Characterising wind power extremes over Kenya using an enhanced process-based reanalysis-driven model

Article

Published Version

Ang'u, C., Bloomfield, H. C., Hirons, L. C. ORCID: https://orcid.org/0000-0002-1189-7576, Woolnough, S. J. ORCID: https://orcid.org/0000-0003-0500-8514, Brayshaw, D. J. ORCID: https://orcid.org/0000-0002-3927-4362, Gitau, W., Masukwedza, G. I. T., Mutemi, J., Ochieng, W., Olago, D., Oludhe, C. and Wainwright, C. M. ORCID: https://orcid.org/0000-0002-7311-7846 (2026) Characterising wind power extremes over Kenya using an enhanced process-based reanalysis-driven model. Renewable Energy, 273. 126048. ISSN 1879-0682 doi: 10.1016/j.renene.2026.126048 Available at https://centaur.reading.ac.uk/130789/

To link to this article DOI: http://dx.doi.org/10.1016/j.renene.2026.126048

Publisher: Elsevier

Renewable Energy 273 (2026) 126048

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Characterising wind power extremes over Kenya using an enhanced process-based reanalysis-driven model

Cohen Ang'u [a,*], Hannah C. Bloomfield [b], Linda C. Hirons [c], Steven J. Woolnough [c], David J. Brayshaw [d], Wilson Gitau [a], Gibbon I.T. Masukwedza [c], Joseph Mutemi [a], Willis Ochieng [e], Daniel Olago [a], Christopher Oludhe [a], Caroline M. Wainwright [f]

[a] *Department of Earth and Climate Sciences, University of Nairobi, P.O. Box 30197-00100, Nairobi, Kenya*
[b] *School of Engineering, Newcastle University, Newcastle Upon Tyne, NE1 7RU, United Kingdom*
[c] *National Centre for Atmospheric Science, University of Reading, Reading, RG6 6ET, United Kingdom*
[d] *Department of Meteorology, University of Reading, Reading, RG6 6ET, United Kingdom*
[e] *Kenya Electricity Generating Company PLC, Kenya*
[f] *University of Leeds, Leeds, LS2 9JT, United Kingdom*

ARTICLE INFO

Keywords:
Wind energy
Reanalysis
Bias correction
Weibull quantile mapping
Persistence
Ramp rates

ABSTRACT

This study presents a robust framework for addressing systematic biases in the ERA5 wind speeds to model long-term, high-resolution wind energy and characterise wind power extremes in data-sparse regions. By integrating Weibull Quantile Mapping, hub-height extrapolation, and dynamic efficiency, the study models hourly output for three Kenyan wind farms: Lake Turkana Wind Power, Kipeto, and Ngong Hills. The model significantly reduced Mean Bias Error and Root Mean Square Error in the reanalysis while preserving temporal rank correlations. The reanalysis-driven model captures the fundamental variability of wind power generation. Persistence and ramp diagnostics using Threshold-Duration Frequency analysis reveal that: LTWP exhibits low variability, with high-output events (>80% Capacity Factor) sustained for durations exceeding 14 days, contrasting with the frequent multi-day droughts and pronounced ramping typical of mid-latitude wind turbine fleets. Kipeto and Ngong Hills sites exhibit strong diurnal cycling, necessitating short-term storage rather than seasonal balancing. While LTWP frequently undergoes large shifts (>60% Δ Capacity Factor) over diurnal cycles, extreme volatility at shorter timescales (3-h) is heavily damped. This framework demonstrates a transferable process for realistic wind power modelling in data-sparse environments, supporting regional energy planning and integration of renewables into developing power systems.

1. Introduction

Wind energy plays an important role in the global transition towards low-carbon power systems, offering a mature, scalable, and increasingly cost-effective alternative to fossil fuels. The technology's rapid evolution and improved turbine design have significantly enhanced the performance and reliability of wind energy systems [1]. In addition, declining turbine costs [2], and advances in wind forecasting have positioned wind energy as a cornerstone of modern renewable energy portfolios. Across Africa, South Africa, Morocco, and Egypt have established themselves as continental leaders, successfully deploying vast utility-scale wind farms. In contrast, most countries on the continent still lack high-quality wind observations and long-term in situ datasets required for investment-grade assessment [3].

Kenya presents a unique case study for wind energy integration because of its considerable wind resource potential, particularly in regions such as Marsabit [4]. In response, the country has incorporated significant wind power generation capacity into its electricity grid, including the 310 MW Lake Turkana Wind Power (LTWP), the 102 MW Kipeto wind farm, and the 25.5 MW Ngong Hills wind farm, highlighting both the promise and challenges of large-scale wind deployment [5]. In 2024, wind power contributed more than 12% of Kenya's national electricity mix [6], up from 1% in 2015, underscoring its growing importance in the country's renewable energy strategy.

One persistent challenge in wind-energy modelling, particularly in developing regions, is the scarcity of continuous, site-specific wind

---

* Corresponding author. P.O. Box 30197-00100, Nairobi, Kenya.
E-mail address: cangu@uonbi.ac.ke (C. Ang'u).

https://doi.org/10.1016/j.renene.2026.126048
Received 3 February 2026; Received in revised form 14 May 2026; Accepted 7 June 2026

measurements. Ground-based anemometer observations, though highly accurate, are typically short-term, incomplete, or geographically sparse because of logistical and maintenance constraints [7]. In contrast, global reanalysis datasets such as ERA5 [8], MERRA-2 [9], and CFSR [10] provide long-term, spatially and temporally consistent atmospheric records at hourly resolution. Reanalysis data are produced via data assimilation, relying on both observations and model-based forecasts to estimate atmospheric conditions [11]. Reanalysis-driven wind-power modelling frameworks have been successfully implemented for Europe [12], in Great Britain [13] and other regions, combining multi-decadal meteorological data with operational generation records [14], and bias correction [15]. However, such reproducible frameworks remain limited in African contexts, where observational coverage is sparse, hence the increasing need for context-relevant evidence to shape a clean energy future [16].

To overcome observational data gaps in Africa, reanalysis-based approaches have been applied. For instance Ref. [17], validated ERA5-derived wind generation against operational data from Ethiopian wind farms, while [18] demonstrated the potential of satellite data in wind resource assessment over Sudan. Although such studies confirm that reanalysis and satellite-based datasets can approximate regional wind power potential, they often omit bias-correction or operational adjustment layers. This is despite the fact that East Africa's atmospheric circulation is strongly influenced by regional climatic variability and complex rainfall dynamics [19]. Recent studies have shown that the Turkana low-level jet plays an important role in controlling regional wind regimes and associated rainfall characteristics across East Africa [20]. Furthermore, observational and reanalysis analyses have demonstrated that the Turkana Jet exhibits a pronounced diurnal cycle, with vital implications for regional wind variability and energy applications [21]. These processes underscore the importance of developing locally calibrated frameworks for wind resource assessment and modelling.

Reanalysis products have become indispensable tools for wind-resource assessment and wind power modelling applications [22]. In addition, comparative evaluations of global reanalysis datasets have highlighted their importance in representing near-surface wind conditions [23]. However, these products frequently exhibit systematic biases arising from coarse spatial resolution and simplified representations of land-surface roughness [24]. In regions influenced by complex mesoscale circulations, such as the Lake Turkana Low Level Jet, there can be further limitations in the ability of the reanalysis products to accurately represent local wind dynamics [25]. This limitation is compounded by the lack of observation input. In complex terrains such as the East African Rift and the Turkana Basin, these shortcomings can translate into substantial uncertainties in modelled energy yields.

To address these discrepancies, statistical bias-correction and downscaling approaches have been widely applied in climate and renewable-energy studies [26]. Trend-preserving bias-correction frameworks have also been developed to improve the consistency of corrected datasets while retaining long-term climate signals [27]. Among the available techniques, Weibull Quantile Mapping (WQM) has demonstrated strong performance for wind-energy applications and wind-power estimation [28]. Its effectiveness largely stems from the fact that wind speed distributions are commonly characterised using the Weibull probability distribution [29]. The Weibull framework characterises wind regimes using shape (k) and scale (c) parameters, which are closely linked to wind variability and energy potential [30]. WQM aligns the cumulative distributions of reanalysis and observed data while preserving their temporal rank structure, thereby correcting biases across the full distribution rather than only the mean or variance [31]. Recent studies have refined this approach using local calibration windows and fallback strategies to improve robustness under sparse data conditions [32].

Once wind speeds are bias-corrected, the next challenge in wind power modelling is wind-to-power conversion, typically represented by turbine power curves that define the nonlinear relationship between wind speed and electrical output. These curves are shaped by turbine-specific design parameters, including cut-in, rated, and cut-out speeds [33]. When aggregated across a wind farm, they form an idealised farm-level power curve that represents potential output under perfect operating conditions. However, real-world production is consistently lower due to operational inefficiencies such as wake effects, yaw misalignment, and curtailment. To capture these dynamics, efficiency factors are applied to improve realism by reducing high-wind overestimation and aligning modelled outputs more closely with observed generation.

Beyond the estimation of aggregate energy yield, the successful integration of wind power into national grids requires a rigorous characterisation of temporal variability, particularly the persistence of extreme states. Grid operators must quantify the frequency and duration of 'wind droughts' - prolonged periods of low generation that necessitate significant backup capacity or storage, as well as the stability of high-output events. Characterising these long-duration persistence phenomena presents a statistical challenge for standard observational campaigns. In-situ mast measurements, typically spanning only one to three years, are often too short to capture the full spectrum of inter-annual variability or to robustly estimate the return periods of rare, multi-day weather regimes [15]. Consequently, multi-decadal reanalysis products have become the standard tool for testing reliability because they provide the long-term temporal scope necessary to populate the 'tails' of the distribution.

Building upon these insights, this study develops an integrated process-based modelling framework that combines WQM, hub-height scaling, farm-level power curve modelling, and dynamic efficiency (dynEff) adjustments to characterise wind power extremes. The framework is validated based on in-situ observations from the LTWP, Ngong Hills and Kipeto wind farms in Kenya, which are characteristic of both high and moderate wind speed climatology. The procedure systematically corrects biases in reanalysis wind data and vertical wind shear, and incorporates a dynEff factor to produce robust estimates of power, capacity factors (CF), and energy-yield estimates for planning. Unlike previous studies that focus primarily on mean wind speeds for resource assessment, this work specifically targets the 'tails' of the distribution. The validated framework is used to characterise the persistence of extremes - both the stability of high-output events and the severity of wind droughts, thus providing Kenyan energy stakeholders with the investment-grade data required to optimise storage solutions, plan reserve margins, and manage the realities of a high-renewables grid.

2. Materials and methods

The datasets, calibration, validation, and analytical steps presented in this section underlie the entire integrated pipeline that bridges the gap between reanalysis-based modelling and site-specific wind energy estimation for characterisation of wind power extremes. The LTWP project in Marsabit County, and the Ngong Hills and Kipeto wind farms near Nairobi were chosen, as they constitute two different wind speed climatologies. Strong, persistent easterly flows associated with the Turkana Low Level Jet (Kinuthia, 1992) dominate at the LTWP, while moderate, topographically influenced winds dominate at Ngong Hills and Kipeto.

To evaluate the engineering and grid-planning value of integrating geographically dispersed wind resources, we quantified the spatial complementarity of the three sites (LTWP, Kipeto, and Ngong). We analysed the hourly CFs of each individual site. The Pearson correlation tests reveal that while the geographically proximate southern sites (Kipeto and Ngong) share a strong temporal alignment ($r = 0.778$), the northern site (LTWP) exhibits only moderate correlation with them ($r = 0.367$ and 0.418, respectively). LTWP is the most stable site (coefficient of variation (COV) $= 0.434$), while Kipeto and Ngong are highly volatile (COV $= 0.646$ and 0.886, respectively). LTWP drops below a 5% CF (a severe wind drought) only 3.27% of the time, Kipeto 14.18% and

Ngong 21.03%. These present very good model transferability cases. The contrasting environments are also suitable for comparing the performance metrics from models and evaluating how the model performs under various climatological and terrain conditions.

2.1. *Wind farm data*

The primary meteorological input is the fifth-generation reanalysis (ERA5) dataset, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service. ERA5 provides globally consistent meteorological data at hourly temporal resolution and $0.25^\circ \times 0.25^\circ$ spatial resolution. Ground-based wind speed and wind energy generation data were obtained from the specific wind farms and used for calibration and validation. Half-hourly in-situ anemometer records were available at both 40 m and 100 m above ground level for the LTWP site, and at 50 m and 20 m for both Kipeto and Ngong Hills wind farms. These in-situ observations were used (1) to bias-correct ERA5 wind speeds through WQM and (2) to validate modelled outputs against actual wind power generation. Half-hourly wind energy records used for model validation are as summarised in Table 1.

Distinct training windows based on data availability were utilised in model calibration: 1 October 2019 to 30 March 2021 for LTWP, and 1 January 2023 to 30 April 2025 for both Kipeto and Ngong. Model validation was performed over the period 1 April 2021 to 30 December 2022 for LTWP; 1 January 2021 to 30 December 2022 for Ngong; and 1 April 2021 to 30 December 2022 for Kipeto. The power curves were empirically verified using the WQM-corrected wind speeds and the observed generation.

2.2. *Modelling framework*

This study develops a seven-step process-based modelling framework to translate ERA5 reanalysis winds into site-representative wind speed, power output, CF, and energy series for wind farms. The framework integrates the statistical robustness of ERA5 with site-specific calibration from local observations. The modelling framework is illustrated schematically in Fig. 1 and includes: (1) data extraction and spatial representativity, (2A) WQM calibration of LTWP ERA5 winds at 100m, (2B) vertical scaling of ERA5 winds to 50m observations level for Kipeto and Ngong, (3A) Scaling the WQM-corrected LTWP winds to hub-height (50m), (3B) WQM calibration of Kipeto and Ngong winds at 50m, (4A) Scaling the WQM-corrected Kipeto/Ngong winds to hub-height, (5) wind-to-power conversion, (6) dynamic efficiency (dynEff) adjustment, and (7) model validation. For LTWP, calibration was performed at 100 m before scaling to the 50 m hub height, whereas for Kipeto and Ngong, ERA5 wind speeds were first adjusted from 100 m to 50 m, calibrated at 50 m, and then scaled to hub height. Each stage is unpacked in the following subsections.

2.2.1. *Wind-shear adjustment*

To account for vertical wind-speed variation between the ERA5 reference level (100m) and turbine hub height, wind-shear adjustment was performed using a time-varying empirical shear exponent. The shear exponent, $\alpha(t)$, was estimated from concurrent anemometer observations at two levels using:

$$\alpha(t)=\frac{\ln(U_{l2}(t)/U_{l1}(t))}{\ln(l2/l1)} \tag{1}$$

where $U_{l2}(t)$ and $U_{l1}(t)$ are the observed wind speeds at level 2 and level 1, respectively, at time t. Wind speed observations at LTWP were collected at 100 m and 40 m above ground level (AGL), whereas measurements at the Kipeto and Ngong wind farms were obtained at 50 m and 20 m AGL.

The resulting $\alpha(t)$ series was then used to vertically adjust the bias-corrected ERA5 wind speed to the target hub height z_h according to:

$$U(z_h,t)=U_h(t)\left(\frac{z_h}{100}\right)^{\alpha(t)} \tag{2}$$

where $U_h(t)$ is the WQM-corrected ERA5 wind speed, and z_h is the target height.

For LTWP, WQM was first applied at 100 m before scaling to hub height because both ERA5 and observed wind speeds were available at that height. For Kipeto and Ngong, ERA5 100 m wind speeds were first scaled to the 50 m observation level, WQM applied at 50 m, before scaling to hub height. Thus, in all cases, bias correction was performed at the observational reference height before final hub-height adjustment.

2.2.2. *Calibration via Weibull Quantile Mapping (WQM)*

The core bias-correction step uses Weibull Quantile Mapping (WQM) to align ERA5 wind speed distributions with observations while preserving temporal rank order. For each calibration stratum, tested options included monthly, hour-of-day, and the combination of both month and hour. The procedure fits Weibull distributions to observed and ERA5 wind speeds using maximum likelihood estimation, computes cumulative distribution functions, and maps each ERA5 wind speed to observations:

$$V_{corr}=F_{obs}^{-1}(F_{ERA5}(V_{ERA5})) \tag{3}$$

where F_{ERA5} and F_{obs} are the cumulative distribution functions for ERA5 and observed data, respectively. Calibration was performed over a training period and applied to a distinct analysis window to ensure independence and avoid overfitting (see Table 1 for details).

2.2.3. *Wind-to-power conversion*

Bias-corrected wind speeds were converted to power using site-specific farm power curves. Rather than applying a generic turbine power relationship, the conversion was based on the power curve corresponding to the actual wind farm under study, thereby preserving differences in turbine configuration, rated capacity, and aggregate farm

Table 1

Details of the site characteristics for the three operational wind farms in Kenya used in this study.

Wind Farm	Coordinates	Elevation (m)	Installed Capacity (MW)	No. of Turbines	Turbine Model	Hub-Height (m)	Data Period Used		
							ERA5	Obs. wind speed (m/s)	Generation (KWh)
Lake Turkana (LTWP)	36.7°E – 36.9°E, 2.4°N - 2.8°N	450 - 470	310	365	Vestas V52-850	50	1979-2025	01-10-2019– 30-03-2021 (Height, 40m & 100m)	01-04-2021–30-12-2022
Ngong Hills	36.5°E–36.75°E, 1.50°S - 1.25°S	1950 - 2100	25.5	30	Gamesa G52-850	50	1979-2025	01-01-2023– 30-04-2025 (Height, 20m & 50m)	01-01-2021– 30-12-2022
Kipeto	36.6°E - 36.9°E, 1.9°S - 1.6ºS	1900 - 2100	102	60	GE Energy 1.7-103	80	1979-2025	01-01-2023– 30-04-2025 (Height, 20m & 50m)	01-04-2021– 30-12-2022

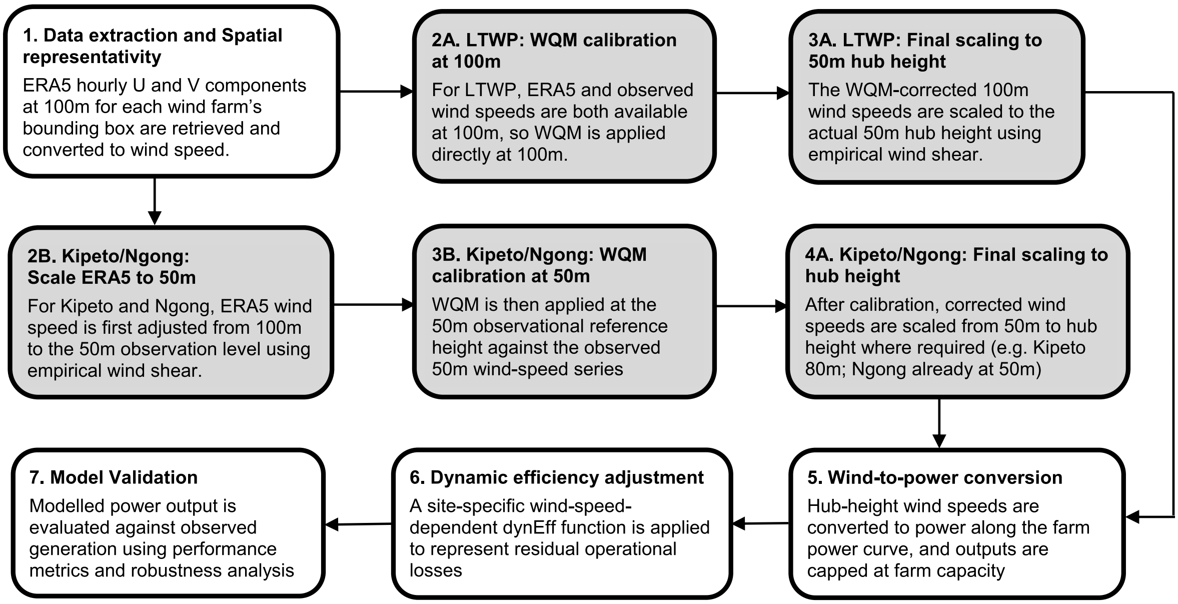

Fig. 1. Seven-step wind-power modelling workflow.

response. Turbine-specific power curves were obtained from manufacturer data at thewindpower.net database. For each site, the corrected wind-speed series $U(t)$, obtained after bias correction and wind shear adjustment were mapped to electrical power using the corresponding farm power curve:

$$P_{base}(t) = P_{curve}(U(t)) \tag{4}$$

where P_{curve} denotes the site-specific farm-level power curve and $P_{\text{base}}(t)$ is the resulting baseline modelled power. The power curve was implemented as a relationship between wind speed and power, with interpolation used between successive wind-speed values to obtain a continuous estimate of power. The resulting baseline power output was further constrained to the installed capacity of the wind farm, so that:

$$0 \le P_{base}(t) \le P_{rated} \tag{5}$$

where P_{rated} is the total installed capacity of the wind farm.

2.2.4. *Dynamic efficiency adjustment*

An efficiency factor was calibrated separately for each wind farm using observed generation data and the site-specific farm power curve. This was done using overlapping observed generation and modelled data over the period 01-01-2023 to 30-12-2023. The dynEff term was introduced after wind-speed correction and height adjustment, so that it captured residual performance effects in power conversion rather than compensating for bias in the wind-speed model itself. Observed generation, $P_{\text{obs}}(t)$, was aligned to the same hourly time step as the modelled series. A raw dynEff ratio was estimated as:

$$\eta_{\text{raw}}(t) = \frac{P_{\text{obs}}(t)}{P_{\text{base}}(t)} \tag{6}$$

This ratio represents the effective fraction of baseline power that was realised in operation. To avoid unstable values under very low-power conditions, calibration was limited to hours where baseline modelled power exceeded 5% of the farm rated capacity. The raw efficiency ratios were then grouped into wind-speed bins of width 0.5 m/s. For each bin, the median efficiency ratio was calculated. The final calibrated efficiency function, $\eta(U)$, yielded a site-specific wind-speed-dependent dynEff relationship.

The final modelled power was obtained by multiplying the baseline power-curve output by the calibrated dynEff function:

$$P_{\text{final}}(t) = \eta(U(t))\, P_{\text{base}}(t) \tag{7}$$

This calibration was applied separately for LTWP, Kipeto, and Ngong, allowing the dynEff term to reflect site-specific operational behaviour.

2.2.5. *Model validation*

Model performance was reported at daily time-scale using point metrics (bias, mean absolute error (MAE), root mean square error (RMSE), Pearson correlation, Kling–Gupta Efficiency (KGE), and Nash–Sutcliffe Efficiency (NSE)) and distributional diagnostics (quantile biases for/P50/P90, duration-curve comparisons, and kernel density overlays. Pearson correlation r for linear association; Bias to show systematic under- or over-production; MAE and RMSE for typical and large errors; The NSE asks whether the model beats a baseline that just uses the mean observation (NSE >0, model is better than using the observed mean, NSE = 0, the model is no better than using the observed mean every time, NSE <0, the model is worse than the mean baseline); KGE blends correlation, variability ratio, and mean ratio into one number. Event diagnostics include ramp statistics and count of extreme overestimation events. To avoid circularity and overlaps with the calibration periods, model validation was done over the period 01-05-2021 to 30-12-2022 (for LTWP), 01-04-2021 to 30-12-2022 (for Kipeto), and 01-01-2021 to 30-12-2022 (for Ngong).

Robustness checks included sensitivity experiments on WQM stratification (monthly, hourly, sectoral) and shear exponents (α), alongside comparisons between constant loss factors and dynEff parameterisations.

A reliability-boundary analysis for the model was also performed under controlled disturbance scenarios. This included the following:

- Missingness curve: the fraction of wind-speed timestamps removed, then time-interpolated
- Noise curve: Gaussian noise added at increasing fractions of wind-speed standard deviation
- Outliers curve: increasing fraction of timestamps hit with ±3σ wind-speed shocks

Missingness was simulated by randomly removing a prescribed fraction of hourly modelled wind-speed values and reconstructing the resulting gaps using time interpolation. Noise was simulated by adding zero-mean Gaussian perturbations with standard deviation defined as a fixed fraction of the standard deviation of the baseline wind-speed series. In each case, the perturbed wind-speed series was converted to power using the farm power curve and evaluated against observed generation.

2.3. Characterising persistence of wind power extremes

A Threshold-Duration Frequency analysis was used to quantify the temporal stability of wind generation. Instead of simple statistics (like mean or standard deviation), this method counts how frequent the power output remains in a specific state (Low or High) for a defined minimum duration. The hourly CF time series is evaluated against a grid of discrete thresholds to define events.

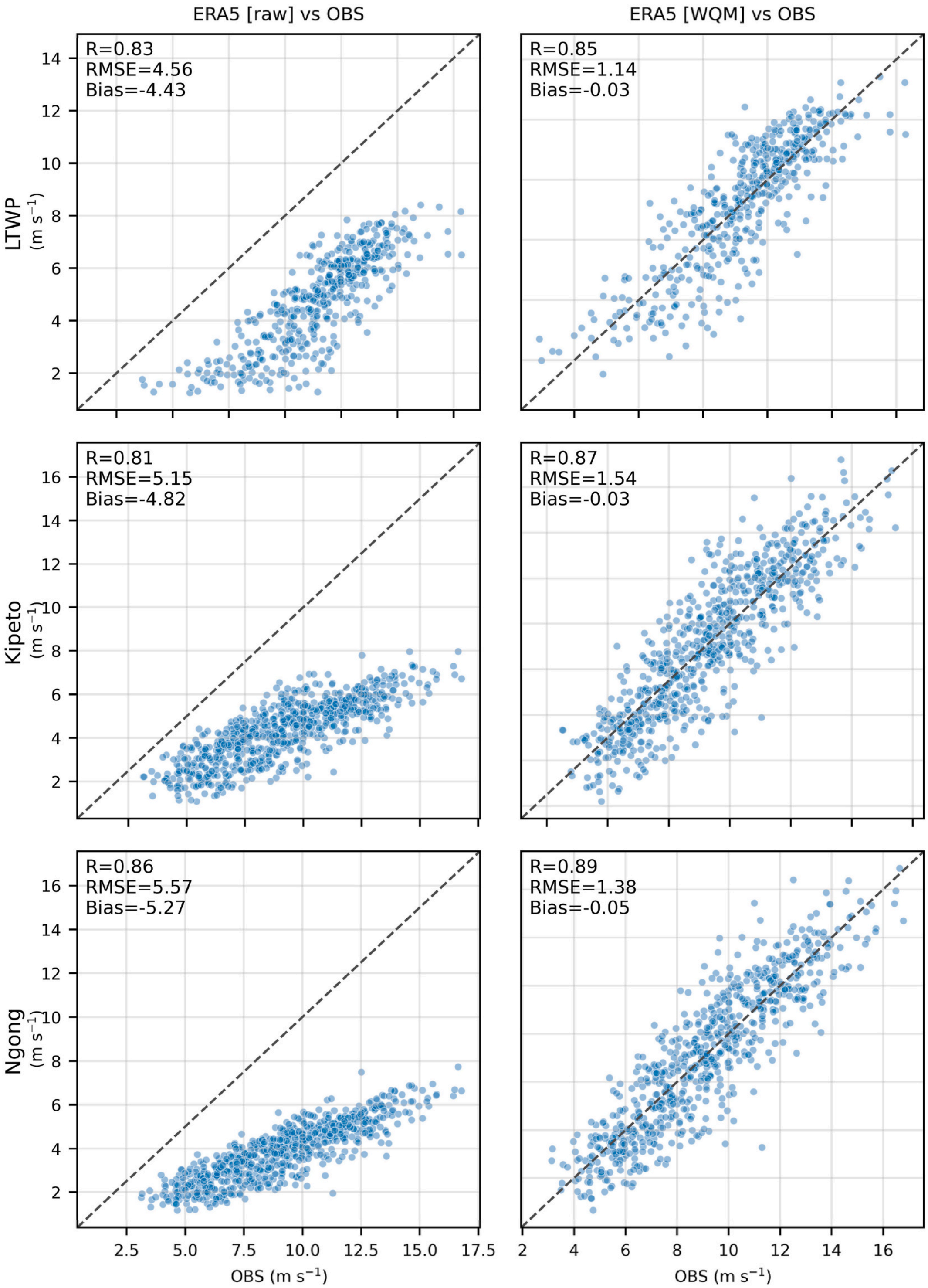

Fig. 2. Scatter plots comparing observed (OBS) and ERA5-derived daily mean wind speeds (m s^{-1}) at 100 m for LTWP, and 50m for Kipeto and Ngong sites before (left column) and after (right column) bias correction using the WQM-monthly method. Each point represents a daily mean value, with the 1:1 dashed line indicating perfect agreement.

Low Persistence (Wind Droughts): A low persistence event is defined as a continuous period where the hourly generation falls below or equals a specific threshold (e.g., CF ≤ 20%). This characterises the risk of supply deficits.

High Persistence (High-output events/reliability): A high persistence event is defined as a continuous period where the hourly generation stays above or is equal to a specific threshold (e.g., CF ≥ 80). This characterises the resource's ability to provide reliable power.

For every threshold level, we identify contiguous blocks of time (runs) where the condition is met. The duration of each block is calculated in hours and converted to days. To visualise the relationship between Intensity (CF %) and Duration (Days), a 2D matrix was constructed for both Low and High events.

3. Results

3.1. Bias correction of ERA5 wind speeds

WQM effectively corrected the higher-order distributional moments in the ERA5 wind field, substantially reducing systematic negative biases at the three wind farms (see Fig. 2). Although the raw ERA5 product resolved the seasonal cycle of wind speeds well, it exhibited a dampened probability density function (PDF) characterised by underestimated means and truncated right-tail extremes. WQM adjustment restored the variance and mean closer to observational values, evidenced by the tightening of scatter around the 1:1 axis (Fig. 2).

Validation metrics confirm a reduction in systematic and random errors (bias and RMSE) with no loss of temporal synchronisation, as correlations increased. The WQM-adjusted ERA5 series (Fig. 3, blue curve) closely follows observed variability (Fig. 3, orange curve) throughout the year, while the raw ERA5 data (grey curve) systematically underestimates wind speeds. Seasonal and site-specific wind speed patterns are well captured after correction, demonstrating improved agreement between reanalysis and observed datasets.

3.2. Conversion to power and sensitivity to hub height and dynEff

Individual wind turbine power curves were applied to convert hub-height winds into idealised plant output. Introducing a wind-speed-dependent (dynEff) parameterisation further reduced the mismatch between ideal generation and observations. The dynEff brings modelled values into close agreement with observations (Fig. 4). Ngong site reveals a distinct operational profile characterised by a strict performance ceiling (Fig. 4). Under idealised conditions (dynEff Off), the model allows generation to approach its theoretical maximum. When realistic dynEff limits are applied, the site's output abruptly caps at approximately 0.7 CF across both 50m and 100m hub heights. LTWP and Kipeto shows highly uniform performance. The model for Kipeto is completely saturated because the 100 m and 80 m model configurations are producing very similar capacity-factor distributions.

Sensitivity tests show that increasing hub height produces a substantial uplift in CF across the mid-range of the distribution, with diminishing gains near the highest CFs where turbines already operate close to rated power (Fig. 4). Across the three sites, the model captures most of the observed temporal variability but with clear site-specific behaviour (Fig. 5). The model generally reproduces the timing and magnitude of major generation fluctuations across the three farms.

The reliability-boundary analysis indicates that LTWP model performance is comparatively insensitive to random missingness, with normalized RMSE remaining nearly unchanged even as the proportion of missing data increases (Appendix C). Sensitivity to Gaussian noise is modest, with only a slight increase in error as noise intensity rises. In contrast, the model is much more sensitive to outliers, for which normalized RMSE increases rapidly.

3.3. Seasonal structure of energy distributions

The seasonal distribution of daily energy generation for LTWP is captured by monthly PDFs of modelled and observed wind energy generation at hub-height (Fig. 6). Results for Kipeto and Ngong are provided in Appendix A (Figures A1 and A2), respectively. At LTWP

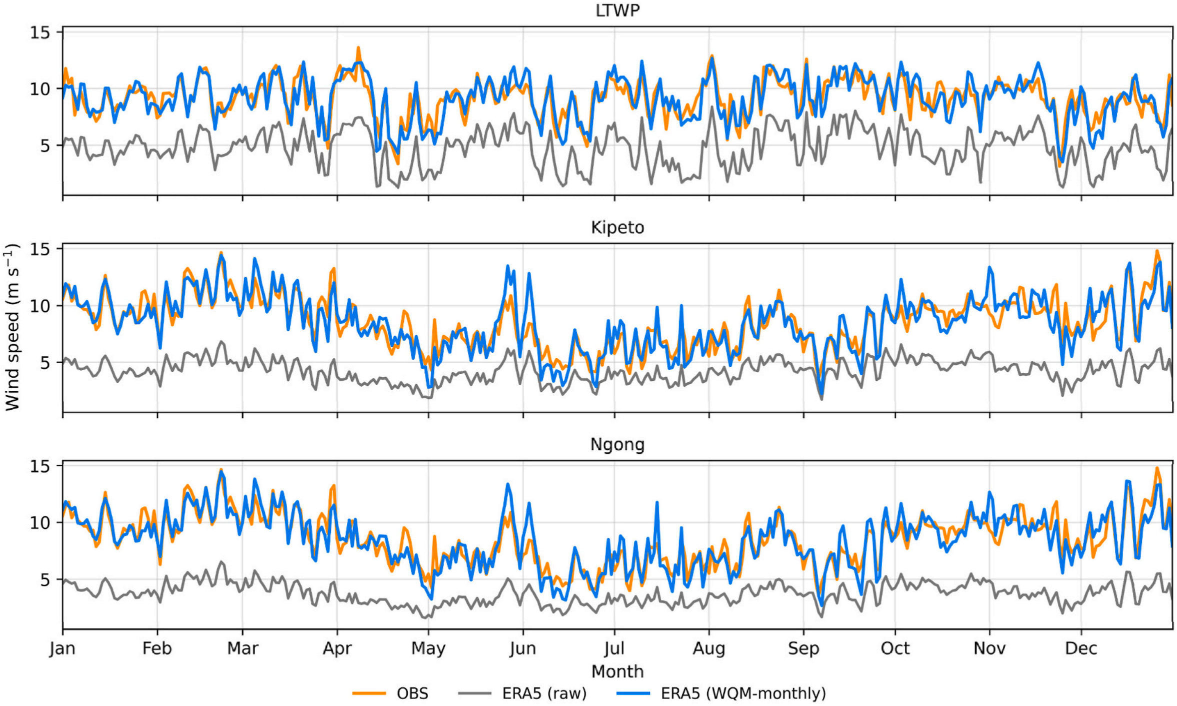

Fig. 3. Daily mean wind speed time series (m s^{-1}) for LTWP (100m), Kipeto (50m), and Ngong (50m) wind farms comparing observations (OBS; orange), raw ERA5 reanalysis data (grey) (100m for LTWP, 50m for Kipeto and Ngong), and bias-corrected ERA5 data (blue) using the WQM-monthly method. The data periods are shown in Table 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

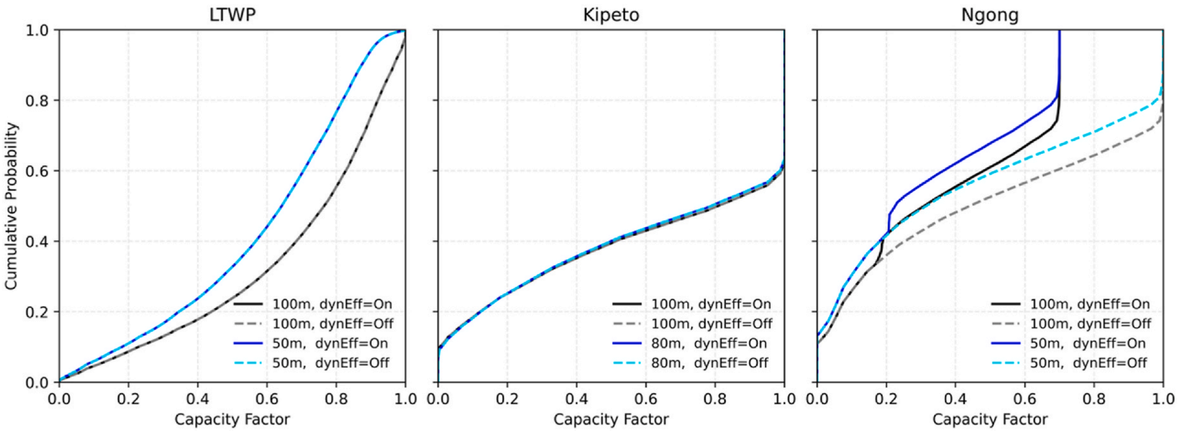

Fig. 4. Sensitivity analysis of modelled CF distributions to hub height and dynEff. Cumulative Distribution Functions (CDFs) are shown for LTWP (left), Kipeto (centre), and Ngong (right). The analysis compares four configurations per site: hub heights of 100 m (black/grey) versus 50 m (blue/cyan) for LTWP and Ngong, and 80m vs 100m hub heights for Kipeto. The plots also show dynEff corrections enabled (solid lines) versus disabled (dashed lines). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

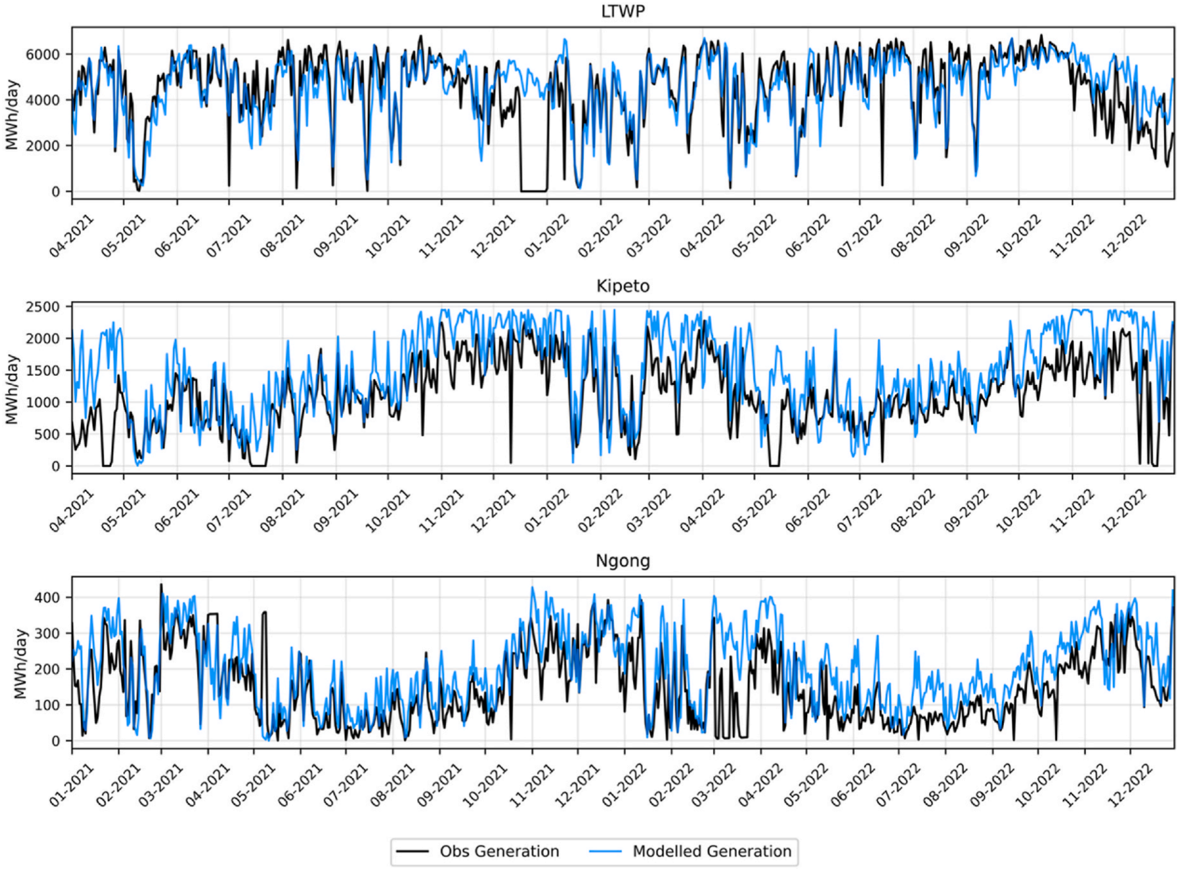

Fig. 5. Comparison of observed (black curve) and modelled (blue curve) daily wind-energy generation at LTWP, Kipeto, and Ngong over the model validation periods. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

(Fig. 6), the model successfully captures the broad seasonal shifts in wind power generation. It correctly identifies the high-generation months (June - October) and the transition to lower-generation periods (November - April). The strongest model performance is achieved in the months of February – May and September - October. Pearson correlations (r) are excellent ($r > 0.90$), and NSE scores are high (e.g., February and October, $NSE = 0.87$), indicating high predictive skill. During the peak Turkana Jet season JJA (June-July-August), the model correctly predicts high average output (note the x-axis peaks shifting right toward the 5000–6000 MWh/day band). While the model generally tracks the seasonal generation profile, significant divergences in December highlight the distinction between meteorological resource availability and actual grid dispatch. Observed output in December was notably erratic and lower than expected. Our consultations with KenGen revealed that August is a designated month for scheduled maintenance at LTWP, while the month of December reflects ‘economic curtailment’ due to low demand because of the holiday season. At Kipeto (Figure A1) and Ngong (Figure A2), the distributions reflect more intermittent wind

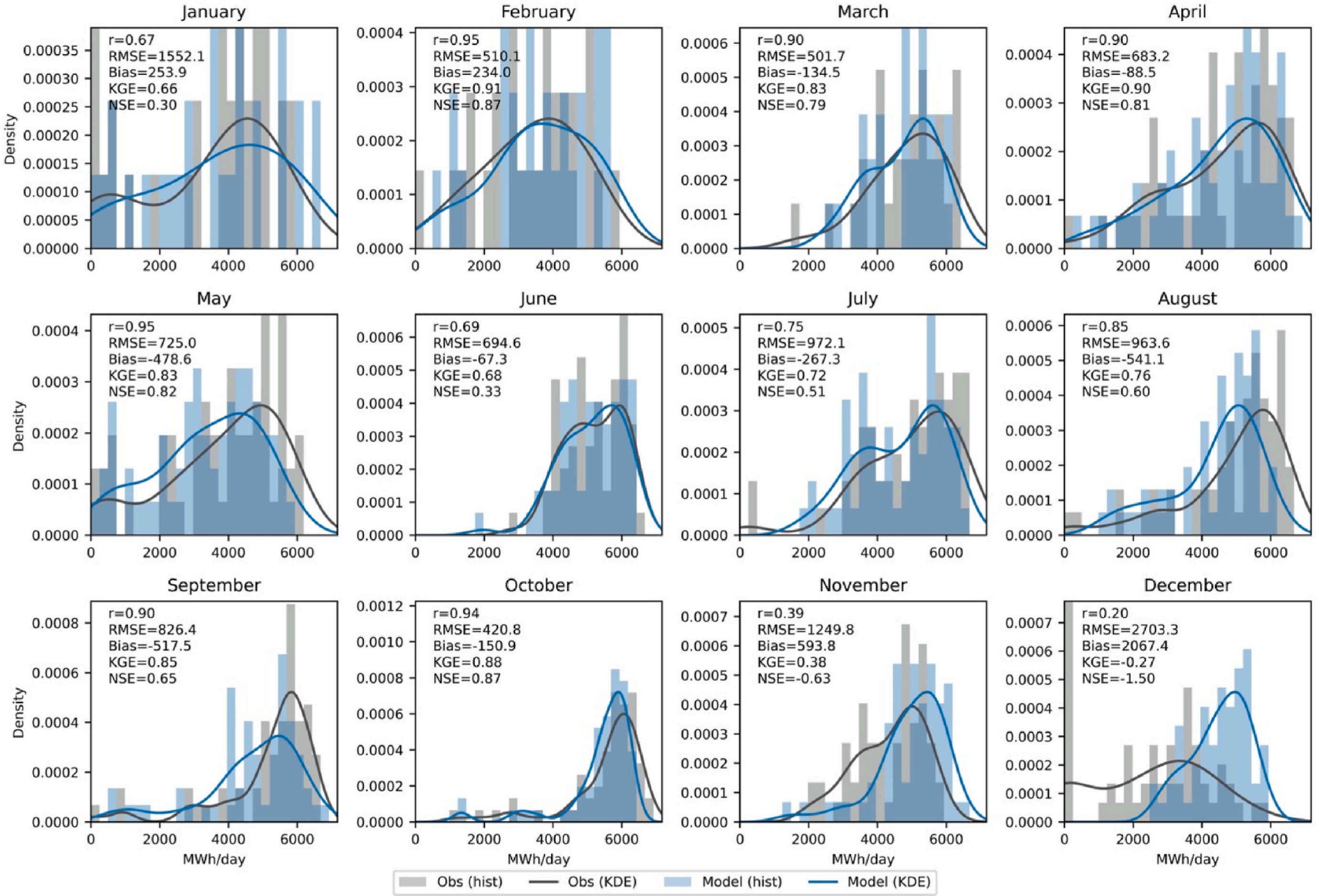

Fig. 6. Monthly probability density functions (PDFs) of daily modelled and observed wind energy at LTWP at 50m hub-height, comparing distributions across January to December. Each subplot displays histograms and kernel density estimates (KDEs) for observed data (grey histogram, black KDE) and modelled data (light blue histogram, blue KDE). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

regimes; here, the model captures the monthly pattern but retains some bias in the tails. Consultations with KenGen-Ngong and Kipeto Energy indicated that May represents a period of reduced wind generation, and routine maintenance was typically scheduled during that month.

3.4. *Effect of non-meteorological factors*

Duration curves (Fig. 7) show that the model captures generation ranking but tends to overestimate output, especially on high-production days. Excluding August and December (for LTWP), and May and December (for Kipeto and Ngong) improves fit, reduces RMSE, and strengthens correlations. This suggests that the model's apparent underperformance is driven by non-meteorological factors rather than inherent flaws. At LTWP, this adjustment greatly improves fit (r rises from 0.72 to 0.93; NSE from 0.50 to 0.87), with the filtered model closely matching observed data (Fig. 8). Kipeto also shows better results (r from 0.72 to 0.85; NSE from −0.17 to 0.31) and Ngong (r from 0.75 to 0.82; NSE from 0.17 to 0.31). However, persistent biases may suggest systematic issues linked to resource dispatch.

3.5. *Persistence of extremes and ramp characteristics*

Persistence-of-extremes and rapid-change diagnostics provide a dynamical perspective on the modelled CF distributions. For each site, the mean annual frequency of persistent low- and high-CF events as a function of duration, and the frequency of ramps exceeding a given ΔCF for 3, 6 and 12-h windows have been quantified (Fig. 9 and Appendix B). Low-CF thresholds correspond to the 10th, 20th and 30th percentiles of the CF distribution, while high-CF thresholds correspond to the 90th, 80th and 70th percentiles. Panels (a) and (d) quantify the risk of "energy droughts" - periods where generation remains consistently below the 10th, 20th, and 30th percentiles (corresponding to CF thresholds of 18.1%, 34.6%, and 47.1%, respectively). At LTWP, persistent low-CF episodes are relatively rare and short, whereas high-CF regimes persist for several hours to days, and large ramps are comparatively infrequent, indicating a strong and stable wind power regime (Fig. 9). The "extreme low" threshold (P10) is 18.1% CF, suggesting that a "crisis" low-wind event still produces power at ~18% CF. Events where the CF stays below 18.1% for more than 2 days occur fewer than 2 times a year (see blue line, Fig. 9, panel d).

Panels (b) and (e) analyse how often the wind farm operates as a "baseload" plant, sustaining output above the 70th, 80th, and 90th percentiles (87.3%, 81.8%, and 76.4% CF, respectively). Unlike the low-wind events (whose curves peak to the left), the high-wind events have slow decay curves. The LTWP farm frequently hits >87% capacity (blue line), sustaining that level for more than 10 h.

Panels (c) and (f) quantify the "ramping" challenge - how fast power output changes (ΔCF) over 3, 6, and 12-h windows (t_{win}). For all time windows (3h, 6h, 12h) in panel c, the curves peak at the left (ΔCF near 0%), confirming that most of the time, the wind power output at LTWP is relatively stable with only minor fluctuations. The 3-h window (blue) curve drops off the fastest. This indicates that large power changes rarely happen quickly. For instance, cumulative changes exceeding 40% of CF within 3 h are not frequent. The 12-h window (black) curve extends

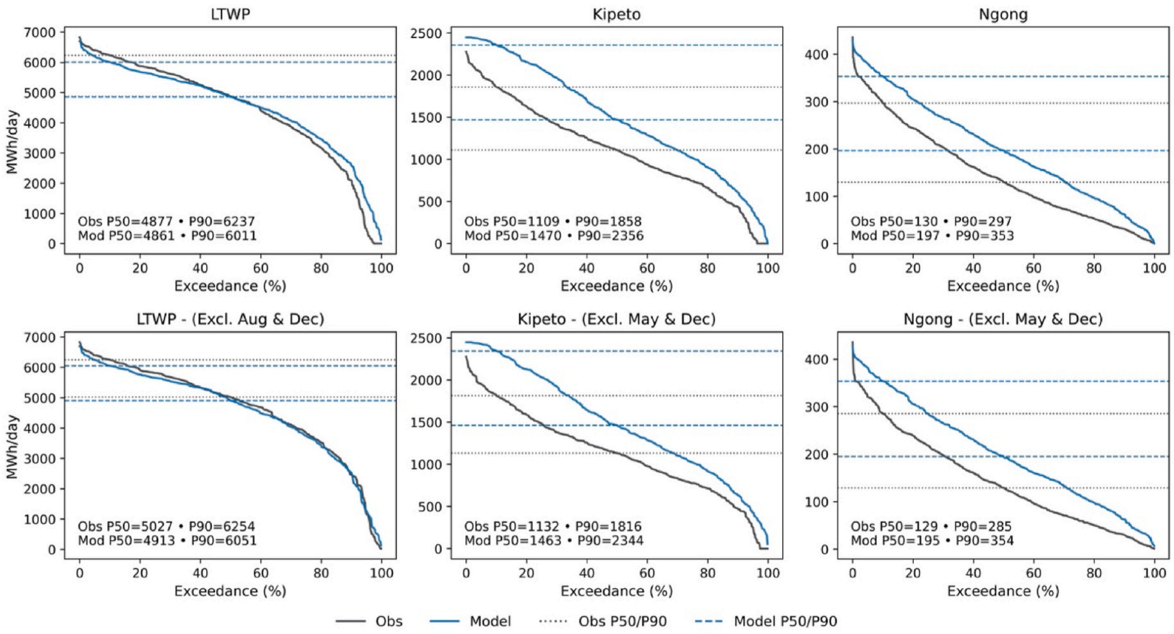

Fig. 7. Duration curves comparing observed and modelled (ERA5-derived) wind power generation for the three Kenyan sites, LTWP, Kipeto, and Ngong, at hub height. The top row shows daily generation for the full validation period, while the bottom row excludes August and December (for LTWP), May and December (for Kipeto and Ngong). The plots show the exceedance probability (%) against daily energy (MWh/day) for both observed (black) and modelled (blue) series. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

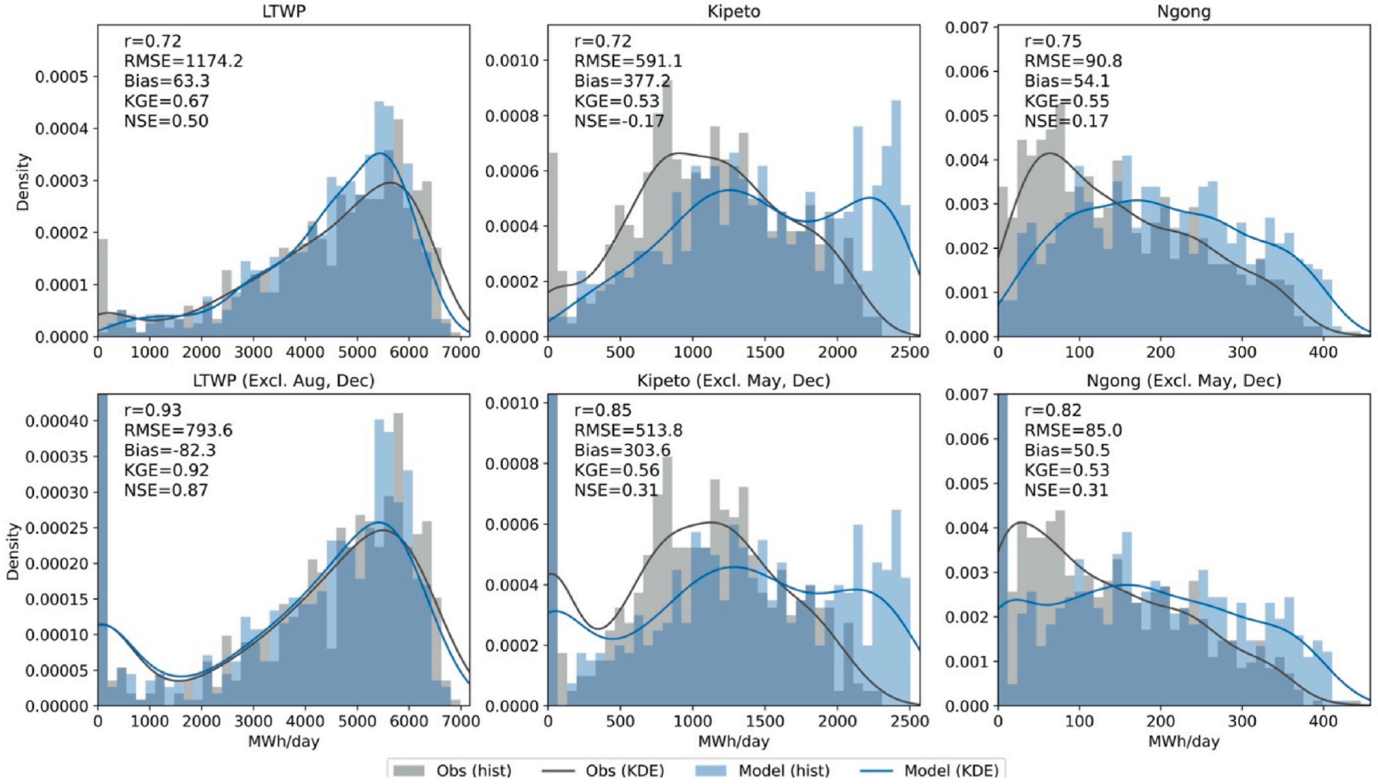

Fig. 8. Probability density functions (PDFs) comparing observed and modelled daily energy generation (MWh/day) for the LTWP, Kipeto, and Ngong wind farms. The top panels show the full validation period comparison, while the bottom panels exclude August and December (for LTWP), May and December (for Kipeto and Ngong). Histograms represent empirical distributions, while the smooth curves denote kernel density estimates (KDEs). Statistical metrics (r, RMSE, Bias, KGE, NSE) indicate model performance relative to observations.

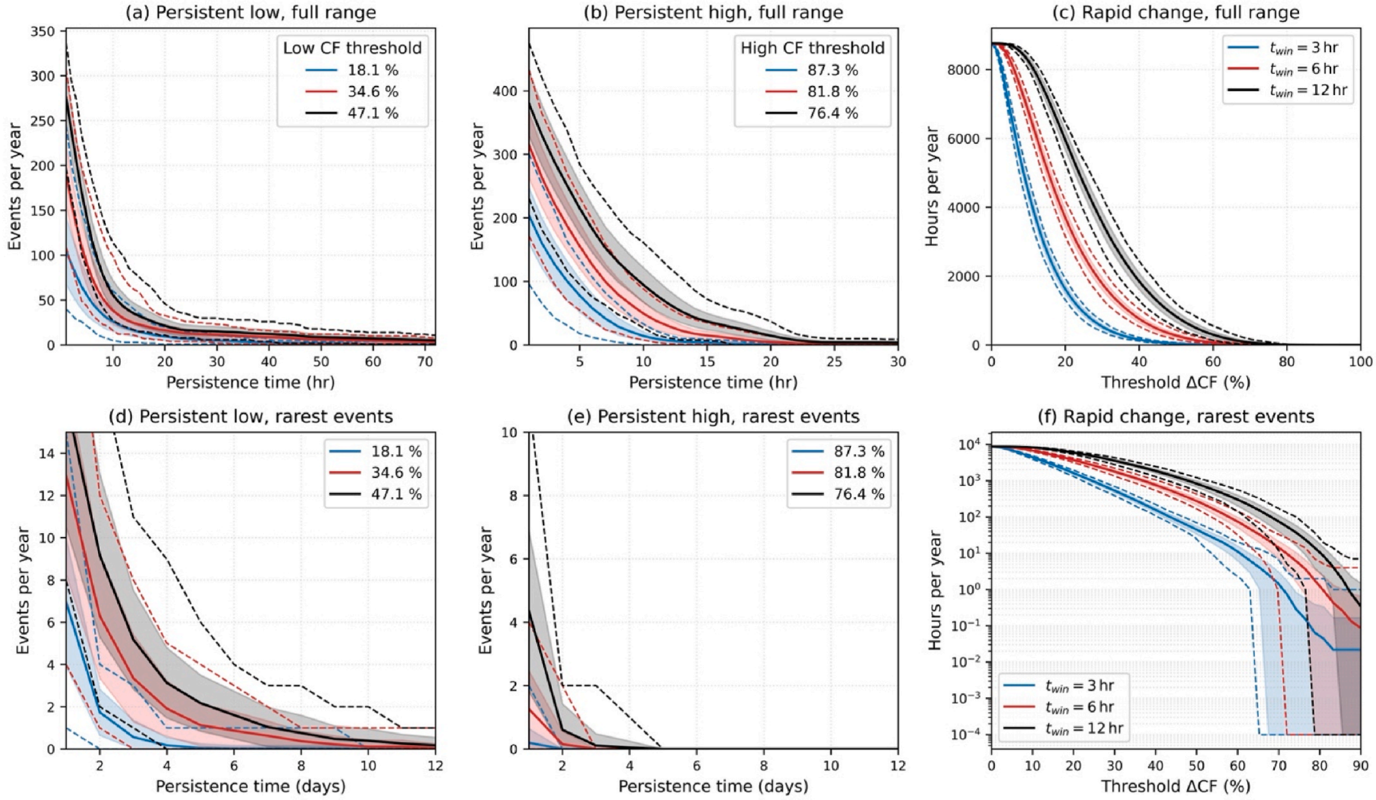

Fig. 9. LTWP - Persistence of extremes and rapid changes in CF. Panels (a,b,d,e) show the mean annual frequency of persistent low- and high-CF events versus duration, with shading and dashed lines indicating interannual spread. Low-CF thresholds correspond to the 10th, 20th and 30th percentiles of the CF distribution, while the high-CF thresholds correspond to the 90th, 80th and 70th percentiles. Panels (c,f) give hours per year with ramps ≥ ΔCF for 3, 6 and 12-h windows (full range and rare tail).

much further to the right. This shows that while the CF doesn't change instantly, it frequently undergoes large swings over half a day, reflecting the natural diurnal cycle of wind power. Panel f uses a logarithmic y-axis to zoom in on the rarest, most extreme events (i.e., the "stress tests" for the grid). The 3-h (blue) curve drops fastest, indicating short-term stability. Ramps exceeding 60% of CF in 3 h are statistically negligible.

Kipeto exhibits more frequent and longer high-CF events, less persistent low-CF states, and more hours per year with large ramps across all time windows, implying a more volatile resource (Appendix B: Figure B1). Ngong experiences frequent short low-CF events, moderate persistence of high-CF episodes, and substantial ramp activity for moderate ΔCF (Appendix B: Figure B2).

Two-dimensional persistence maps (Fig. 10) highlight grid reliability patterns. At Kipeto and Ngong, high-generation events fade quickly with duration, while LTWP shows rare low-wind persistence (<1 event/year) and exceptional high-wind persistence, sustaining >70% capacity for over 5 days. Aggregated fleet analysis (Fig. 11) reveals resilience in the overall wind power fleet: deep generation lapses at individual sites are offset collectively, making simultaneous low output (<10% CF) unlikely (Fig. 11a). High-generation persistence is moderated compared to LTWP alone, but the fleet still sustains >50% capacity for weeks (Fig. 11b).

The seasonal 2-D maps show strong intra-annual structure (Appendix B: Figures B3 and B4). At LTWP, JJA is characterised by almost no long low-CF events and very persistent high-CF states (Figure B3), consistent with the season when the Turkana jet is strongest (Nicholson, 2016). In contrast, December-January-February (DJF) and March-April-May (MAM) show more and longer low-CF spells (Figure B4). Kipeto also exhibits seasonality, with enhanced high-CF persistence in DJF and SON and more common low-CF events in MAM and JJA. Ngong shows the least persistent extremes overall, with relatively frequent short low-CF spells in all seasons and only modest extensions of high CF persistence in DJF. Together, these diagnostics confirm LTWP as the most stable and consistently productive site, while Kipeto and especially Ngong are more prone to both prolonged low-generation periods and sharper ramps.

4. Discussion

These results demonstrate that while raw ERA5 data provides a reliable synoptic signal for East Africa, it fails to capture the wind speed magnitude required for bankable energy assessments. This deficiency aligns with recent findings by Ref. [34], who emphasised that reanalysis products require site-specific transformations to account for local complexities. While [34] demonstrated that ERA5 is sufficiently reliable for predicting temporal patterns in offshore and flat onshore locations, they noted significant spatial limitations in complex terrain. Our results extend this dialogue by showing that in East Africa, the primary hurdle is a systematic amplitude bias rather than a failure in the temporal cadence. The success of WQM in this study reinforces the "optimal strategy" argument by Ref. [15] and further confirms that the transformation of reanalysis data, as advocated by Ref. [34], is a global necessity for reducing uncertainty in regions with high orographic complexity.

A pivotal finding of this study is the explicit decoupling of atmospheric model error from operational performance. Previous literature, e.g., assessment by Ref. [35], noted that mismatches between models and observed data are frequently attributed to "model error." By identifying specific deviations in August, May and December, corresponding to scheduled maintenance and grid curtailment, we demonstrate that these are not failures of the atmospheric model, but rather technical and operational constraints. This distinction is critical for the "Gross-to-Net" loss frameworks defined in IEC 61400-15 [36]. Our work suggests that pre-construction yield assessments may fail in emerging markets, not

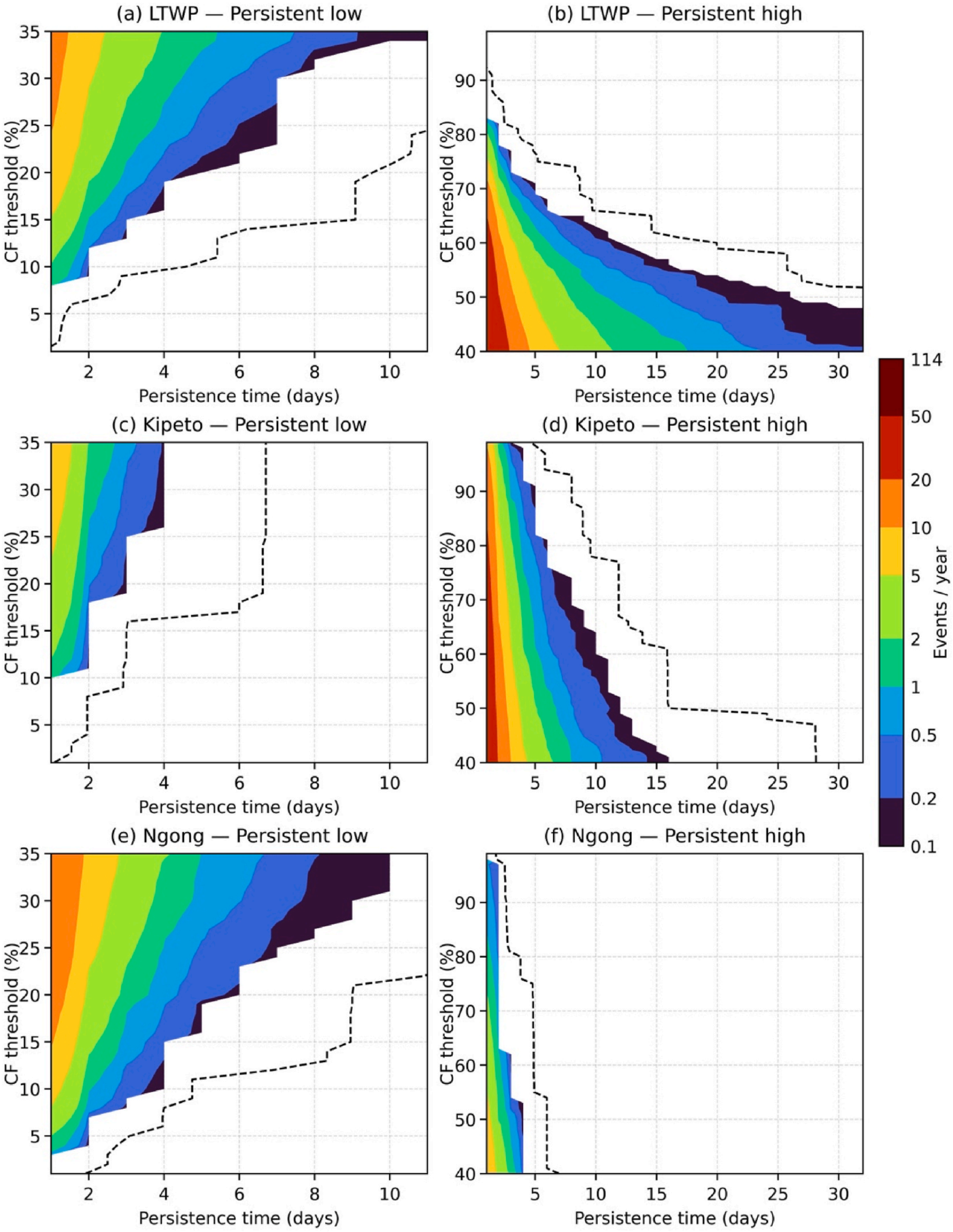

Fig. 10. Persistent of extremes in modelled CF at LTWP, Kipeto and Ngong. Panels (a, c, e) show the mean number of persistent low-generation events per year, expressed as a function of the CF threshold below which generation remains for at least the indicated persistence time (x-axis). Panels (b, d, f) show the corresponding persistent high-generation events, for CF above the indicated threshold. Shading gives the mean frequency of events (events per year, logarithmic scale; common colour bar), and the black dashed curve in each panel marks the maximum persistence observed in the full multi-year record for each threshold. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

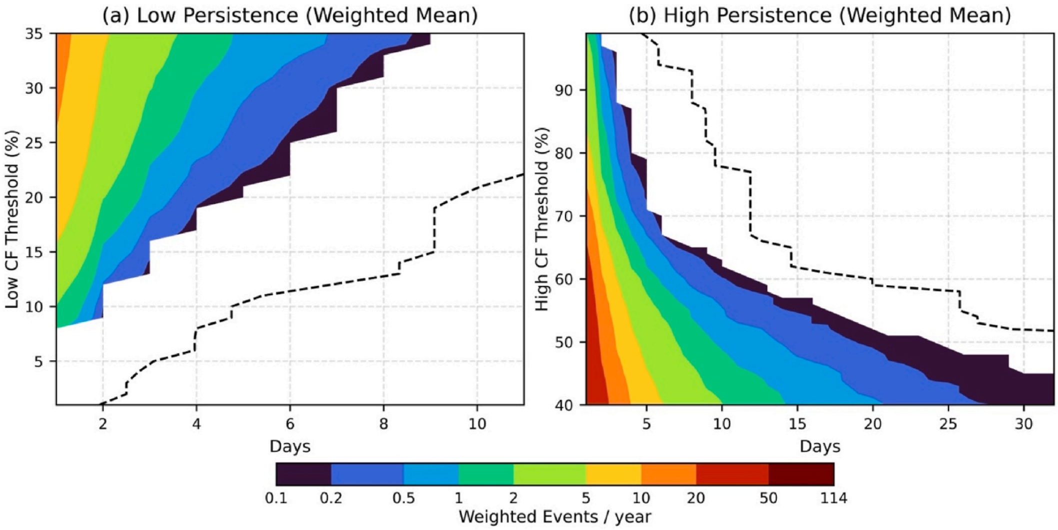

Fig. 11. Aggregate weighted persistence of wind power extremes. The panels display the averaged frequency of continuous generation events across the three sites (LTWP, Kipeto, Ngong), providing a representative "regional weighted mean" profile for Kenya. The left panel ("Low Persistence") tracks the mean annual frequency of wind droughts where the CF remains continuously below a given threshold. The right panel ("High Persistence") tracks baseload-like events where the CF remains continuously above a threshold.

because the meteorological models are inherently flawed, but because the non-meteorological availability losses are poorly parameterised. This adds a necessary layer of nuance to the discourse regarding pre-construction yield overestimation. We argue that model validation is incomplete unless it can differentiate between meteorological error and operational loss.

Sensitivity analysis identifies vertical scaling as the first-order control on model performance. This extreme sensitivity is a direct fingerprint of the Turkana Low-Level Jet (LLJ). The Turkana Low Level Jet presents a velocity maximum that is often situated well above the surface layer represented by 10 m reanalysis outputs [37]. In recent field campaigns, ERA5 has been unable to resolve the full intensity of this jet core [38], confirming our conclusion that the extrapolation of hub-height is not a mere refinement of the method but a prerequisite to fully capturing the potential resources within the region. The remarkable sensitivity to hub-height scaling highlights the importance of resource estimates in jet-dominated regions, such as Turkana, that go beyond standard reanalysis inputs. Yield estimates without site-specific vertical profiles risk systematic underestimation, a major blow for investment confidence as well as grid planning.

The persistence and ramp-based diagnostics challenge traditional wind power archetypes that typically attribute high stability largely to offshore settings. For instance, [39] demonstrated that offshore wind in the Iberian Peninsula exhibits significantly higher availability and lower seasonal variability than its onshore counterparts. However, our findings reveal that LTWP exhibits a persistence profile that rivals or even exceeds these European offshore and onshore [13] benchmarks. While [39] emphasises that offshore wind can reduce grid dependence on a single energy source due to its high CF thresholds, LTWP provides a similar stability in an onshore tropical setting. Moreover, the geographic smoothing between the northern (Turkana) and southern (Kipeto/Ngong) clusters can lead to a "portfolio effect" – referring to the increase of the overall stability of system [40]. Such decorrelation of the continuous Turkana Jet with thermally driven southern flows is also consistent with complementarity strategies proposed for the Iberian grid, suggesting that spatial diversification is just as much of a core demand driver for Kenya's grid reliability as it is for mature European markets.

5. Conclusion

This study provides empirical validation of a robust process-based framework for converting global ERA5 reanalysis data into investment-grade wind energy products and characterisation of wind power extremes in data-scarce areas. The approach assumes (1) observation timestamps are correct and representative of farm-average conditions, (2) a Weibull distribution is a reasonable fit for wind speed within groups, (3) the calibration window is representative of the longer application period, and (4) the dynEff curve is a pragmatic approximation of operational effects. These are transparent assumptions, and the pipeline is designed so each can be revisited: change the training window, swap the shear model, or update the efficiency curve as new evidence emerges.

This composite approach succeeded in overcoming the scale gap between coarse atmospheric models and farm-level operation by incorporating Weibull Quantile Mapping with dynEff scaling and hub-height extrapolation to reconstruct the hourly and daily wind power generation profiles of Kenya's wind farms. We derive three significant conclusions from our results.

First, statistical bias correction is essential but insufficient on its own; while WQM restores the distributional fidelity of the wind resource, correcting the "thin tails" typical of reanalysis products, it cannot resolve discrepancies driven by non-meteorological factors. The persistence of residual errors in August and December quantifies the significant impact of scheduled maintenance and grid curtailment on realised yield. This underscores that future resource assessments must explicitly decouple meteorological availability from technical availability to avoid inflating performance expectations.

Second, the sensitivity analysis confirms that vertical wind shear is the dominant driver of economic viability in the region. The dramatic improvement in CF distribution when scaling from 50 m to 100 m highlights the necessity of high-tower technology to access the accelerated core of the Turkana Low-Level Jet. Dynamic efficiency parameterisation serves as a critical secondary constraint, ensuring that these high-wind projections remain physically realistic. The reliability test confirms that the modelling framework is robust to incomplete data and moderate random noise, but its reliability degrades under anomalous spikes or extreme erroneous inputs. The results therefore identify outlier contamination as the dominant factor constraining the model's reliability boundary in engineering applications.

Finally, the persistence and ramp diagnostics reveal a strategic opportunity for grid diversification. The Lake Turkana Wind Power (LTWP) site exhibits an exceptional “baseload” characteristic, with high-output events sustained for weeks, contrasting sharply with the volatile, synoptic-scale wind droughts often observed in European fleets. Kipeto and Ngong Hills exhibit strong diurnal cycling, characterised by persistent intermittency. To system planners, this suggests that in Kenya, grid stability is better achieved not by treating wind as a monolith, but by leveraging the spatial complementarity between the continuous Turkana Jet and the thermally driven flows to the south to balance seasonal transmission needs with diurnal storage solutions. Ultimately, this framework provides a replicable pathway for unlocking wind power potential across the African continent, reducing reliance on sparse ground-based campaigns while maintaining rigorous standards for uncertainty analysis.

Although this framework improves the physical realism of wind-to-power conversion, non-meteorological influences on wind farm output remain only partially represented. In particular, the calibrated dynEff term absorbs part of the residual difference between baseline modelled power and observed generation, but it does not explicitly separate the effects of curtailment, turbine availability, maintenance outages, wake-management actions, grid dispatch constraints, or other operational controls. As a result, some of the remaining model–observation mismatch may reflect non-meteorological processes that are not independently resolved in this framework. Future work can compare Weibull-based WQM with alternative bias-correction methods and develop explicit representations of non-meteorological loss factors using operational datasets such as SCADA, curtailment, and turbine-availability records.

Availability of data and code

The ERA5 reanalysis data used in this study are publicly available from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu) under the Copernicus Climate Change Service license. Operational generation and in-situ wind speed measurements from the LTWP, Kipeto and Ngong Hills wind farms were provided by the respective operators under data-sharing agreements and are not publicly available due to confidentiality restrictions. Derived and aggregated datasets supporting the findings of this study are available from the corresponding author upon reasonable request. All data processing and modelling routines were implemented in Python (v3.10) using open-source scientific libraries. Example scripts and documentation supporting this workflow are available from the corresponding author upon reasonable request and will be deposited in a public repository (GitHub) upon publication.

Declaration of Generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used Microsoft Copilot in order to identify, organise and synthesise literature. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Cohen Ang'u: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Hannah C. Bloomfield:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Linda C. Hirons:** Conceptualization, Funding acquisition, Methodology, Project administration, Writing – review & editing. **Steven J. Woolnough:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **David J. Brayshaw:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **Wilson Gitau:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **Gibbon I.T. Masukwedza:** Conceptualization, Formal analysis, Methodology, Writing – review & editing. **Joseph Mutemi:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **Willis Ochieng:** Conceptualization, Funding acquisition, Methodology, Resources, Writing – review & editing. **Daniel Olago:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **Christopher Oludhe:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing. **Caroline M. Wainwright:** Conceptualization, Funding acquisition, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research was supported by UKRI through the International Science Partnership Fund (ISPF) as part of the UKRI469: Potential of sub-seasonal Operational Weather and climate information for building Energy Resilience in Kenya (POWER-Kenya) project, through the OPP680: STFC Africa UK Physics Partnership Collaborative Research Projects Opportunity. Woolnough, Hirons and Masukwedza are supported by the National Centre for Atmospheric Science (NCAS) through the NERC National Capability International Programme Award (NE/X006263/1). The authors acknowledge the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing ERA5 reanalysis data through the Copernicus Climate Data Store. We thank the Kenya Electricity Generating Company (KenGen) for granting access to site-level operational and meteorological data. The views expressed are those of the authors and do not necessarily reflect those of the supporting institutions. This work used JASMIN, the UK's collaborative data analysis environment (https://www.jasmin.ac.uk).

Appendix A. Monthly verification statistics at the Kipeto and Ngong wind farms:

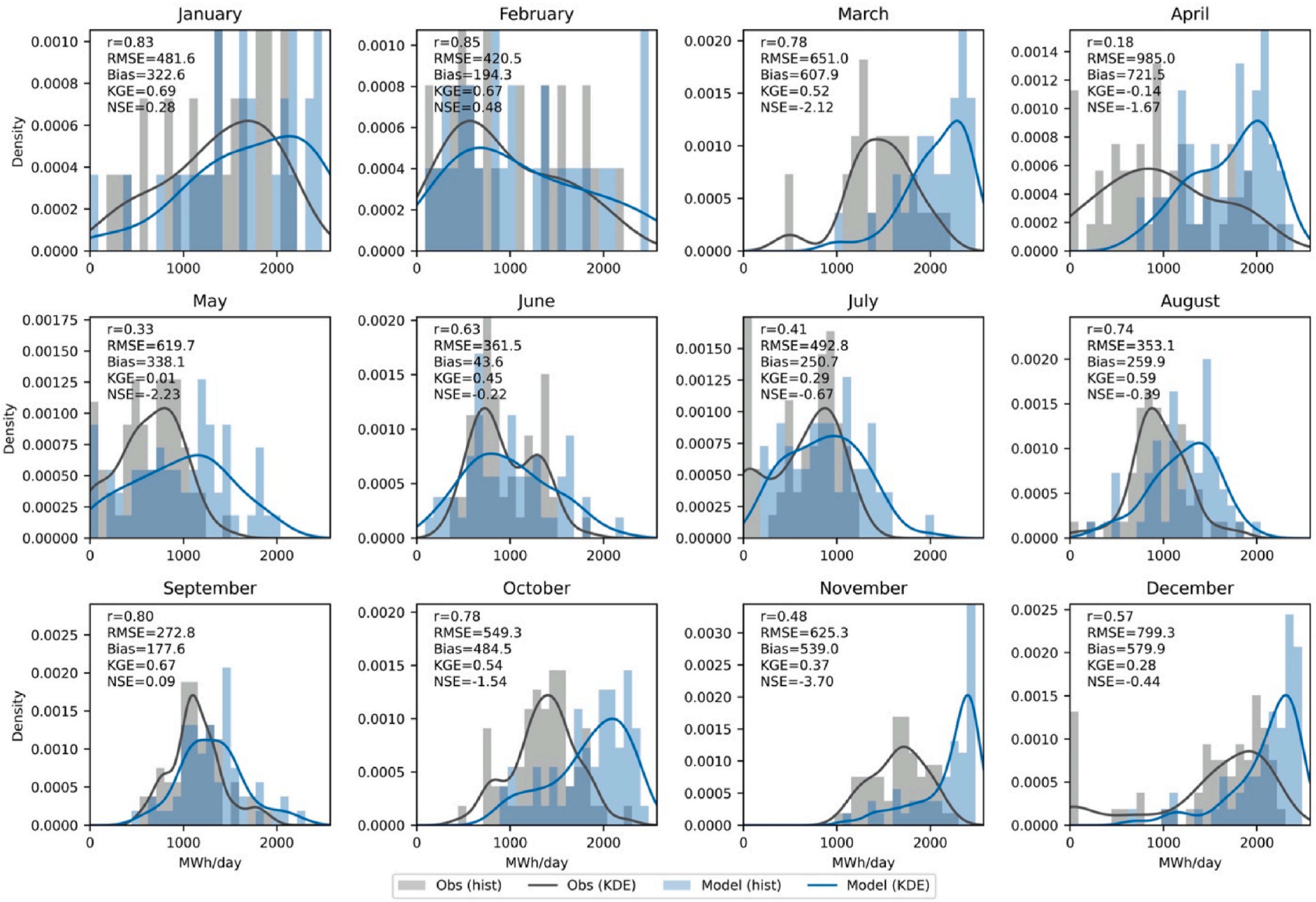

Fig. A1. Monthly probability density functions (PDFs) of daily modelled and observed MWh/day values for Kipeto at 80m hub-height, comparing distributions across January to December. Each subplot displays histograms and kernel density estimates (KDEs) for observed data (grey histogram, black KDE) and modelled data (light blue histogram, blue KDE).

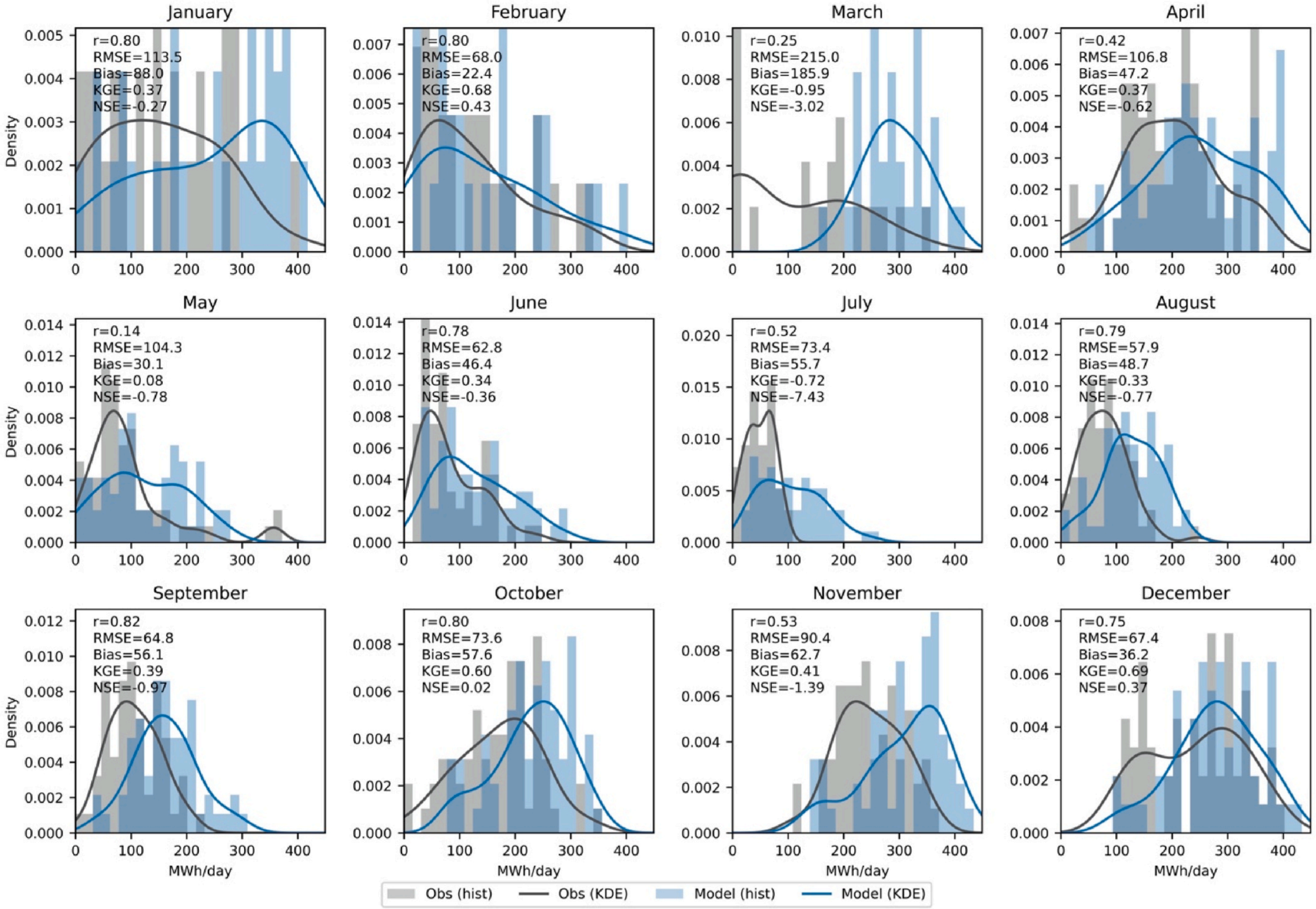

Fig. A2. Monthly probability density functions (PDFs) of daily modelled and observed MWh/day values at Ngong at 50m hub-height, comparing distributions across January to December. Each subplot displays histograms and kernel density estimates (KDEs) for observed data (grey histogram, black KDE) and modelled data (light blue histogram, blue KDE).

Appendix B. Extreme events at the Kipeto and Ngong wind farms:

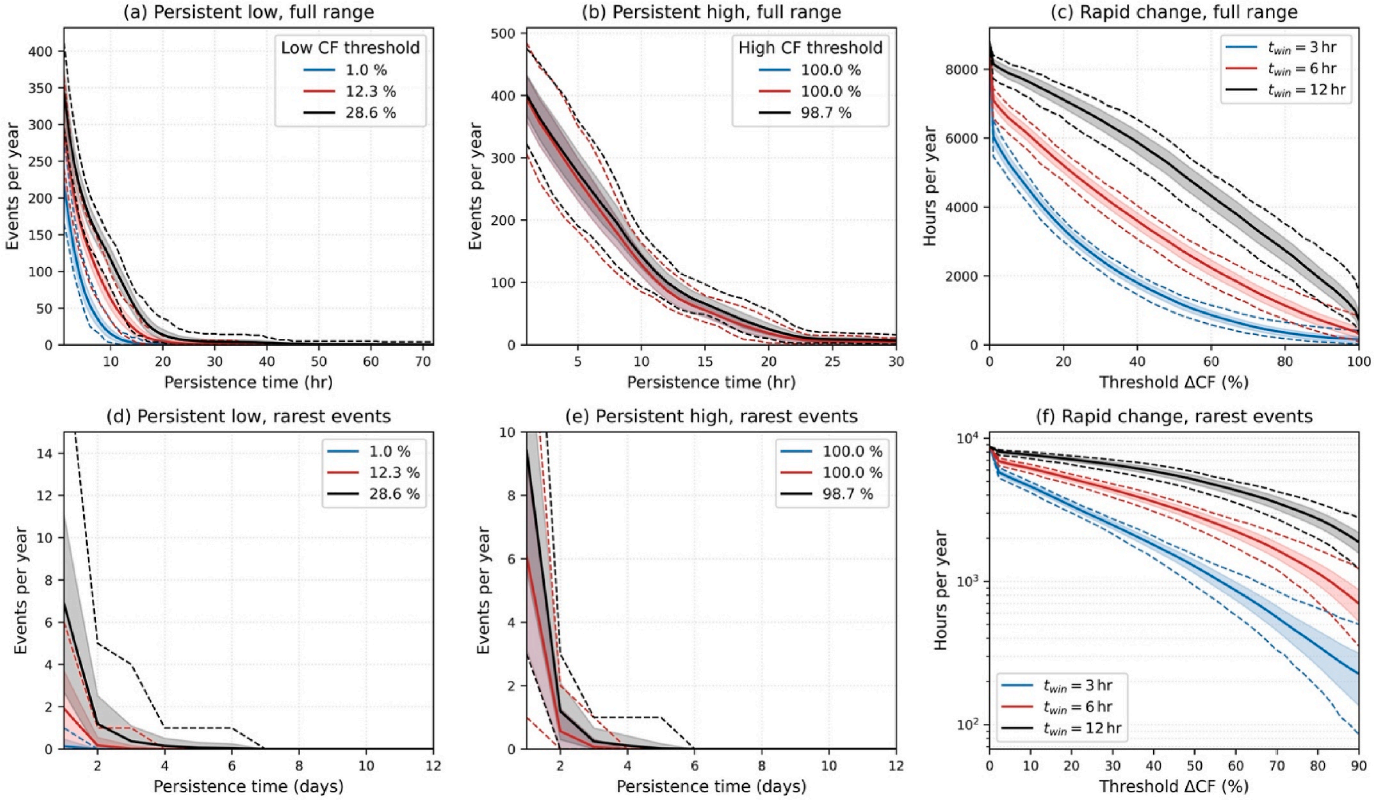

Fig. B1. Kipeto - Persistence of extremes and rapid changes in CF. Panels as in Fig. 8. Kipeto shows more frequent and longer low-CF events, less persistent high-CF states, and more hours per year with large ramps across all time windows.

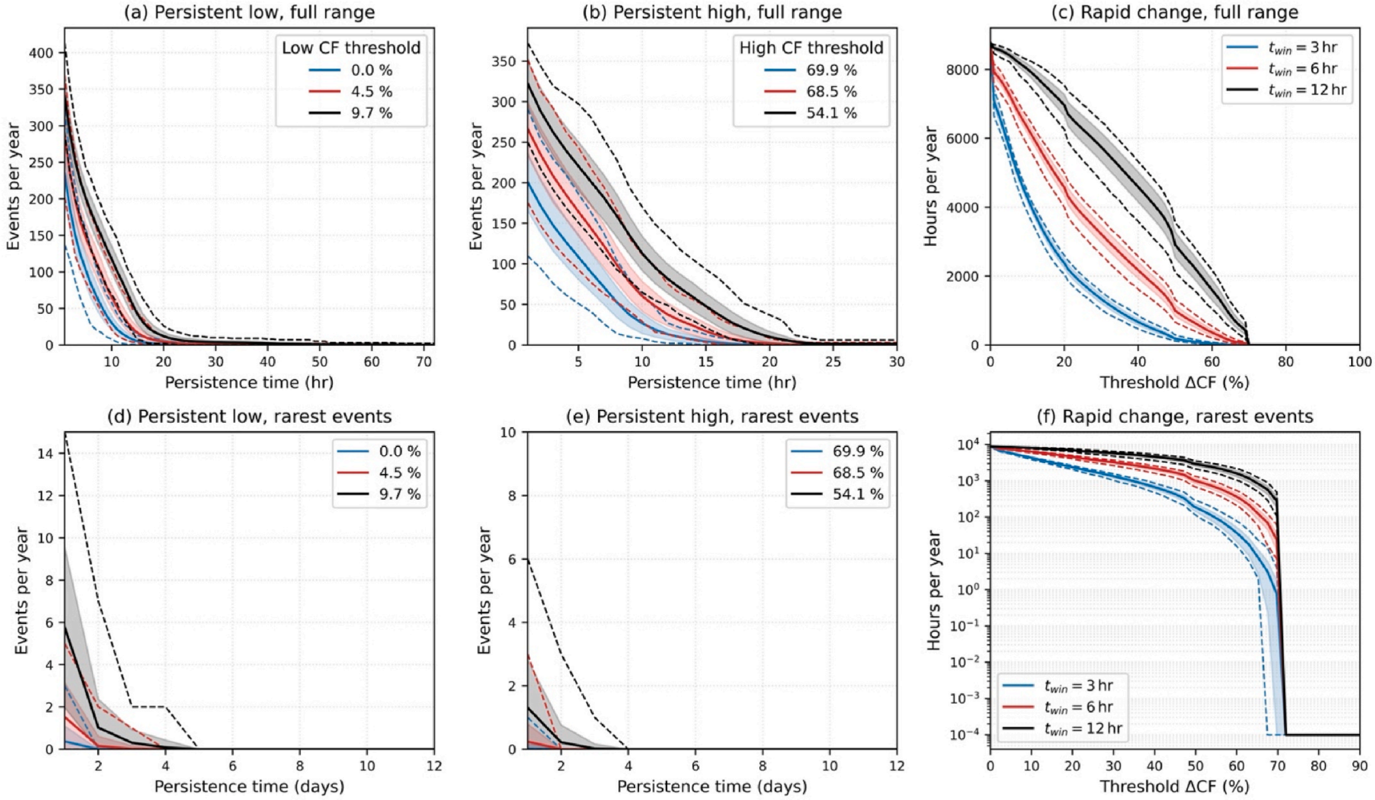

Fig. B2. Ngong - Persistence of extremes and rapid changes in CF. Panels as in Fig. 8. Ngong exhibits frequent short low-CF events, moderate persistence of high-CF episodes, and substantial ramp activity, especially for moderate ΔCF.

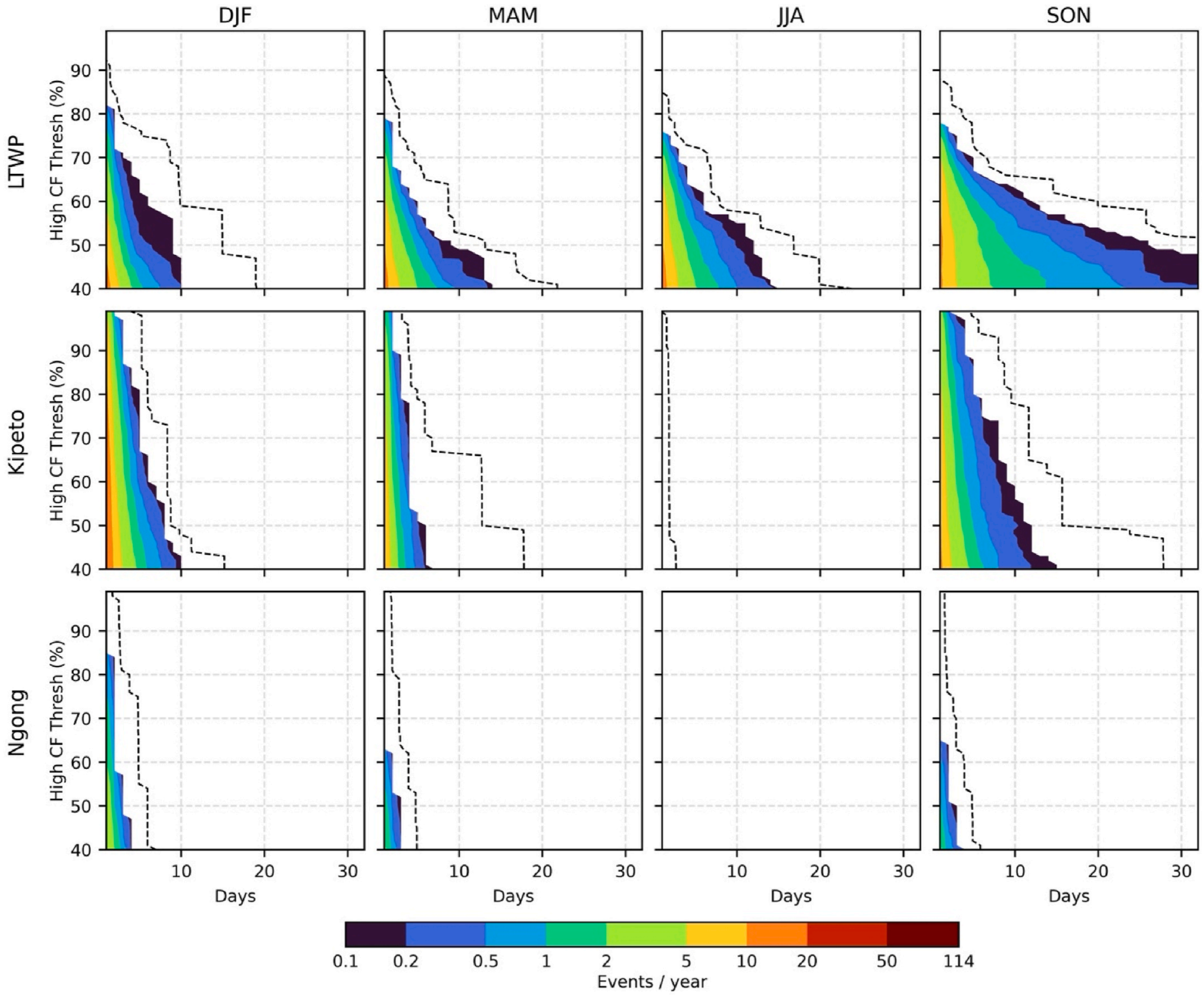

Fig. B3. Seasonal persistent high extremes in CF at LTWP, Kipeto and Ngong. Rows show sites (top–bottom: LTWP, Kipeto, Ngong); columns show seasons and event type (DJF, MAM, JJA, SON; low-CF extremes on the left pair and high-CF extremes on the right pair of each season). Each panel gives the mean number of events per year (colour shading, log-spaced 0.1–114 events yr^{-1}) as a function of CF threshold (y-axis) and persistence time in days (x-axis). Dashed lines mark the longest single event observed over the full record for each threshold.

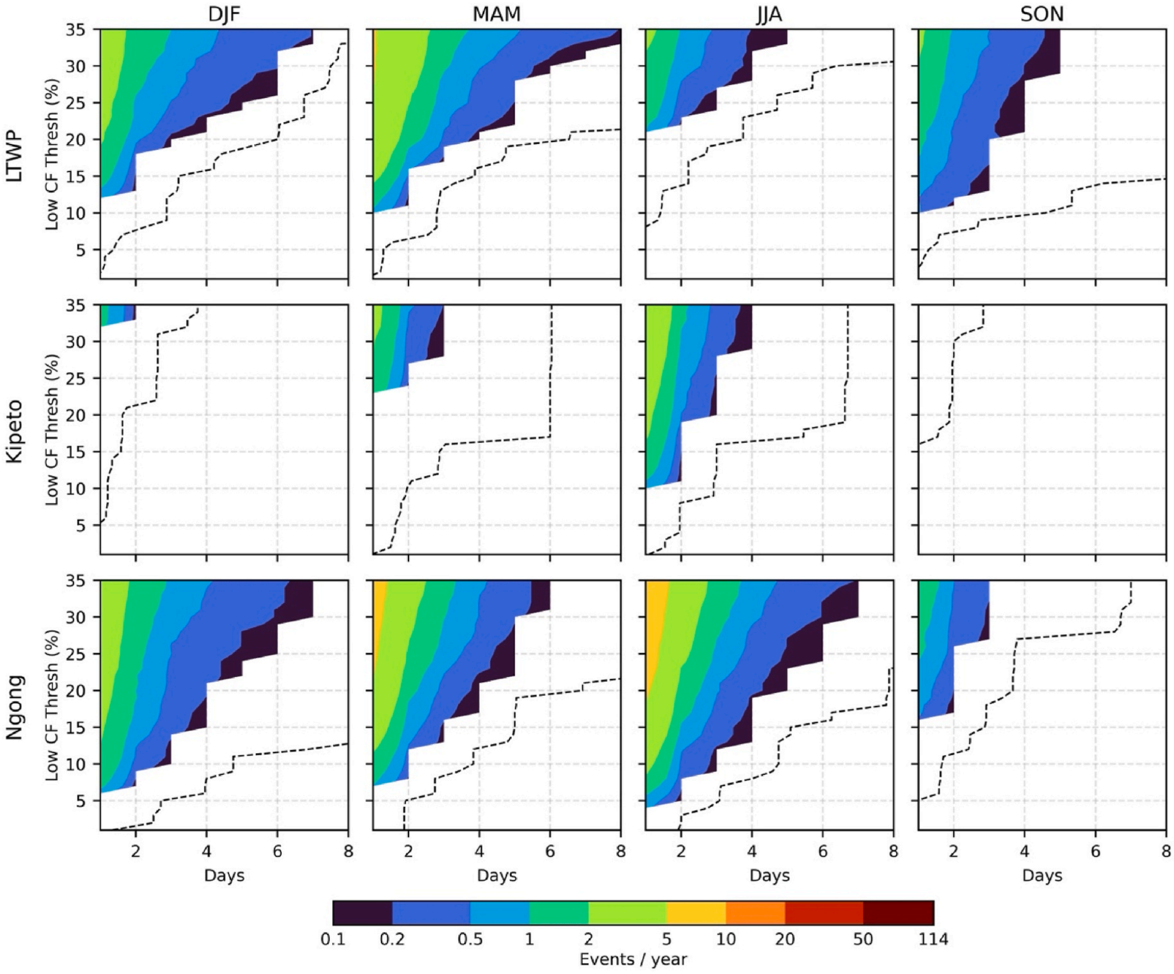

Fig. B4. Seasonal persistent low extremes (wind droughts) in CF at LTWP, Kipeto and Ngong. Rows show sites (top–bottom: LTWP, Kipeto, Ngong); columns show seasons and event type (DJF, MAM, JJA, SON; low-CF extremes on the left pair and high-CF extremes on the right pair of each season). Each panel gives the mean number of events per year (colour shading, log-spaced 0.1–114 events yr^{-1}) as a function of CF threshold (y-axis) and persistence time in days (x-axis). Dashed lines mark the longest single event observed over the full record for each threshold.

Appendix C. Reliability-Boundary Test

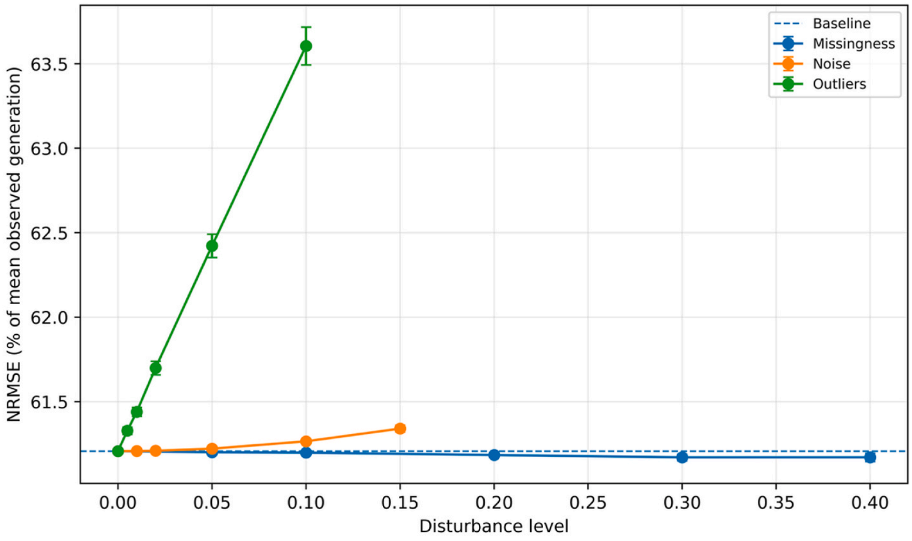

Fig. C1. Reliability-boundary analysis for the LTWP model under controlled disturbance scenarios. The x-axis shows disturbance level, and the y-axis shows normalized RMSE as a percentage of mean observed generation. Curves represent missingness, Gaussian noise, and outlier contamination, while the dashed line indicates baseline performance under undisturbed conditions.

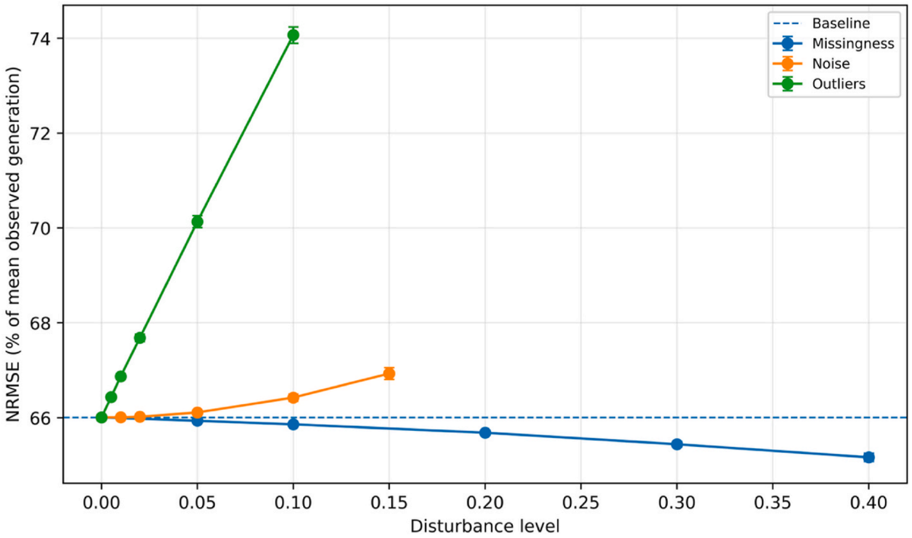

Fig. C2. Reliability-boundary analysis for the Kipeto model under controlled disturbance scenarios. The x-axis shows disturbance level, and the y-axis shows normalized RMSE as a percentage of mean observed generation. Curves represent missingness, Gaussian noise, and outlier contamination, while the dashed line indicates baseline performance under undisturbed conditions.

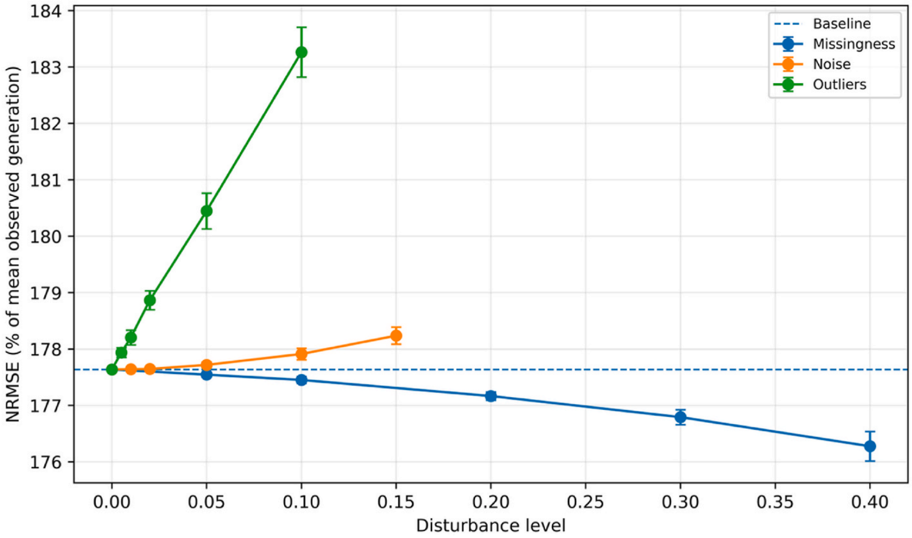

Fig. C3. Reliability-boundary analysis for the Ngong model under controlled disturbance scenarios. The x-axis shows disturbance level, and the y-axis shows normalized RMSE as a percentage of mean observed generation. Curves represent missingness, Gaussian noise, and outlier contamination, while the dashed line indicates baseline performance under undisturbed conditions.

References

[1] Nwankwo Charles Uzondu, Dominic Dummene Lele, Comprehensive analysis of advancements in wind turbine design and offshore wind energy integration: technological innovations, economic viability, and environmental impacts, Int. J. Appl. Res. Soc. Sci. 6 (8) (Aug. 2024) 1538–1556, https://doi.org/10.51594/ijarss.v6i8.1369.

[2] R. McKenna, et al., System impacts of wind energy developments: key research challenges and opportunities, Joule 9 (1) (Jan. 2025) 101799, https://doi.org/10.1016/j.joule.2024.11.016.

[3] K.B. Karnauskas, J.K. Lundquist, L. Zhang, Southward shift of the global wind energy resource under high carbon dioxide emissions, Nat. Geosci. 11 (1) (Jan. 2018) 38–43, https://doi.org/10.1038/s41561-017-0029-9.

[4] J.N. Kamau, R. Kinyua, J.K. Gathua, 6 years of wind data for Marsabit, Kenya average over 14m/s at 100m hub height; an analysis of the wind energy potential, Renew. Energy 35 (6) (Jun. 2010) 1298–1302, https://doi.org/10.1016/j.renene.2009.10.008.

[5] J.K. Kiplagat, R.Z. Wang, T.X. Li, Renewable energy in Kenya: resource potential and status of exploitation, Renew. Sustain. Energy Rev. 15 (6) (Aug. 2011) 2960–2973, https://doi.org/10.1016/j.rser.2011.03.023.

[6] EPRA, Energy and petroleum statistics report; for the financial year ended 30th June 2024, Energy and Petroleum Regulatory Authority (EPRA) (2024). Nairobi, Kenya.

[7] S. Emeis, Wind energy meteorology, in: Green Energy and Technology, Springer International Publishing, Cham, 2018, https://doi.org/10.1007/978-3-319-72859-9.

[8] H. Hersbach, et al., The ERA5 global reanalysis, Q. J. R. Meteorol. Soc. 146 (730) (Jul. 2020) 1999–2049, https://doi.org/10.1002/qj.3803.

[9] R. Gelaro, et al., The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), J. Clim. 30 (14) (Jul. 2017) 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

[10] S. Saha, et al., The NCEP climate forecast system version 2, J. Clim. 27 (6) (Mar. 2014) 2185–2208, https://doi.org/10.1175/JCLI-D-12-00823.1.

[11] W.S. Parker, Reanalyses and observations: what's the difference? Bull. Am. Meteorol. Soc. 97 (9) (Sep. 2016) 1565–1572, https://doi.org/10.1175/BAMS-D-14-00226.1.

[12] H.C. Bloomfield, D.J. Brayshaw, P.L.M. Gonzalez, A. Charlton-Perez, Sub-seasonal forecasts of demand and wind power and solar power generation for 28 European countries, Earth Syst. Sci. Data 13 (5) (May 2021) 2259–2274, https://doi.org/10.5194/essd-13-2259-2021.

[13] D.J. Cannon, D.J. Brayshaw, J. Methven, P.J. Coker, D. Lenaghan, Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in Great Britain, Renew. Energy 75 (Mar. 2015) 767–778, https://doi.org/10.1016/j.renene.2014.10.024.

[14] L. Dubus, et al., C3S energy: a climate service for the provision of power supply and demand indicators for Europe based on the ERA5 reanalysis and ENTSO-E data, Meteorol. Appl. 30 (5) (Sep. 2023) e2145, https://doi.org/10.1002/met.2145.

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

[16] Y. Mulugetta, et al., Africa needs context-relevant evidence to shape its clean energy future, Nat. Energy 7 (11) (Oct. 2022) 1015–1022, https://doi.org/10.1038/s41560-022-01152-0.

[17] K.L. Nefabas, L. Söder, M. Mamo, J. Olauson, Modeling of Ethiopian wind power production using ERA5 reanalysis data, Energies 14 (9) (Apr. 2021) 2573, https://doi.org/10.3390/en14092573.

[18] Y. Kassem, H. Gökçekuş, M.H.A. Abdalla, Wind energy resource assessment based on the use of multiple satellite data for sustainable energy production in Sudan, Environ. Dev. Sustain. 27 (3) (Nov. 2023) 7779–7815, https://doi.org/10.1007/s10668-023-04221-w.

[19] S.E. Nicholson, Climate and climatic variability of rainfall over eastern Africa, Rev. Geophys. 55 (3) (Sep. 2017) 590–635, https://doi.org/10.1002/2016RG000544.

[20] L. Oscar, et al., Characteristics of the Turkana low-level jet stream and the associated rainfall in CMIP6 models, Clim. Dyn. 62 (9) (Sep. 2024) 8371–8387, https://doi.org/10.1007/s00382-022-06499-4.

[21] C. Munday, et al., The Turkana jet diurnal cycle in observations and reanalysis, J. Clim. 37 (18) (Sep. 2024) 4633–4645, https://doi.org/10.1175/JCLI-D-23-0325.1.

[22] J. Olauson, ERA5: the new champion of wind power modelling? Renew. Energy 126 (Oct. 2018) 322–331, https://doi.org/10.1016/j.renene.2018.03.056.

[23] J. Ramon, L. Lledó, V. Torralba, A. Soret, F.J. Doblas-Reyes, What global reanalysis best represents near-surface winds? Q. J. R. Meteorol. Soc. 145 (724) (Oct. 2019) 3236–3251, https://doi.org/10.1002/qj.3616.

[24] G. Gualtieri, Reliability of ERA5 reanalysis data for wind resource assessment: a comparison against tall towers, Energies 14 (14) (Jul. 2021) 4169, https://doi.org/10.3390/en14144169.

[25] J.H. Kinuthia, Horizontal and vertical structure of the Lake Turkana jet, J. Appl. Meteorol. 31 (11) (Nov. 1992) 1248–1274, https://doi.org/10.1175/1520-0450(1992)031%3C1248:HAVSOT%3E2.0.CO;2.

[26] A.J. Cannon, S.R. Sobie, T.Q. Murdock, Bias correction of GCM precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? J. Clim. 28 (17) (Sep. 2015) 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1.

[27] S. Hempel, K. Frieler, L. Warszawski, J. Schewe, F. Piontek, A trend-preserving bias correction – the ISI-MIP approach, Earth Syst. Dynam. 4 (2) (Jul. 2013) 219–236, https://doi.org/10.5194/esd-4-219-2013.

[28] A. Maciel-Tiburcio, O. Martínez-Alvarado, O. Rodríguez-Hernández, Evaluating bias correction methods for wind power estimation using numerical meteorological models, Renew. Energy 247 (Jul. 2025) 122927, https://doi.org/10.1016/j.renene.2025.122927.

[29] J.A. Carta, P. Ramírez, S. Velázquez, A review of wind speed probability distributions used in wind energy analysis, Renew. Sustain. Energy Rev. 13 (5) (Jun. 2009) 933–955, https://doi.org/10.1016/j.rser.2008.05.005.

[30] M.F. Zambak, C.I. Cahyadi, J. Helmi, T.M. Sofie, S. Suwarno, Evaluation and analysis of wind speed with the weibull and rayleigh distribution models for energy potential using three models, Int. J. Energy Econ. Pol. 13 (2) (Mar. 2023) 427–432, https://doi.org/10.32479/ijeep.12775.

[31] M.R. Tye, D.B. Stephenson, G.J. Holland, R.W. Katz, A weibull approach for improving climate model projections of tropical cyclone wind-speed distributions, J. Clim. 27 (16) (Aug. 2014) 6119–6133, https://doi.org/10.1175/JCLI-D-14-00121.1.

[32] X. Costoya, A. Rocha, D. Carvalho, Using bias-correction to improve future projections of offshore wind energy resource: a case study on the Iberian Peninsula, Appl. Energy 262 (Mar. 2020) 114562, https://doi.org/10.1016/j.apenergy.2020.114562.

[33] T. Burton, N. Jenkins, D. Sharpe, E. Bossanyi, Wind Energy Handbook, first ed., Wiley, 2011 https://doi.org/10.1002/9781119992714.

[34] J.A. Carta, P. Cabrera, Transformation of reanalysis data for improved long-term estimation of wind speed and direction at a target site, Renew. Energy 261 (Apr. 2026) 125280, https://doi.org/10.1016/j.renene.2026.125280.

[35] J.C.Y. Lee, M.J. Fields, An overview of wind-energy-production prediction bias, losses, and uncertainties, Wind Energy Sci 6 (2) (Mar. 2021) 311–365, https://doi.org/10.5194/wes-6-311-2021.

[36] A. Clifton, A. Smith, M. Fields, 'Wind plant preconstruction energy estimates, Current Practice and Opportunities', NREL/TP–5000-64735, 1248798 (Apr. 2016), https://doi.org/10.2172/1248798.

[37] S. Nicholson, The Turkana low-level jet: mean climatology and association with regional aridity, Int. J. Climatol. 36 (6) (May 2016) 2598–2614, https://doi.org/10.1002/joc.4515.

[38] C. Munday, et al., Observations of the Turkana jet and the East African dry tropics: the RIFTJet field campaign, Bull. Am. Meteorol. Soc. 103 (8) (Aug. 2022) E1828–E1842, https://doi.org/10.1175/BAMS-D-21-0214.1.

[39] N. López-Franca, M.Á. Gaertner, E. Sánchez, C. Gutiérrez, M. Ortega, C. Gallardo, Offshore wind energy in the Iberian peninsula: a comparative analysis of availability, persistence, and complementarity with onshore wind and solar photovoltaic generation, Renew. Energy 258 (Feb. 2026) 124903, https://doi.org/10.1016/j.renene.2025.124903.

[40] W. Katzenstein, E. Fertig, J. Apt, The variability of interconnected wind plants, Energy Policy 38 (8) (Aug. 2010) 4400–4410, https://doi.org/10.1016/j.enpol.2010.03.069.