mode (PC1,
450 accounting for 74.2% of variance) represents a common climatological signal, with consistent loading patterns across spatial categories, suggesting that this fundamental signal is robust regardless of surface characteristics. This likely reflects their shared response to large-scale synoptic forcing from ERA-Interim boundary conditions, such as the passage of extratropical cyclones and frontal systems that dominate extreme wind generation across central Europe.

455 The disparities shown by PC2 and PC3 demonstrate how the models vary in their extreme wind simulations. PC2 identifies the main axis of disagreement among models, showing how CNRM and CMCC diverge in opposite directions while ETH stays close to the overall consensus. On the other hand, PC3 reveals a secondary divergence in which ETH separates from the CNRM-CMCC duo, as evidenced by persistent sign patterns in the loading distributions across all spatial categories (Fig. 4d). Because PC2 and PC3 are orthogonal, these divergences are statistically independent. This implies that the systematic differences
460 among models originate from various uncorrelated sources within the CPMs formulations, rather than being driven by a single dominant factor. These independent divergence patterns indicate that systematic differences originate from various uncorrelated aspects of model formulation. While we cannot definitively attribute these to specific parameterisations without sensitivity experiments, as illustrative examples, PC2 could potentially relate to sub-grid schemes and vertical momentum transport, while PC3 might involve boundary layer mixing and surface flux calculations; both crucial areas where CPMs settings can
465 differ significantly (Coppola et al., 2020; Fosser et al., 2024).

Furthermore, the consistent loadings across diverse surface types suggest that each CPM maintains its characteristic behaviour independent of local surface conditions, pointing to systematic (not random or specific) differences in model physics

as the primary drivers of inter-model uncertainty. These results establish that ensemble approaches are necessary because single-model representations of wind fields could provide complementary information.

470 5.3 Seasonal and spatial dependence in model agreement

A pronounced seasonal cycle in inter-model correlations of the monthly 99th percentile is observed (Fig. 5), providing clear evidence for the physical mechanisms controlling extreme wind generation and revealing fundamental differences in how CPMs represent diverse meteorological regimes. Winter months exhibit consistently higher correlations across most spatial categories, reflecting the dominance of extratropical cyclones and large-scale synoptic systems in generating extreme winds. 475 This synoptic-scale inter-model agreement is the result of the ERA-Interim boundary forcing, which provides the common meteorological boundaries that all three models represent similarly. In contrast, correlations are lower during summer months, especially July, reflecting a shift in the physical processes generating extreme winds. During summer, large-scale synoptic systems are weakened, and extreme wind events become increasingly driven by localised convective processes, thermal circulations, and mesoscale phenomena. In the summer, in the absence of strong, large-scale synoptic forcing, each CPM internally 480 develops its own mesoscale meteorology, including convective systems, land-sea breezes, and orographically influenced flows. The reduced summer correlations reflect both differences in convective physics representation and potential spatial displacement of similar events. This independent representation of sub-grid phenomena explains why model agreement decreases when these internally driven processes dominate the generation of extreme winds.

The spatial behaviour of seasonal correlation patterns reveals that CNRM-CMCC consistently exhibits the lowest correlations across all months and surface categories. At the same time, ETH-CMCC and ETH-CNRM show comparable and alternating correlation levels throughout the seasonal cycle (Fig. 5). This pattern is consistent with the PC2 results, where CNRM and CMCC systematically diverged in opposite directions while ETH remained near the consensus. Rather than surface-specific variations driving inter-model agreement, the almost consistent correlation patterns across diverse surface types reinforce that systematic model differences dominate over surface-atmosphere interaction effects. However, while spatial variations in 490 correlations are typically modest (~ 0.2 range), certain months, such as September and February, show substantially larger spatial dependency (~ 0.4 range), suggesting that specific seasonal conditions may amplify surface-related effects on inter-model agreement. Lower winter correlations in cold/tundra categories likely arise from model contrasting treatments of stable boundary-layer dynamics and/or orographic flow interactions under temperature inversions, and systematic model biases in high-elevation regions where terrain-induced processes amplify inter-model uncertainties (Correa-Sánchez et al., 2025a).

495 Future climate projections of extreme winds in convectively active regimes will likely carry greater uncertainty and would benefit significantly from multi-model ensemble approaches to capture the range of possible mesoscale responses to changing thermodynamic conditions. In addition, particular care should be taken regarding the possible changes in the seasonality of large-scale systems, such as extra-tropical cyclones.

5.4 Estimation of extreme wind return levels

500 The application of SMEV to estimate extreme wind return levels from CPMs is a novel contribution of this study. The systematic differences among individual models and the ensemble’s central tendency provide a benchmark for future CPM evaluation studies, while the spatial category approach enables targeted assessment of model performance across different surface types relevant to wind energy development.

The advantage of using SMEV instead of a traditional extreme value approach comes from the short duration of CPM
505 simulations. While traditional extreme value analysis depends on annual maximum values, which are just 10 values from the length of the simulations, or a still limited number of peak values above a very high threshold, SMEV uses the complete set of independent events. This provides abundant information for estimating extremes from the available short time series, thereby addressing this critical limitation of CPM simulations. This is particularly important for wind energy, where design standards require estimates for 50-year return periods that typically need to be based on limited simulation periods.

510 The assumption of Weibull tails for the distribution of the wind ordinary events, while well-supported for wind data (Harris and Cook, 2014), may not fully capture wind field complexity in heterogeneous terrain where multiple meteorological mechanisms contribute to extreme generation (Harris and Cook, 2014). However, this potential limitation is mitigated in the present study through the use of left-censoring, by which the potential impact of heterogeneous wind processes with different typical magnitudes is reduced.

515 Using ensemble approaches offers clear benefits over relying on single-model assessments for extreme wind evaluations. The individual models in Fig. 6a-c follow the same consistent hierarchy ($ETH > CNRM > CMCC$) across all spatial categories, as evidenced in previous results. This reinforces the idea that the way models are formulated leads to differences between them, rather than specific surface effects. In this context, CMCC shows greater internal variability, as indicated by its wider interquartile ranges, compared to ETH and CNRM. This suggests varying levels of dispersion within categories that warrant
520 further investigation.

The CPM ensemble analysed here differs in multiple aspects of model configuration: dynamical cores (COSMO vs AROME), PBL parameterisations (TKE vs CBR), vertical discretisation (50-60 levels), horizontal diffusion approaches (fourth-order Smagorinsky vs none vs SLHD), and horizontal grid spacing (2.2-3.0 km), as explained in Section 2. A key finding from our inter-model comparison is that CMCC and ETH show U_{50} differences that are similar in size to those between models using
525 different PBL schemes, even though they have the same PBL parameterisations. For instance, the U_{50} spread between CMCC or ETH (both using TKE-based PBL) is similar to the spread between CMCC and CNRM (TKE vs CBR). This pattern suggests that uncertainty in extreme wind estimates among models arises from the combined effects of various configuration choices, including, but not limited to, PBL parameterisation and fine-scale heterogeneity. These inter-model differences highlight the importance of using ensemble-based approaches for reliable extreme wind assessment in wind energy applications

530 Some spatial categories, like Ar —arid climates and R_1 —water surfaces, show less variability among the models. This probably reflects the more stable and limited range of meteorological conditions typical of these environments. Additionally, the ensemble approach in Fig. 6d-f gives more balanced estimates by averaging these systematic model differences and varying

internal sensitivities. This helps avoid the potential bias that could arise from using a single model. These patterns confirm that while the model hierarchy is consistent, individual models respond differently to sub-grid processes, highlighting the importance of ensemble approaches in capturing the full spectrum of physical process representations. While previous validation confirms that these models are physically sound (Correa-Sánchez et al., 2025b) and multi-model ensembles quantify structural uncertainty, observational validation remains essential to establish absolute accuracy and to detect potential shared systematic biases beyond the structural uncertainty quantified by multi-model ensembles.

The bootstrap confidence intervals in Fig. 6d-f provide strong statistical uncertainty bounds that are much narrower than the spatial variability seen within ensemble estimates across each category. This offers wind energy practitioners a more reliable range for design purposes, compared to the broader variability found in the full ensemble data.

The intra-category variability in U_{50} estimates shows the combined effects of meteorological sampling over the 10-year period and fine-scale spatial heterogeneity. During these temporal windows, different sites may have varying extreme meteorological conditions, with some experiencing more severe convective storms or synoptic extremes randomly, leading to different sets of ordinary events and consequently varying U_{50} estimates. Additionally, local surface characteristics (fine-scale roughness, microtopography) vary within categories; although this heterogeneity is less than the heterogeneity among categories, it is physically reasonable because categories inevitably capture dominant patterns, but not absolute uniformity. In this sense, both factors represent features of high-resolution climate data rather than methodological limitations.

5.5 Implications for wind energy applications

The consistent differences in how extreme winds are represented among CPMs and their spatial patterns offer insights for wind energy applications that go beyond traditional meteorological analysis. The variability of seasonal agreement among models provides a framework for assessing uncertainty and enables wind industry experts to adjust the confidence levels based on the season. For example, one could consider applying higher uncertainty margins during the summer months, when convective processes are more prevalent and inter-model correlations tend to decrease. In contrast, higher confidence can be placed in model consensus during winter periods, when synoptic processes dominate and correlations increase. The uniformity of model differences across all surface types indicates that ensemble approaches are essential, regardless of the project's location.

In terms of wind resource assessment, this systematic model hierarchy allows for risk-based approaches. For critical infrastructure design, conservative estimates using ETH can be employed, while ensemble estimates offer balanced projections for general applications. Given the consistent differences among models across all spatial categories, adopting ensemble approaches should be standard practice for evaluating extreme winds, irrespective of the type of surface or geographic location. Insights into the systematic behaviours of the models identified in this study can assist wind energy professionals in understanding the inherent uncertainties in CPM-based projections and in implementing appropriate ensemble strategies that align with their risk tolerance and application needs.

The intra-category variability of U_{50} has practical implications for multi-scale wind resource assessment and long-term planning. Our CPM ensemble approach provides robust climatological information at the regional and categorical scales (3 km horizontal grid spacing), as evidenced by the strong common signal (74.2% of variance), which is valuable for regional

resource mapping and preliminary site identification. However, when transitioning to site-specific turbine design, wind energy specialists require micrositing assessments at sub-grid scales (1-100 m) to capture fine-scale terrain features, local obstacles, and surface roughness variations not resolved by CPMs. This site-specific refinement follows established industry protocols and is necessary regardless of the regional climate data source employed.

Our U_{50} estimates over R_1 should be interpreted cautiously for offshore wind energy applications. First, considering our study area, the R_1 sample represents semi-enclosed Mediterranean/Adriatic waters rather than the unlimited-fetch open-ocean conditions characteristic of major offshore wind farms in other open-basin seas, such as the North Sea, where persistent westerly flow and unrestricted fetch typically generate systematically higher extreme wind speeds. Second, CORDEX-FPS CPMs employ coarse-resolution SST forcing (~ 80 km monthly means), creating scale mismatches with the fine-resolution (2–3 km) atmospheric grid (Caillaud et al., 2021), which may contribute to conservative marine wind estimates. These geographic and technical factors, combined with the physical process whereby water surfaces experience occasional intense synoptic events but lack the high-frequency convective events enriching terrestrial tail distributions, result in 18% lower scale parameter λ in R_1 compared to land categories R_3 – R_5 (Fig. E1), translating approximately linearly to lower U_{50} estimates.

5.6 Limitations of this study

A fundamental limitation of this study is the lack of observational datasets with sufficient spatial coverage, temporal extent, open accessibility, and consistent measurement heights to enable direct validation of CPM simulations. This observational gap limits our ability to validate the physical realism of the observed model hierarchy and to assess whether the U_{50} estimates align with real-world extremes across spatial categories. Additionally, the relatively short 10-year simulation period, while sufficient for obtaining reasonable estimates with SMEV, represents a limitation for characterising long-term climate variability. Nevertheless, since identical forcing isolates systematic model differences from climatic variations, 10 years is sufficient as well for inter-model comparison purposes. Also, an increased number of ensemble members is needed for a more robust assessment of model uncertainty.

Future validation efforts would greatly benefit from extensive observational networks that provide long-term wind measurements at hub height across diverse surface types. Specific validation campaigns in representative spatial categories could be helpful to assess absolute model performance and bias characteristics on the typical wind conditions, but would hardly provide actionable information about the extreme return levels needed for turbine design. To this end, priority should be given to the already available observational datasets, which should be made available to the community as open data. Until such validation becomes possible, the framework presented here serves as a methodological baseline for understanding CPM capabilities and inter-model uncertainties in extreme wind assessment.

The 10-year simulation period provides enough temporal coverage for SMEV application, as demonstrated by the strong inter-model consensus. However, this simulation period captures a finite sample of possible meteorological events, and some intra-category spread in U_{50} estimates reflects this meteorological sampling variability. Future studies with longer simulation periods (20-30 years) would further refine the separation between systematic spatial patterns and temporal sampling effects.

This is the first application of CORDEX-FPS CPMs for assessing extreme wind at turbine heights. We first investigated whether there are discrepancies among the models and how well the models agree in representing winds at 100 m. We demonstrated substantial inter-model differences in extreme wind representation, with systematic variations exceeding $15\text{--}20\text{ m s}^{-1}$ in annual maxima values in complex terrain despite identical geographic domains and atmospheric forcing. Furthermore, we provided a framework based on model inter-comparison for evaluating the CPM simulations of extreme wind for different spatial categories in the absence of long-term observational data. We then combined these simulations with the Simplified Metastatistical Extreme Value approach (SMEV) to assess extreme wind return levels at turbine heights.

We organise the analyses into spatial categories based on climate, terrain roughness, and topographic complexity to isolate the impact of these features and check if they had an effect on the divergences of CPM simulations. This approach provides category-specific estimates of extreme wind, along with category-specific confidence intervals, which can offer practical information for wind energy development and turbine design in various physiographic settings.

We applied the Principal Component Analysis (PCA) to understand the relative agreement among models, and show that 74.2% of variance represents a strong common climatological signal while the rest of the variance captures systematic and independent divergences among CPMs that remain uniform across all spatial categories. This consistent nature of the differences across all surface types indicates that model formulations, rather than the surface characteristics, are what mainly drive the inter-model uncertainty in extreme wind projections. This highlights the need for ensemble approaches, regardless of geographic location. Moreover, the significant seasonal variation in model agreement (with higher correlations in winter and lower agreement in summer) reflects the different effects of synoptic and convective processes in creating extreme winds, which also offers valuable insight for quantifying seasonal uncertainty in wind energy applications.

We find a consistent hierarchy in the simulated extreme wind speeds across all spatial categories, which is useful for designing risk strategies in wind resource assessment. Furthermore, the application of SMEV to short CPM simulations extends its applicability for obtaining extreme wind estimates from limited-duration datasets, establishing its viability for extreme wind assessment in the wind energy context.

Overall, our work establishes a methodological framework for evaluating inter-model agreement and uncertainties in CPMs' extreme winds, providing baseline knowledge that could inform future work in wind resource assessment, turbine site placement, and grid integration planning across various surface conditions and climates. Given that individual CPM members hold differences in representing extreme wind patterns, approaches based on an ensemble of models are fundamental.

Key limitations include the lack of observational datasets that provide enough spatial and time coverage and that are easily accessible for direct validation. Since multi-decadal simulations are useful for planning, turbine design, and investment strategies, there are other limitations regarding the relatively short 10-year simulation period that includes only three ensemble members. Future efforts should focus on expanding observational networks at hub heights, using longer simulations with more ensemble members, and carrying out targeted measurement campaigns across relevant spatial categories.

CPM projections are already available for near- and far-future scenarios. Combining these projections with the framework presented here could be a viable way to conduct future climate change impact assessments for the wind energy sector.

635 *Code availability.* The scripts for the formal analyses carried out in this work are available at Correa-Sánchez (2025). Codes to run the SMEV model and test its applicability are available from Marra et al. (2019) and Marra (2022).

Data availability. The CPM data used in the study cannot be shared by the authors, but they are available at the *Metagrid* user interface of the infrastructure software Earth System Grid Federation (ESGF) at <https://esgf-metagrid.cloud.dkrz.de/search> for the CORDEX-FPS CONV project.

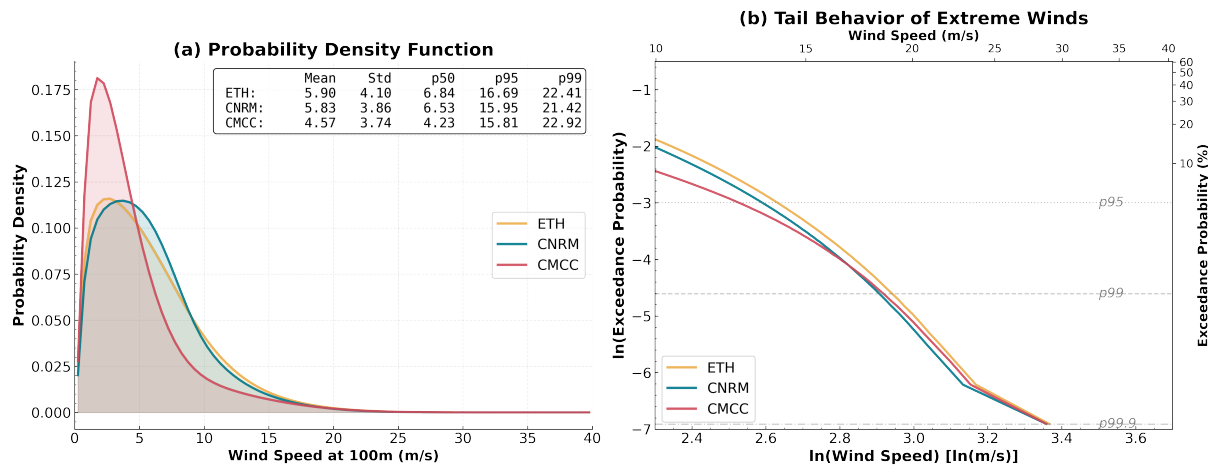


Figure A1. Statistical distributions of hourly wind speed at 100 m for the full 10-year time series (2000-2009). (a) Probability density function showing the overall distribution with summary statistics. (b) Exceedance probability in log-log scale, highlighting tail behaviour and focusing on the extreme wind range (10-40 m/s) where the grey horizontal lines indicate the 95th, 99th, and 99.9th percentiles.

Appendix B: Annual Maximum Statistics Across Spatial Categories

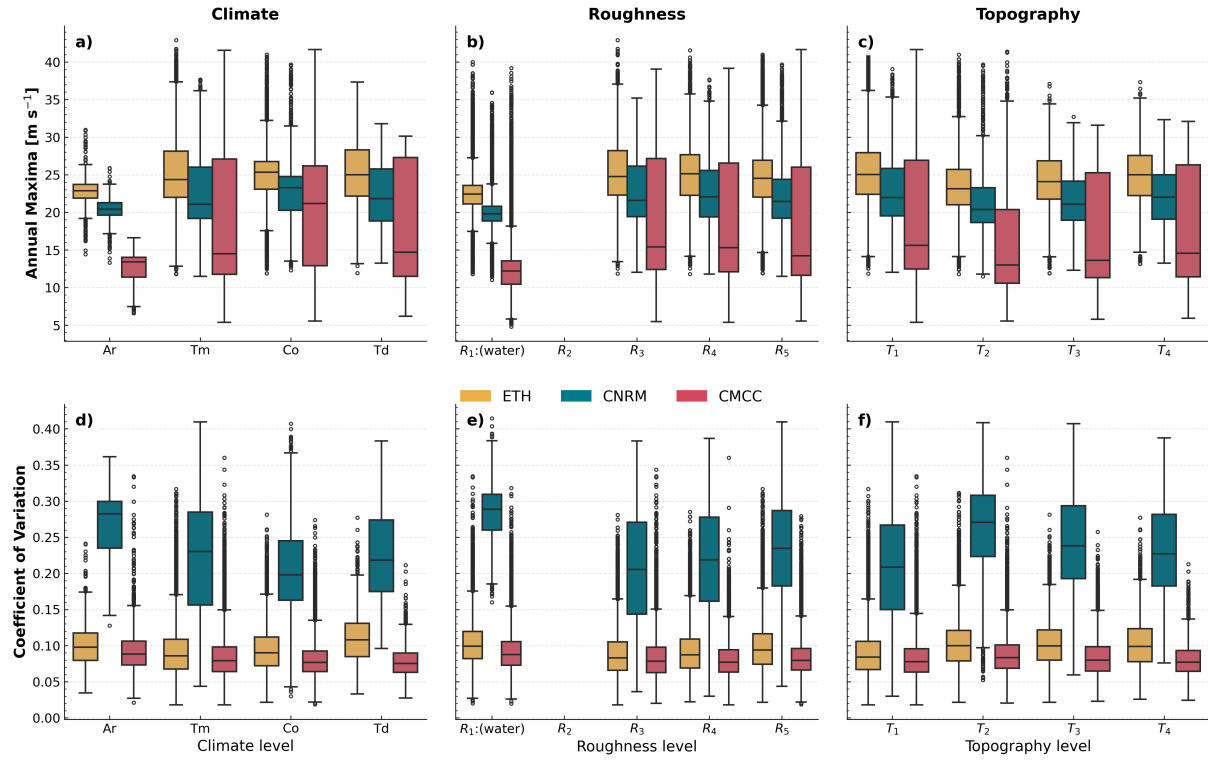


Figure B1. Distribution of annual maximum wind speeds and coefficient of variation (CV) across spatial categories. Boxplots show the complete distributions of (a-c) 10-year average annual maxima and (d-f) coefficient of variation for each CPM, disaggregated by climate zones, roughness classes, and topographic levels.

Appendix C: Spatial Patterns in Wind Speed Decorrelation Time

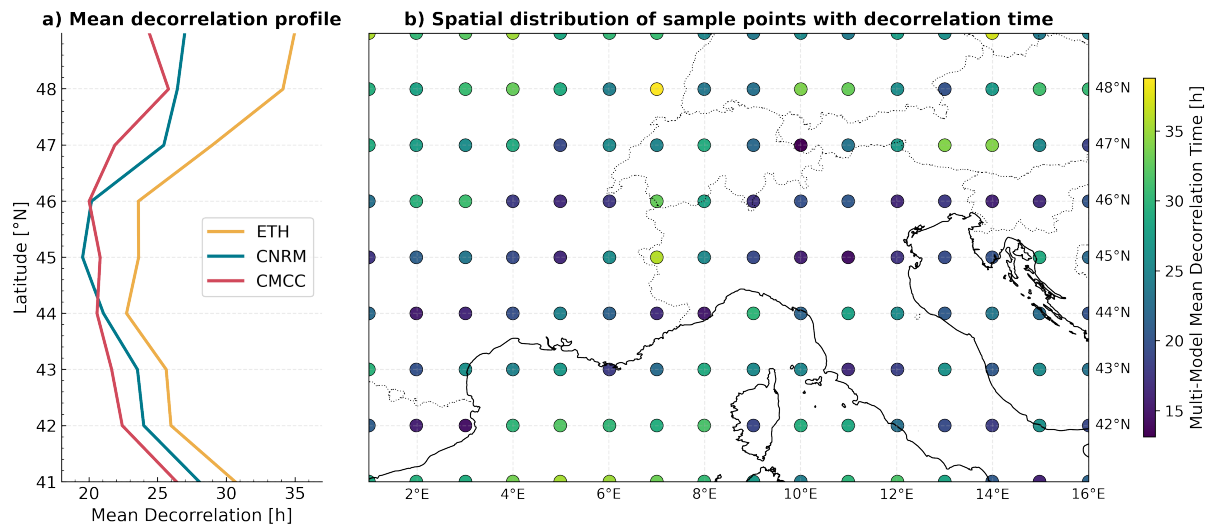


Figure C1. Temporal decorrelation time for 100 m wind speed in each CPM. (a) Latitudinal profile of zonally averaged decorrelation time for each CPM. (b) Multi-model mean decorrelation time averaged across ETH, CNRM, and CMCC. Circles indicate the $1^\circ \times 1^\circ$ grid points where decorrelation was computed for this illustrative analysis.

Appendix D: Sample Weibull Plots from Random Locations per CPM

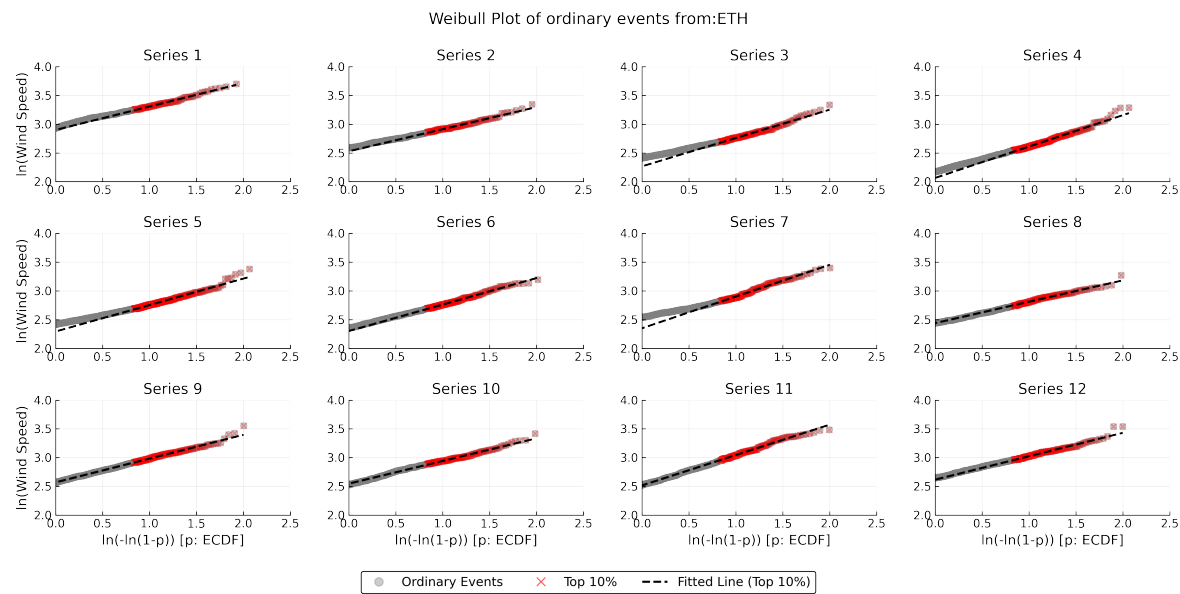


Figure D1. Weibull probability plots for ordinary events from the ETH model. Each panel shows ordinary events (gray), the top 10 % data portion (in red), and the linear fit (black dashed line) of that top portion in randomly selected locations.

Table D1. Spatial categories and location of sample series selected to illustrate the Weibull fit to ordinary events.

Series	Climate	Roughness	Topography	Longitude	Latitude
Series 1	Co	R ₅	T ₄	6.9836	45.2795
Series 2	Ar	R ₃	T ₁	0.5996	41.4725
Series 3	Ar	R ₃	T ₁	16.0656	41.1755
Series 4	Tm	R ₃	T ₂	15.6856	44.2805
Series 5	Co	R ₅	T ₃	12.0756	46.7375
Series 6	Co	R ₅	T ₃	6.1096	45.7115
Series 7	Co	R ₅	T ₃	5.4636	44.7665
Series 8	Tm	R ₃	T ₂	0.5236	43.0655
Series 9	Co	R ₅	T ₂	14.5076	46.5215
Series 10	Tm	R ₃	T ₂	11.6956	42.9845
Series 11	Td	R ₅	T ₄	13.4056	47.0615
Series 12	Co	R ₅	T ₂	15.0016	46.6025

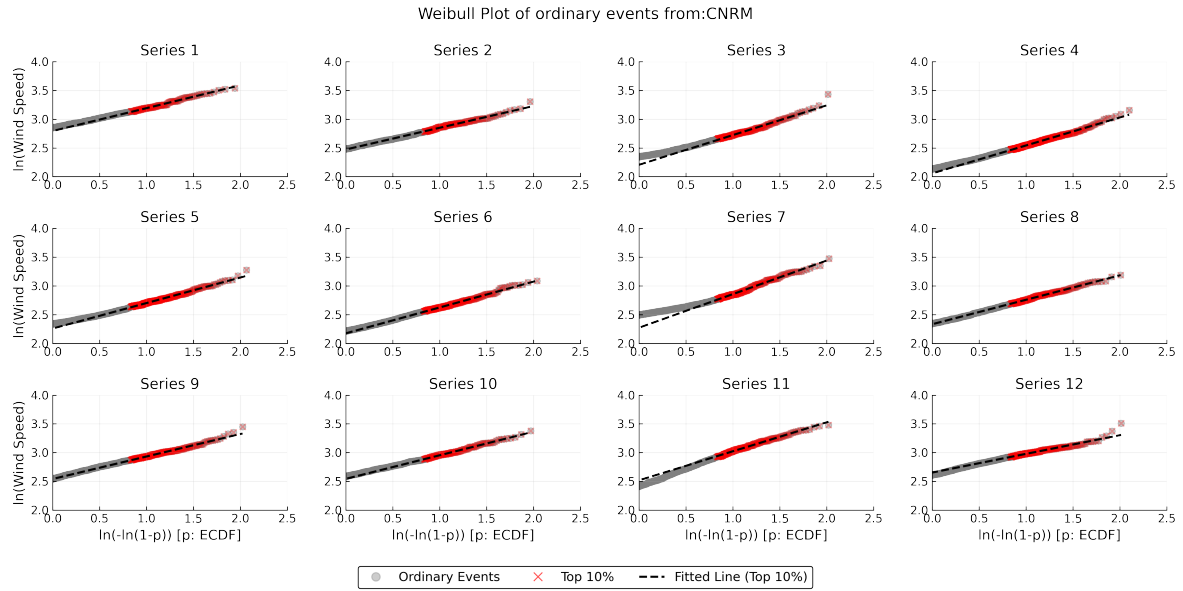


Figure D2. Weibull probability plots for ordinary events from the CNRM model. Each panel shows ordinary events (grey), the top 10 % data portion (in red), and the linear fit (black dashed line) of that top portion in randomly selected locations.

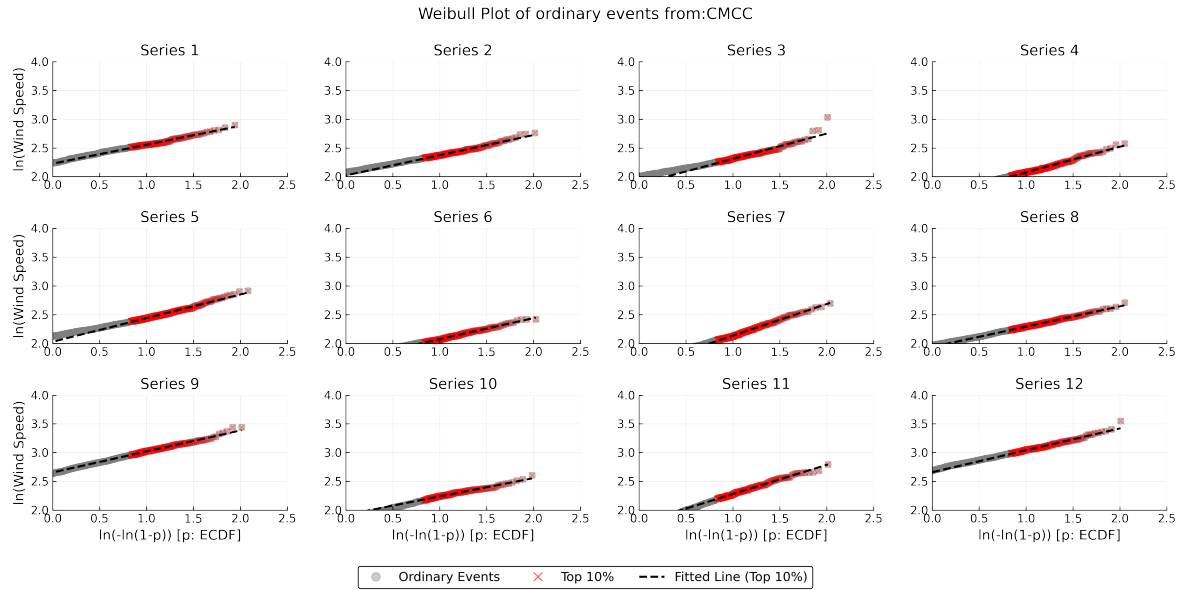


Figure D3. Weibull probability plots for ordinary events from the CMCC model. Each panel shows ordinary events (grey), the top 10 % data portion (in red), and the linear fit (black dashed line) of that top portion in randomly selected locations.

Appendix E: Weibull Parameter Distributions by Spatial Category

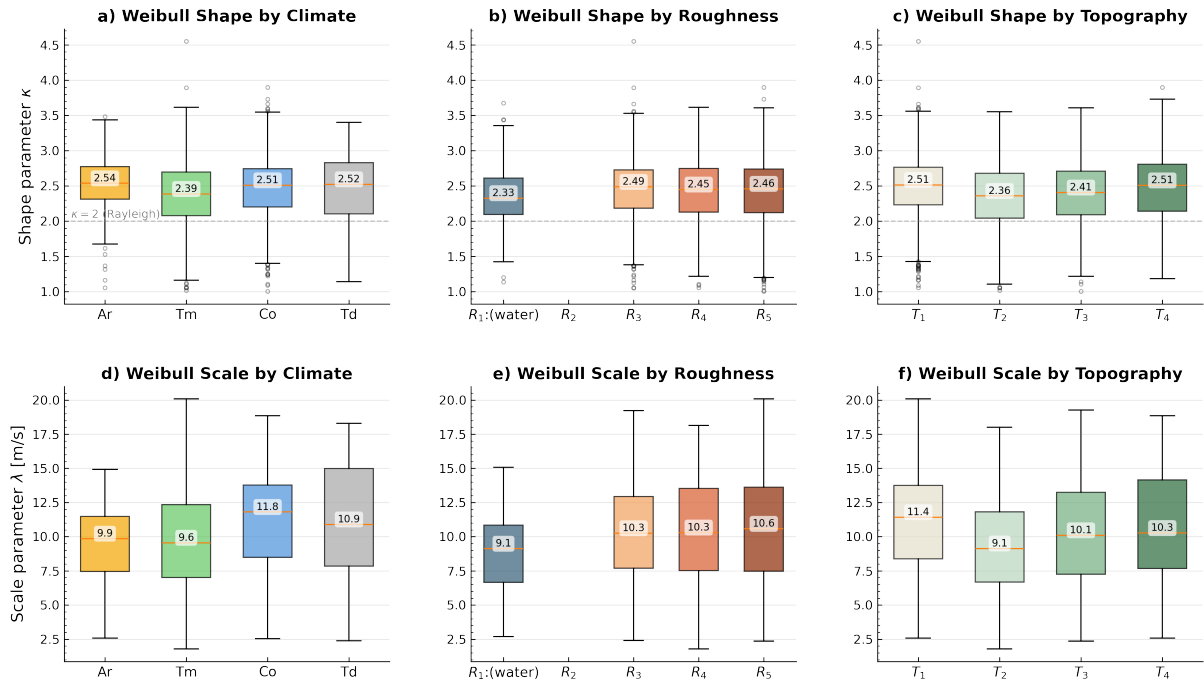


Figure E1. Distribution of Weibull parameters across disaggregated spatial categories. Panels (a-c) show shape parameter k . Panels (d-f) show scale parameter λ [m/s]. Each boxplot represents 100 locations per category. The dashed line in (a-c) marks $k=2$ (Rayleigh distribution); median values are annotated within boxes.

645 *Author contributions.* NCS, Data curation, Formal analysis, Software, Visualisation, Writing (original draft preparation); NCS, XGL, FM
Conceptualisation, Methodology, Investigation; XGL, FM Supervision; MB Funding acquisition; NCS, XGL, ED, MB, FM Writing (review
and editing).

Competing interests. The authors declare that they have no conflict of interest.

650 *Acknowledgements.* This study was partially supported by the CARIPARO Foundation through the Excellence Grant 2021 to the “Re-
silience” Project. FM was supported by the “The Geosciences for Sustainable Development” project (Budget Ministero dell’Università e
della Ricerca–Dipartimenti di Eccellenza 2023–2027 C93C23002690001). ED was supported by the RETURN Extended Partnership and
received funding from the European Union Next-Generation EU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2,
Investment 1.3 – D.D. 1243 2/8/2022, PE0000005). XL acknowledges support from Horizon Europe DTWO project (101146689). We thank
Marianna Adinolfi for providing the roughness layer used in the COSMO model for this study.

- Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., and Schmidt, G. A.: ESD reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing, *Earth System Dynamics*, 10, 91–105, 2019.
- Adinolfi, M., Raffa, M., Reder, A., and Mercogliano, P.: Evaluation and Expected Changes of Summer Precipitation at Convection Permitting Scale with COSMO-CLM over Alpine Space, *Atmosphere*, 12, 54, <https://doi.org/10.3390/ATMOS12010054>, 2020.
- Archer, C. L. and Jacobson, M. Z.: Evaluation of global wind power, *Journal of Geophysical Research: Atmospheres*, 110, 2005.
- Avissar, R. and Schmidt, T.: An Evaluation of the Scale at which Ground-Surface Heat Flux Patchiness Affects the Convective Boundary Layer Using Large-Eddy Simulations, *Journal of the Atmospheric Sciences*, 55, 2666 – 2689, [https://doi.org/10.1175/1520-0469\(1998\)055<2666:AEOTSA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1998)055<2666:AEOTSA>2.0.CO;2), 1998.
- Beck, H. E., McVicar, T. R., Vergopolan, N., Berg, A., Lutsko, N. J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, A. I. J. M., and Miralles, D. G.: High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections, *Scientific Data*, 10, 724, <https://doi.org/10.1038/s41597-023-02549-6>, data Descriptor, 2023.
- Bellucci, A., Haarsma, R., Gualdi, S., Athanasiadis, P., Caian, M., Cassou, C., Fernandez, E., Germe, A., Jungclaus, J., Kröger, J., et al.: An assessment of a multi-model ensemble of decadal climate predictions, *Climate Dynamics*, 44, 2787–2806, 2015.
- Bodini, N., Optis, M., Redfern, S., Rosencrans, D., Rybchuk, A., Lundquist, J. K., Pronk, V., Castagneri, S., Purkayastha, A., Draxl, C., Krishnamurthy, R., Young, E., Roberts, B., Rosenlieb, E., and Musial, W.: The 2023 National Offshore Wind data set (NOW-23), *Earth System Science Data*, 16, 1965–2006, <https://doi.org/10.5194/essd-16-1965-2024>, 2024.
- Bougeault, P. and Lacarrere, P.: Parameterization of orography-induced turbulence in a mesobeta-scale model, *Monthly weather review*, 117, 1872–1890, 1989.
- Caillaud, C., Somot, S., Alias, A., Bernard-Bouissières, I., Fumière, Q., Laurantin, O., Seity, Y., and Ducrocq, V.: Modelling Mediterranean heavy precipitation events at climate scale: An object-oriented evaluation of the CNRM-AROME convection-permitting regional climate model, *Climate Dynamics*, 56, 1717–1752, <https://doi.org/10.1007/S00382-020-05558-Y>, 2021.
- Chen, J., Yang, S., Li, H., Zhang, B., and Lv, J.: Research on geographical environment unit division based on the method of natural breaks (Jenks), *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 47–50, 2013.
- Chen, S., Poll, S., Hendricks Franssen, H.-J., Heinrichs, H., Vereecken, H., and Goergen, K.: Convection-permitting ICON-LAM simulations for renewable energy potential estimates over Southern Africa, *Journal of Geophysical Research: Atmospheres*, 129, e2023JD039 569, 2024.
- Cochran, W. G.: *Sampling techniques*, Johan Wiley & Sons Inc, 1977.
- Coles, S., Bawa, J., Trenner, L., and Dorazio, P.: *An introduction to statistical modeling of extreme values*, vol. 208, Springer, 2001.
- Cook, N. J.: Consolidation of analysis methods for sub-annual extreme wind speeds, *Meteorological Applications*, 21, 403–414, 2014.
- Cook, N. J., Harris, R. I., and Whiting, R.: Extreme wind speeds in mixed climates revisited, *Journal of Wind Engineering and Industrial Aerodynamics*, 91, 403–422, 2003.
- Coppola, E., Sobolowski, S., Pichelli, E., Raffaele, F., Ahrens, B., Anders, I., Ban, N., Bastin, S., Belda, M., Belusic, D., Caldas-Alvarez, A., Cardoso, R. M., Davolio, S., Dobler, A., Fernandez, J., Fita, L., Fumiere, Q., Giorgi, F., Goergen, K., et al.: A first-of-its-kind multi-model convection permitting ensemble for investigating convective phenomena over Europe and the Mediterranean, *Climate Dynamics*, 55, 3–34, <https://doi.org/10.1007/s00382-018-4521-8>, 2020.

Correa-Sánchez, N., Dallan, E., Marra, F., Fosser, G., and Borga, M.: Orographic control on bias and uncertainty in extreme sub-daily precipitation simulations from a convection-permitting ensemble, *Journal of Hydrology*, 659, 133–324, 2025a.

Correa-Sánchez, N., Larsén, X. G., Fosser, G., Dallan, E., Borga, M., and Marra, F.: Brief communication: Enhanced representation of the power spectra of wind speed in Convection-Permitting Models, *Wind Energy Science Discussions*, 2025, 1–14, 2025b.

Correa-Sánchez, N.: Extreme wind analysis scripts (Version 1.0.0) [Scripts], <https://doi.org/10.5281/zenodo.17459959>, 2025.

Courault, D., Drobinski, P., Brunet, Y., Lacarrere, P., and Talbot, C.: Impact of surface heterogeneity on a buoyancy-driven convective boundary layer in light winds, *Boundary-layer meteorology*, 124, 383–403, 2007.

Cuxart, J., Bougeault, P., and Redelsperger, J.-L.: A turbulence scheme allowing for mesoscale and large-eddy simulations, *Quarterly Journal of the Royal Meteorological Society*, 126, 1–30, 2000.

Dallan, E., Marra, F., Fosser, G., Marani, M., Formetta, G., Schär, C., and Borga, M.: How well does a convection-permitting regional climate model represent the reverse orographic effect of extreme hourly precipitation?, *Hydrology and Earth System Sciences*, 27, 1133–1149, <https://doi.org/10.5194/hess-27-1133-2023>, 2023.

Dallan, E., Borga, M., Fosser, G., Canale, A., Roghani, B., Marani, M., and Marra, F.: A method to assess and explain changes in sub-daily precipitation return levels from convection-permitting simulations, *Water Resources Research*, 60, e2023WR035 969, 2024.

Davis, N. N., Badger, J., Hahmann, A. N., Hansen, B. O., Mortensen, N. G., Kelly, M., Larsén, X. G., Olsen, B. T., Floors, R., Lizcano, G., Casso, P., Lacave, O., Bosch, A., Bauwens, I., Knight, O. J., van Loon, A. P., Fox, R., Parvanyan, T., Hansen, S. B. K., Heathfield, D., Onninen, M., and Drummond, R.: The Global Wind Atlas: A High-Resolution Dataset of Climatologies and Associated Web-Based Application, *Bulletin of the American Meteorological Society*, 104, E1507 – E1525, <https://doi.org/10.1175/BAMS-D-21-0075.1>, 2023.

Dee, D. P., Uppala, S., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the royal meteorological society*, 137, 553–597, 2011.

Demuzere, M., De Ridder, K., and van Lipzig, N. P. M.: Modeling the energy balance in Marseille: Sensitivity to roughness length parameterizations and thermal admittance, *Journal of Geophysical Research: Atmospheres*, 113, D16 120, <https://doi.org/10.1029/2007JD009113>, 2008.

Donat, M. G., Leckebusch, G. C., Pinto, J. G., and Ulbrich, U.: Examination of wind storms over Central Europe with respect to circulation weather types and NAO phases, *International Journal of Climatology*, 30, 1289–1300, <https://doi.org/10.1002/joc.1982>, 2010.

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

Dörenkämper, M., Olsen, B. T., Witha, B., Hahmann, A. N., Davis, N. N., Barcons, J., et al.: The making of the new European wind atlas—part 2: Production and evaluation, *Geoscientific Model Development*, 13, 5079–5102, 2020.

European Commission: In focus: Wind energy powering the clean transition, Tech. rep., European Commission, https://energy.ec.europa.eu/news/focus-wind-energy-powering-clean-transition-2025-03-17_en, 2025.

European Environment Agency: CORINE Land Cover 2018, version 2020_20u1, <https://land.copernicus.eu/en/products/corine-land-cover>, 2020.

Fosser, G., Gaetani, M., Kendon, E. J., Adinolfi, M., Ban, N., Belušić, D., Caillaud, C., Careto, J. A., Coppola, E., Demory, M.-E., et al.: Convection-permitting climate models offer more certain extreme rainfall projections, *NPJ Climate and atmospheric science*, 7, 51, 2024.

- Frisius, T., Cetin, I. I., Keup-Thiel, E., and Rechid, D.: Analysis of high-resolution convection-permitting regional climate simulations with respect to the impact of spatial resolution on simulated wind climate, in: *Journal of Physics: Conference Series*, vol. 2767, p. 092023, IOP Publishing, 2024.
- Harris, R. I. and Cook, N. J.: The parent wind speed distribution: Why Weibull?, *Journal of Wind Engineering and Industrial Aerodynamics*, 131, 72–87, <https://doi.org/10.1016/j.jweia.2014.05.005>, 2014.
- Hennessey Jr, J. P.: Some aspects of wind power statistics, *Journal of Applied Meteorology* (1962-1982), pp. 119–128, 1977.
- Jackson, P. S. and Hunt, J. C. R.: Turbulent wind flow over a low hill, *Quarterly Journal of the Royal Meteorological Society*, 101, 929–955, <https://doi.org/10.1002/qj.49710143015>, 1975.
- Jolliffe, I. T.: *Principal component analysis*, Springer, New York, 2nd edn., ISBN 978-0-387-95442-4, <https://doi.org/10.1007/b98835>, 2002.
- Kaimal, J. C. and Finnigan, J. J.: *Atmospheric boundary layer flows: their structure and measurement*, Oxford university press, 1994.
- Kang, S.-L. and Bryan, G. H.: A Large-Eddy Simulation Study of Moist Convection Initiation over Heterogeneous Surface Fluxes, *Monthly Weather Review*, 139, 2901 – 2917, <https://doi.org/10.1175/MWR-D-10-05037.1>, 2011.
- Kristensen, L., Kirkegaard, P., and Mann, J.: Sampling Statistics of Atmospheric Observations, *Wind Energy*, 5, 301–313, <https://doi.org/10.1002/we.75>, 2002.
- Larsén, X. G., Mann, J., Rathmann, O., and Jørgensen, H. E.: Uncertainties of the 50-year wind from short time series using generalized extreme value distribution and generalized Pareto distribution, *Wind Energy*, 18, 59–74, 2015.
- Larsén, X. G. and Mann, J.: The effects of disjunct sampling and averaging time on maximum mean wind speeds, *Journal of Wind Engineering and Industrial Aerodynamics*, 94, 581–602, <https://doi.org/10.1016/j.jweia.2006.01.020>, 2006.
- Leutwyler, D., Fuhrer, O., Lapillonne, X., Lüthi, D., and Schär, C.: Towards European-scale convection-resolving climate simulations with GPUs: A study with COSMO 4.19, *Geoscientific Model Development*, 9, 3393–3412, <https://doi.org/10.5194/GMD-9-3393-2016>, 2016.
- Leutwyler, D., Lüthi, D., Ban, N., Fuhrer, O., and Schär, C.: Evaluation of the convection-resolving climate modeling approach on continental scales, *Journal of Geophysical Research: Atmospheres*, 122, 5237–5258, <https://doi.org/10.1002/2016JD026013>, 2017.
- Ma, X., Li, Y., and Li, Z.: The projection of canadian wind energy potential in future scenarios using a convection-permitting regional climate model, *Energy Reports*, 8, 7176–7187, 2022.
- Ma, X., Li, Y., Huo, F., and Li, Z.: Convection-permitting WRF simulation of extreme winds in Canada: Present and Future scenarios, *Weather and Climate Extremes*, p. 100777, 2025.
- MacFerrin, M., Amante, C., Carignan, K., Love, M., and Lim, E.: The Earth Topography 2022 (ETOPO 2022) global DEM dataset, *Earth System Science Data*, 17, 1835–1849, <https://doi.org/10.5194/essd-17-1835-2025>, 2025.
- Manning, C., Kendon, E. J., Fowler, H. J., Roberts, N. M., Berthou, S., Suri, D., and Roberts, M. J.: Extreme windstorms and sting jets in convection-permitting climate simulations over Europe, *Climate Dynamics*, 58, 2387–2404, 2022.
- Marani, M. and Ignaccolo, M.: A metastatistical approach to rainfall extremes, *Advances in Water Resources*, 79, 121–126, 2015.
- Marra, F.: A Unified Framework for Extreme Sub-daily Precipitation Frequency Analyses based on Ordinary Events - data codes, <https://doi.org/10.5281/zenodo.3971557>, 2020.
- Marra, F.: A test for the hypothesis: block maxima are samples from a parent distribution with Weibull tail, <https://doi.org/10.5281/zenodo.7234707>, 2022.
- Marra, F., Zoccatelli, D., Armon, M., and Morin, E.: A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes, *Advances in Water Resources*, 127, 280–290, <https://doi.org/10.1016/j.advwatres.2019.04.002>, 2019.

- 765 Marra, F., Armon, M., and Morin, E.: Coastal and orographic effects on extreme precipitation revealed by weather radar observations, *Hydrology and Earth System Sciences*, 26, 1439–1458, <https://doi.org/10.5194/hess-26-1439-2022>, 2022.
- Marra, F., Amponsah, W., and Papalexiou, S. M.: Non-asymptotic Weibull tails explain the statistics of extreme daily precipitation, *Advances in Water Resources*, 173, 104388, <https://doi.org/10.1016/j.advwatres.2023.104388>, available online 30 January 2023, 2023.
- Mellor, G. L. and Yamada, T.: Development of a turbulence closure model for geophysical fluid problems, *Reviews of Geophysics*, 20,
770 851–875, 1982.
- Miniussi, A. and Marra, F.: Estimation of extreme daily precipitation return levels at-site and in ungauged locations using the simplified MEV approach, *Journal of Hydrology*, 603, 126946, 2021.
- Nabat, P., Somot, S., Cassou, C., Mallet, M., Michou, M., Bouniol, D., Decharme, B., Drugé, T., Roehrig, R., and Saint-Martin, D.: Modulation of radiative aerosols effects by atmospheric circulation over the Euro-Mediterranean region, *Atmospheric Chemistry and Physics*,
775 20, 8315–8349, <https://doi.org/10.5194/ACP-20-8315-2020>, 2020.
- National Renewable Energy Laboratory: Wind Energy Grows Up, <https://www.energy.gov/eere/wind/articles/wind-energy-grows>, 2023.
- Outten, S. and Sobolowski, S.: Extreme wind projections over Europe from the Euro-CORDEX regional climate models, *Weather and Climate Extremes*, 33, 100363, 2021.
- Palutikof, J. P., Brabson, B. B., Lister, D. H., and Adcock, S. T.: A review of methods to calculate extreme wind speeds, *Meteorological Applications*, 6, 119–132, 1999.
780
- Pergaud, J., Masson, V., Malardel, S., and Couvreur, F.: A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction, *Boundary-layer meteorology*, 132, 83–106, 2009.
- Petersen, R. L.: A wind tunnel evaluation of methods for estimating surface roughness length at industrial facilities, *Atmospheric Environment*, 31, 45–57, [https://doi.org/10.1016/S1352-2310\(96\)00154-9](https://doi.org/10.1016/S1352-2310(96)00154-9), 1997.
- 785 Pfahl, S.: Characterising the relationship between weather extremes in Europe and synoptic circulation features, *Natural Hazards and Earth System Sciences*, 14, 1461–1475, <https://doi.org/10.5194/nhess-14-1461-2014>, 2014.
- Prein, A. F.: Thunderstorm straight line winds intensify with climate change, *Nature Climate Change*, 13, 1353–1359, 2023.
- Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Keller, M., Tölle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N. P. M., and Leung, R.: A review on regional convection-permitting climate modeling: Demonstrations,
790 prospects, and challenges, *Reviews of Geophysics*, 53, 323–361, <https://doi.org/10.1002/2014RG000475>, 2015.
- Pryor, S. C., Barthelmie, R. J., Bukovsky, M. S., Leung, L. R., and Sakaguchi, K.: Climate change impacts on wind power generation, *Nature Reviews Earth & Environment*, 1, 627–643, 2020.
- Raschendorfer, M.: The new turbulence parameterization of LM, *COSMO newsletter*, 1, 89–97, 2001.
- Riley, S. J., DeGloria, S. D., and Elliot, R.: Index that quantifies topographic heterogeneity, *Intermountain Journal of Sciences*, 5, 23–27,
795 1999.
- Rockel, B., Will, A., and Hense, A.: The Regional Climate Model COSMO-CLM (CCLM), *Meteorologische Zeitschrift*, 17, 347–348, <https://doi.org/10.1127/0941-2948/2008/0309>, 2008.
- Ruiz, S. A. G., Barriga, J. E. C., and Martínez, J. A.: Assessment and validation of wind power potential at convection-permitting resolution for the Caribbean region of Colombia, *Energy*, 244, 123127, 2022.
- 800 Schmidt, O. T., Mengaldo, G., Balsamo, G., and Wedi, N. P.: Spectral Empirical Orthogonal Function Analysis of Weather and Climate Data, *Monthly Weather Review*, 147, 2979 – 2995, <https://doi.org/10.1175/MWR-D-18-0337.1>, 2019.

- Schär, C., Ban, N., Lenderink, G., Buonomo, E., Lüthi, D., and Coppola, E.: Kilometer-Scale Climate Models: Prospects and Challenges, *Surveys in Geophysics*, 42, 759–794, 2020.
- 805 Siena, M., Levizzani, V., and Marra, F.: A method to derive satellite-based extreme precipitation return levels in poorly gauged areas, *Journal of Hydrology*, 626, 130 295, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2023.130295>, 2023.
- Smith, G. L., Mlynczak, P. E., and Potter, G. L.: A technique using principal component analysis to compare seasonal cycles of Earth radiation from CERES and model computations, *Journal of Geophysical Research: Atmospheres*, 117, 2012.
- Smith, R. B.: The influence of mountains on the atmosphere, in: *Advances in geophysics*, vol. 21, pp. 87–230, Elsevier, 1979.
- 810 Soares, P. M. M., Cardoso, R. M., Miranda, P. M. A., de Medeiros, J., Belo-Pereira, M., and Espirito-Santo, F.: WRF high resolution dynamical downscaling of ERA-Interim for Portugal, *Climate Dynamics*, 39, 2497–2522, <https://doi.org/10.1007/s00382-012-1315-2>, 2012.
- Stull, R. B.: *An Introduction to Boundary Layer Meteorology*, vol. 13 of *Atmospheric and Oceanographic Sciences Library*, Springer, Dordrecht, ISBN 978-90-277-2769-5, <https://doi.org/10.1007/978-94-009-3027-8>, 1988.
- Thompson, S. K.: *Sampling*, John Wiley & Sons, Hoboken, NJ, 3 edn., ISBN 1118162943, 9781118162941, 2012.
- 815 Tibshirani, R. J. and Efron, B.: *An introduction to the bootstrap*, Monographs on statistics and applied probability, 57, 1–436, 1993.
- Troen, I. and Lundtang Petersen, E.: *European Wind Atlas*, Risø National Laboratory, Roskilde, Denmark, 1989.
- Tuller, S. E. and Brett, A. C.: The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis, *Journal of Climate and Applied Meteorology*, pp. 124–134, 1984.
- 820 Tye, M. R., Stephenson, D. B., Holland, G. J., and Katz, R. W.: A Weibull approach for improving climate model projections of tropical cyclone wind-speed distributions, *Journal of Climate*, 27, 6119–6133, 2014.
- Ulbrich, U., Leckebusch, G. C., and Pinto, J. G.: Extra-tropical cyclones in the present and future climate: a review, *Theoretical and applied climatology*, 96, 117–131, 2009.
- Webster, R. and Oliver, M. A.: *Geostatistics for Environmental Scientists*, John Wiley & Sons, ISBN 9780470028582, 9780470517277, <https://doi.org/10.1002/9780470517277>, 2007.
- 825 Yan, X. and Tang, Y.: An analysis of multi-model ensembles for seasonal climate predictions, *Quarterly Journal of the Royal Meteorological Society*, 139, 1179–1198, 2013.