



# Structural defects in the Kelmarsh open SCADA dataset: a provenance audit with file reconstruction and downstream consequences

Adam Park<sup>1</sup>, Luke Yin<sup>2</sup>, Cooper Lee<sup>1</sup>, Andrew Wong<sup>3</sup>, and Yoo Choi<sup>4</sup>

<sup>1</sup>Illinois Mathematics and Science Academy, Aurora, Illinois, USA

<sup>2</sup>Princeton University, Princeton, New Jersey, USA

<sup>3</sup>University of Michigan, Ann Arbor, Michigan, USA

<sup>4</sup>Cornell Tech, New York, New York, USA

**Correspondence:** Adam Park (apark1@imsa.edu)

**Abstract.** The Kelmarsh wind farm dataset, covering six turbines in the United Kingdom from 2016 to 2024, has been used in several dozen wind energy studies since 2022. The studies we read that use it check its quality by removing implausible values, which cannot reveal errors in how the files are organised or how measurements are defined. We examined all 54 annual turbine files for such errors with nine reproducible checks, fixing our hypotheses in advance, and found four kinds. First, most measurements start partway through the record, in one of twelve steps between 2016 and 2023, so whole-year missing-data rates overstate real gaps by up to 73 percentage points. Second, between November 2017 and April 2018 the main wind speed measurement switches from the average of two sensors to a single sensor, making the turbines appear 1.2 to 9.1 % more productive at a given wind speed; the single sensor alone shows a change of  $-0.6$  to  $+1.5$  %. Third, the 2023 and 2024 files store each year in 81 overlapping segments with measurements only in the first copy of each time step, leaving 97.6 % of rows without wind speed or power values. Keeping the first copy recovers the year; keeping the last retains less than 1 % of the data and overstates the annual capacity factor (output as a share of the maximum possible) by 25 to 53 percentage points. Files tested from a second wind farm by the same owner share this structure. Fourth, the energy, voltage, wind speed and temperature records contain smaller inconsistencies and placeholder values. One pre-specified prediction failed, seven others failed as originally worded and one could not be evaluated as specified; all are reported. We provide corrected files, a table flagging suspect values, and the analysis code.

## 1 Introduction

Openly licensed SCADA records from operating wind farms are scarce. The Kelmarsh and Penmanshiel releases by Cubico Sustainable Investments (Plumley and Takeuchi, 2025a, b) are among the open wind datasets catalogued by Effenberger and Ludwig (2022) and have supported work on forecasting (Deignan et al., 2025), condition monitoring (Fiocchi et al., 2025), drivetrain life extension (Tartt et al., 2024), imputation (Hammer and Barber, 2025), yaw misalignment (Jacquet et al., 2026) and neighbour-informed power prediction (Pozzobon et al., 2026). A full-text search of OpenAlex on 2026-09-15 for works mentioning Kelmarsh and wind returned 8, 6, 6, 20 and 38 works for the years 2022 to 2026 (the last to mid-September), of



which we judge roughly 55 to 65 to be wind energy research outputs that use or discuss the data. Formal citations of the four versions of the dataset DOI number 19 in OpenAlex, so most reuse is invisible to citation counts.

25 In the works we read, the data quality of these records is treated as a value-level problem. Martín-Calzada et al. (2025) use the Kelmarsh 1 and 2 records and, as for their other datasets, remove erroneous values and outliers in preprocessing; Hammer and Barber (2025) study the imputation of missing wind speeds from neighbouring turbines. The wider literature on SCADA data quality addresses, for example, anomalies in the power curve (Morrison et al., 2022), physics-guided clustering for the quality assurance of sensor data (Cui et al., 2026) and inconsistent alarm and maintenance logs (Leahy et al., 2019). Value-  
30 level methods share an assumption: that the file is a faithful, if noisy, record of the sensors, so that a defect shows up as an implausible value. Some defects do not: a signal that is absent for its first two years because it was not yet logged has no implausible values; a channel whose definition changes partway through the record produces plausible values on both sides of the change; a file in which each year is written many times over, with the measurements in one copy and NaN in the others, has no outliers at all; and a 10-minute energy value that does not match the 10-minute mean power stays within the range  
35 of plausible energies. Such defects belong to the provenance of the file, the chain of logging, storage, export and packaging that separates the sensor from the reader. Clifton et al. (2023) name the creation of FAIR data frameworks as the first of three grand challenges in the digitalisation of wind energy, and Barber and Ding (2024) discuss the practice of data sharing through a power curve benchmarking case study that cites the Kelmarsh and Penmanshiel releases; the FAIR principles (Wilkinson et al., 2016) ask for provenance to be documented, but they do not say how a user should verify it.

40 This paper defines a set of nine provenance checks that require only the files themselves and reports their results for every file of the Kelmarsh v4 release, with each hypothesis, tolerance and decision rule fixed before the confirmatory analysis was run (Appendix A). It then provides a reconstruction of the defective 2023 and 2024 files, reproduced on the companion Penmanshiel release, a flag table for the whole record, and a quantification of what the defects do to three ordinary downstream analyses (annual capacity factor, year-to-year change in the binned power curve, and missing-data rates) when they are not  
45 corrected. The audit found defects of four kinds: a staircase of signal onsets, a redefinition of the primary wind speed channel, a paginated export that inflates two years of files 41-fold, and a set of energy-export, scale and sentinel artefacts. One pre-specified prediction failed by its decision rule, seven others failed as worded and one could not be evaluated as specified, and we report each as such. The paper is written for users and maintainers of these and similar releases, and for reviewers of work that depends on them.

## 50 2 Data

### 2.1 The Kelmarsh v4 release

Kelmarsh wind farm (UK, 52.40° N, 0.94° W) consists of six Senvion MM92 turbines of 2050 kW rated power and 92 m rotor diameter; turbines 1, 2, 4 and 5 have a hub height of 78.5 m and turbines 3 and 6 of 68.5 m (static table of the release). Version 4 of the release (Plumley and Takeuchi, 2025a; published 12 August 2025, CC BY 4.0) contains, per calendar year 2016 to  
55 2024, a zip archive with one `Turbine_Data` CSV file and one status-log file per turbine, the former at 10-minute resolution,



**Table 1.** Inventory of the 54 Turbine\_Data files of the Kelmarsh v4 release: files per year, columns, rows per file as delivered, unique timestamps, file size and first timestamp.

| Year | Files | Columns (incl. timestamp) | Rows per file | Unique timestamps | File size (MB) | First timestamp |
|------|-------|---------------------------|---------------|-------------------|----------------|-----------------|
| 2016 | 6     | 299                       | 52,416        | 52,416            | 90 to 91       | 2016-01-03      |
| 2017 | 6     | 299                       | 52,560        | 52,560            | 112            | 2017-01-01      |
| 2018 | 6     | 299                       | 52,560        | 52,560            | 141 to 142     | 2018-01-01      |
| 2019 | 6     | 299                       | 52,560        | 52,560            | 147 to 151     | 2019-01-01      |
| 2020 | 6     | 299                       | 52,704        | 52,704            | 204 to 206     | 2020-01-01      |
| 2021 | 6     | 303                       | 52,560        | 52,560            | 206 to 208     | 2021-01-01      |
| 2022 | 6     | 303                       | 52,560        | 52,560            | 210 to 212     | 2022-01-01      |
| 2023 | 6     | 312                       | 2,174,760     | 52,560            | 2895 to 2941   | 2023-01-01      |
| 2024 | 6     | 312                       | 2,174,904     | 52,704            | 2894 to 2904   | 2024-01-01      |

together with a static turbine table, a signal mapping table, phasor measurement unit data and grid meter data. This paper concerns the 54 Turbine\_Data files. Each file was exported from the owner's secondary SCADA system (Greenbyte); the file header records the export date, which is 27 or 28 January 2022 for the 2016 to 2020 files, 15 August 2023 for the 2021 and 2022 files, and 11 August 2025 for the 2023 and 2024 files. The Zenodo record carries one caveat, "not all signals are available for the entire period", and asks users to tell the owner where, how and why the data are used.

Table 1 gives the inventory. The 2016 file begins on 3 January 2016 and the six turbines record their first power values between 21 January and 5 February 2016. The number of data columns after the timestamp is 298 for 2016 to 2020, 302 for 2021 and 2022, and 311 for 2023 and 2024; the union over all files has 323 column names, of which 12 are the same acceleration signals under two unit spellings ( $\text{mm/s s}$  until 2022,  $\text{mm/s}^2$  from 2023), leaving 311 distinct signals. Four columns contain no value in any file, so 319 columns (307 distinct signals) carry values. The header of each 2016 file states the interval 1 January 2016 to 1 January 2017, but the rows begin on 3 January 2016, so 288 of the stated 52,704 steps are absent from the six 2016 files (Table S1); all other files match their stated interval. The columns hold 10-minute statistics of measured signals (mean, minimum, maximum, standard deviation) and quantities derived by the SCADA system (availability ratios, lost production, potential power, counters and budgets). The 2023 and 2024 files are 2.9 GB each against 90 to 212 MB for the earlier years, a first sign of the export defect described in Sect. 4.3.

## 2.2 The Penmanshiel v3 release

Penmanshiel wind farm (14 Senvion MM82 turbines, no turbine WT03) is released in the same format (Plumley and Takeuchi, 2025b; published 13 August 2025). We use it only to test whether the export defect is specific to Kelmarsh, on 13 files: the monthly export for February 2023 (nine turbines) and the full-year 2023 and 2024 exports for turbines WT11 and WT15.



## 75 2.3 Reading conventions

All files were read with the same rule: ten header lines, a header row beginning with # , a trailing comma on every data row (so the row must be read without treating its last field as an index), UTC timestamps, and the string NaN for missing or erroneous values as stated in the header. No value was altered or removed in any audit analysis. All analyses use Python with pandas (McKinney, 2010) and NumPy (Harris et al., 2020); scripts, outputs and figures are provided (Code availability).

## 80 3 Audit procedure

### 3.1 Nine checks

Each check is a deterministic test on the file contents with a stated decision rule. Together they target the ways in which a 10-minute SCADA export can misrepresent the sensors without producing implausible values.

- 85 1. **Grid completeness.** Row count equals the number of 10-minute steps in the interval stated in the file header; timestamps are unique, monotonic and parseable. Failure indicates duplication, gaps, a header that does not describe the content, or an export artefact.
2. **Timestamp multiplicity.** For files failing check 1, the multiplicity of each timestamp, the position of restarts (a row whose timestamp precedes its predecessor), the length of the resulting blocks and which occurrence carries values.
3. **Per-signal onset.** For every column, turbine and file, the first and last timestamp with a value and the count of values.  
 90 Onsets that coincide to the 10-minute step across turbines are attributed to the monitoring system; onsets that differ by days are attributed to installation or commissioning.
4. **Redundant-channel identity.** Where a quantity is stored both as a primary channel and as individual sensors, the monthly mean of the difference between the primary channel and each candidate definition (here the mean of the two anemometers, and sensor 2 alone) is compared with a tolerance of  $0.02 \text{ m s}^{-1}$ , well below the smallest inter-sensor offset  
 95 of  $0.18 \text{ m s}^{-1}$ .
5. **Scale ratio.** For a quantity stored on two scales, the ratio of the channels (here `Grid voltage` divided by the mean of the three phase voltages) is classified as near 1, near  $\sqrt{3}$  or other, per interval.
6. **Sentinel scan.** Exact and neighbouring occurrences of physically impossible values (negative wind speed, temperature at or near  $-273.2 \text{ }^{\circ}\text{C}$ , rotor speed above 20 RPM, pitch above  $180^{\circ}$ , power outside  $-100$  to  $2300 \text{ kW}$ , wind speed above  
 100  $40 \text{ m s}^{-1}$ ).
7. **Counter consistency.** The 10-minute energy export is compared with mean power times one sixth of an hour, per interval, per complete day and per year.



8. **Constant-column classification.** Columns that are constant, or constant within a year, are separated from measured signals (used in Sect. 4.1 to separate constant and flag-like columns from measured signals).

105 9. **Cross-release replication.** Checks 1 and 2 applied to a second release from the same export system.

### 3.2 Pre-specified hypotheses and statistical conventions

Six hypotheses (H1 to H6) with predictions, tolerances, robustness variants and disconfirmation rules were written down before any confirmatory analysis was run, on the basis of exploratory work on the 2016 to 2022 files and on seven of the twelve 2023 and 2024 files (the 2023 files of turbines 1 to 4 in full, and the block structure of the 2024 files of turbines 1 to 3); the plan is reproduced in Appendix A with the outcome of each prediction. The held-out material, as listed in the plan, was the 2023 T5, T6 and 2024 T4 to T6 files for H3, turbines 2 to 6 for the H2 findings first seen on turbine 1, and the years 2023 and 2024 for the H2a, H4 and H5 windows; H1 and H6 were evaluated on the full record without held-out material. Every departure from the plan is recorded in a deviations log that accompanies the archived code (see Code availability) and is summarised in Sect. 4.6.

115 All statistics are identities, counts or effect sizes. No null-hypothesis tests are used. Uncertainty for the binned-power steps is an i.i.d. bootstrap within bins (1000 replicates for the main estimates and 300 for the robustness variants of Appendix C, seed 20260915, percentile intervals); uncertainty for statistics of the 10-minute series is a moving-block bootstrap with daily blocks, checked against 7-day blocks; the 7-day intervals are reported (they are the wider ones for the exceedance share in every year, and for the annual difference their width differs from the 1-day intervals by at most 0.03 percentage points). Intervals are unadjusted for multiplicity. Robustness variants (tolerances of 0.01, 0.02 and 0.05  $\text{m s}^{-1}$ ; bin centres 7, 8 and 9  $\text{m s}^{-1}$  with widths 0.5 and 1.0  $\text{m s}^{-1}$ ; three operating-state filters; exclusion of turbine-years with logged anemometer defects; a second reference window for the power step) were pre-specified and are reported in Appendix C; two departures from the plan's wording (the second reference window and the bin centres) are noted in Appendix C, and a third (the comparison window of H6(ii)) in Appendix A. The planned test of the assumption that the three wind speed channels share a time base was run: the cross-correlation between the primary channel and each sensor peaks at zero lag on every turbine ( $r = 0.995$  to  $0.998$ , against 0.961 to 0.971 one step either side).

## 4 Results

### 4.1 Signals become available in a staircase of system-wide onsets (H1)

Of the 319 columns with at least one value, 74 have their first value in 2016 before May: five at the start of the file on all turbines, 64 at turbine-specific times in January and February spread over 14 to 22 days, consistent with commissioning, four with spreads of about one day, and one derived column (Potential power learned PC) that starts in February 2016 on five turbines and in December 2016 on turbine 6. The remaining 245 columns begin on twelve later dates or short date ranges (Table 2, Fig. 1). For 242 of the 245 (98.8 %) the first valid timestamp is identical on all six turbines to the 10-minute step; the

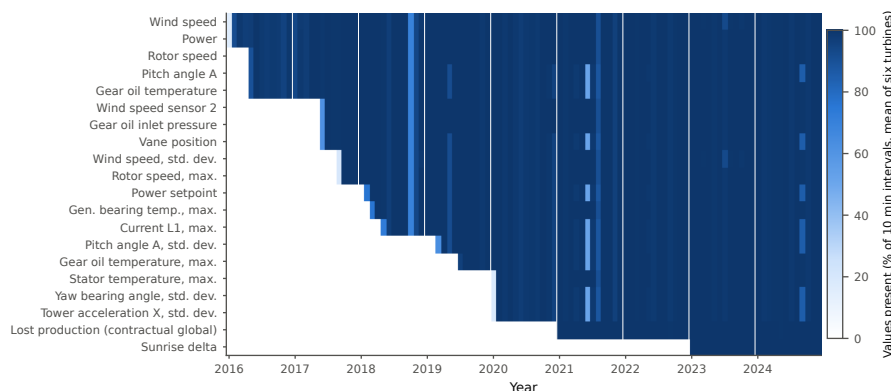


**Figure 1.** Share of 10-minute intervals with a value, by month and signal, mean of the six turbines, for 20 representative columns of the 307 distinct signals with values in the release. Rows are ordered by onset date. The white vertical lines mark calendar years.

three exceptions are derived performance and contractual metrics (Operating Performance Ratio, 20 minutes apart; MTBF and MTTR (Contractual Global), 5.7 days apart) whose first value depends on a turbine-specific event. The 5.7-day spread of the last two exceeds the one-day threshold of the plan's disconfirmation clause, but the clause is stated for a group as a whole, and their group, the 2023 file-schema group, meets the prediction with 19 of 21 columns synchronous; we therefore do not count them as a disconfirmation. The plan phrased the prediction as columns sharing their group's onset to within one step; four monitoring-system groups turned out to span 1.8 to 7.0 days with several sub-onsets (and the 2023 file-schema group 1.6 days), so the prediction was applied as synchrony of each column across the six turbines. The 245 columns correspond to 233 distinct signals, because the 12 renamed acceleration columns appear a second time in 2023; the 311 distinct signals of Sect. 2.1 include the four that never carry a value. The largest groups are 32 operational signals on 2016-05-03 09:40, 15 signals (individual anemometers, vane, motor currents, oil pressures) on 2017-06-12 09:00, 39 ten-minute statistics on 2017-09-25 00:00, 22 statistics between 2019-06-27 and 2019-07-04, and 81 statistics between 2020-01-24 and 2020-01-26. Two further groups coincide with a change of file schema (nine columns from 2021-01-01, 21 from 2023-01-01, of which 12 are the renamed acceleration columns). The plan named seven of the twelve groups; the 2016-08-22, 2018-03-06 to 03-08 and 2018-05-09 groups and the two file-schema groups appeared in the confirmatory run. The pre-specified prediction (at least 90 % of columns in each group sharing the onset to one step on all six turbines) holds for all ten monitoring-system groups (100 %) and for the 2023 file-schema group (19 of 21); the 2021 file-schema group is at 8 of 9 (88.9 %), below the threshold, because one derived ratio first takes a value 20 minutes later on turbine 5. Check 8 adds that 10 of the 319 columns hold a single value (zero) in every file (nine lost-production-to-curtailment categories and compensated lost production) and four are binary or ternary flags (Table S2).

We read the onsets as properties of the monitoring system's logging configuration, because a sensor installed on six machines would not begin logging on all six within the same 10-minute interval. The practical consequence is that the record is a staircase, and any missing-data statistic must be computed from each signal's onset date (Sect. 4.5).



**Table 2.** Post-commissioning signal onset groups: onset date or date range (UTC), number of columns, number of columns whose first valid timestamp is identical on all six turbines, and example columns.

| Onset (UTC)                           | Columns | Synchronous<br>across the six<br>turbines | Examples                                                                                                                                                |
|---------------------------------------|---------|-------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2016-05-03 09:40                      | 32      | 32 of 32                                  | Ambient temperature (converter); Blade angle (pitch position) A;<br>Blade angle (pitch position) B                                                      |
| 2016-08-22 07:50                      | 1       | 1 of 1                                    | Potential power secondary reference turbines                                                                                                            |
| 2017-06-12 09:00                      | 15      | 15 of 15                                  | CPU temperature; Gear oil inlet pressure; Gear oil pump pressure                                                                                        |
| 2017-09-25 00:00                      | 39      | 39 of 39                                  | Front bearing temperature, Max; Front bearing temperature, Min;<br>Front bearing temperature, Standard deviation                                        |
| 2018-02-07 16:50                      | 1       | 1 of 1                                    | Turbine Power setpoint                                                                                                                                  |
| 2018-03-06 to 03-08                   | 7       | 7 of 7                                    | Generator bearing front temperature, Max; Generator bearing<br>front temperature, Min; Generator bearing rear temperature, Max                          |
| 2018-05-09 10:00                      | 3       | 3 of 3                                    | Current L1 / U, StdDev; Current L1 / U, max; Current L1 / U, min                                                                                        |
| 2019-03-12 to 03-14                   | 14      | 14 of 14                                  | Blade angle (pitch position) A, Max; Blade angle (pitch position)<br>A, Min; Blade angle (pitch position) A, Standard deviation                         |
| 2019-06-27 to 07-04                   | 22      | 22 of 22                                  | Current L2 / V, StdDev; Current L2 / V, max; Current L2 / V, min                                                                                        |
| 2020-01-24 to 01-26                   | 81      | 81 of 81                                  | Ambient temperature (converter), Max; Ambient temperature<br>(converter), Min; Ambient temperature (converter), StdDev                                  |
| 2021-01-01 (new file schema)          | 9       | 8 of 9                                    | Investment Performance Ratio; Lost Production (Contractual Cus-<br>tom); Lost Production (Contractual Global)                                           |
| 2023-01-01 to 01-02 (new file schema) | 21      | 19 of 21                                  | Available Capacity for Production (Planning deviation) (%);<br>Available Capacity for Production (Planning deviation) (kW);<br>Drive train acceleration |

## 4.2 The primary wind speed channel changed definition in 2018 (H2)

The files contain three 10-minute wind speed channels: Wind speed, present from 2016, and Wind speed Sensor 1 and Wind speed Sensor 2, present from 2017-06-12. Sensor 1 reads above sensor 2 by 0.18 to 0.50 m s<sup>-1</sup> on average over the complete months July to October 2017, depending on the turbine, so the three channels can be told apart by their monthly means even though their 10-minute values are averaged by different paths and rarely agree to the last digit.

For every turbine and every complete month from June to October 2017 (July to October, since the sensors appear on 12 June), the monthly mean of Wind speed minus the mean of the two sensors is within 0.0043 m s<sup>-1</sup> of zero; for every complete month from May 2018 to December 2024 (78 or 79 months per turbine), Wind speed minus sensor 2 is within 0.0071 m s<sup>-1</sup> of zero (Table 3, Fig. 2a). There are no violations of the 0.02 m s<sup>-1</sup> tolerance, and none at 0.01 m s<sup>-1</sup> either.

The primary channel therefore equalled the mean of the two anemometers from July to October 2017 and has equalled sensor 2 alone since May 2018, on all six turbines (H2a confirmed).



The transition is not a single date. A day-level classification with the same tolerance, adopted after the monthly criterion of the plan proved undefined (Appendix A), shows the last day on which the primary channel equals the sensor mean to be 2018-04-10 on all six turbines, and the first day on which it equals sensor 2 alone to fall between 2017-11-13 and 2017-12-08 depending on the turbine. Between those dates the definition alternates from day to day: of the 124 to 149 days between the first "sensor 2" day and 2018-04-10, 46 to 70 per turbine are classified "mean", 26 to 39 "sensor 2", 2 to 29 "both" (the sensors agree too closely to discriminate) and 11 to 48 neither (Fig. 2a, dashed and solid lines converging). The switch is not explained by a failure of sensor 1: it carries values in every complete month from May 2018 to December 2024 on all six turbines and still reads above sensor 2 by 0.15 to 0.43  $\text{m s}^{-1}$  on average per turbine (0.18 to 0.50  $\text{m s}^{-1}$  from July to October 2017). The synchronous end of the mean definition on 2018-04-10 is consistent with a configuration change in the monitoring system; the alternation before it is an exploratory finding for which we have no explanation from the data.

Because sensor 1 reads high, the redefinition lowers the reported wind speed at a given true wind speed, and the binned power curve on the primary channel shifts upward. The mean power in the 7.5 to 8.5  $\text{m s}^{-1}$  bin of the primary channel (a 1  $\text{m s}^{-1}$  bin, twice the width of the bins of IEC 61400-12-1; IEC, 2022), restricted to intervals with time-based system availability equal to 1 and power above zero, is higher in 2018 to 2022 than in 2017 on every turbine, by 12.1 [7.9, 15.6] kW on turbine 4 to 89.4 [85.2, 93.4] kW on turbine 1, that is by 1.2 to 9.1 % (Table 4c, Fig. 2b, d). The same bin computed on sensor 2 shows steps of  $-6.1 [-11.5, -1.0]$  to  $+15.8 [10.8, 21.1]$  kW, with the interval including zero on four of six turbines (Fig. 2c, d). The sign prediction of H2c (positive on all turbines) holds, and the control prediction (sensor 2 steps within  $\pm 15$  kW of zero) holds on five turbines and fails by 0.8 kW on turbine 1 (15.8 kW).

The quantitative prediction of H2c fails. The step predicted from the local slope of the 2017 power curve at 8  $\text{m s}^{-1}$  (348 to 365 kW per  $\text{m s}^{-1}$ ) times half the sensor offset (0.09 to 0.25  $\text{m s}^{-1}$ ) is 33.6 to 85.5 kW; it lies inside the bootstrap interval of the observed step only for turbine 1. On turbines 2 to 6 the observed step is 9 to 22 kW smaller than predicted. By the pre-specified rule (prediction outside the interval on two or more turbines) this part of H2c is disconfirmed. A plausible reason, not tested here, is that the delta-method prediction assumes the power curve itself did not change between 2017 and 2018 to 2022, whereas the sensor 2 steps show that it did by a few kW in both directions. The robustness variants do not change the picture: with either operating-state filter, at all three bin centres and both widths, and with or without the anemometer-defect years, the primary-channel step is positive on all six turbines (range +7.3 to +116.4 kW, +0.7 to +10.5 %) and the sensor 2 step lies between  $-9.4$  and  $+18.1$  kW. Without any operating-state filter the comparison is confounded by year-to-year differences in downtime; the primary-channel step is then negative on turbine 4 and the sensor 2 step is negative in every variant on turbines 4 to 6 (Appendix C).

A decomposition added after the prediction failed, and therefore exploratory, separates the two contributions. Because the primary channel equals sensor 2 from May 2018, the primary-channel step from June to December 2017 to 2018 to 2022 is, up to sampling differences, the sum of a redefinition term (the 8  $\text{m s}^{-1}$  bin mean on sensor 2 minus that on the primary channel within June to December 2017) and the sensor 2 step. The redefinition term measured this way is 28.4 [21.8, 35.2] kW on turbine 4 to 88.0 [80.5, 95.3] kW on turbine 1 (Table 4c), and the decomposition closes to within 4.2 kW on every turbine. The delta-method prediction lies within 6 kW of this term on five turbines and 11 kW above it on turbine 3. Table 4e splits the

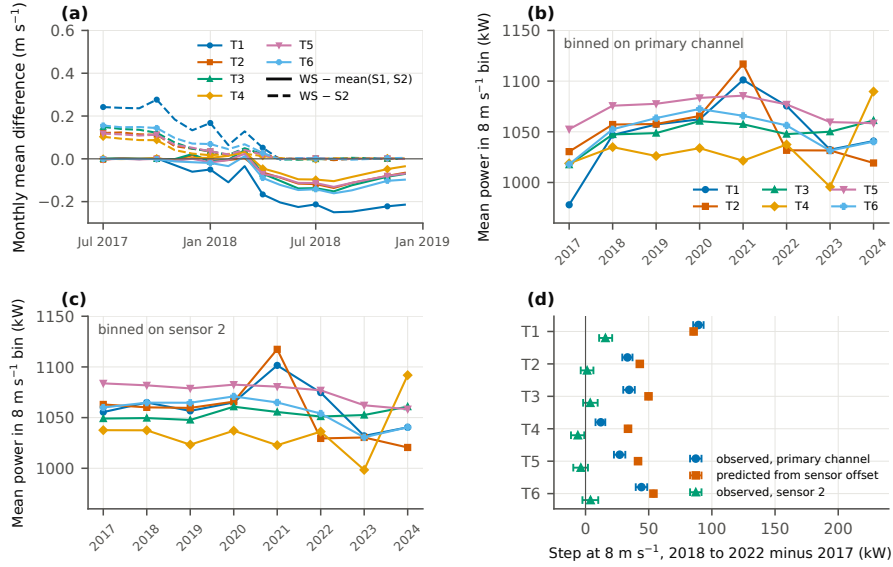


**Figure 2.** Redefinition of the primary wind speed channel. (a) Monthly mean of Wind speed minus the mean of sensors 1 and 2 (solid) and minus sensor 2 (dashed), six turbines, complete months from July 2017 to December 2018; colours and markers identify the turbines as in (b). (b) Mean power in the  $7.5$  to  $8.5 \text{ m s}^{-1}$  bin of the primary channel by year, and (c) the same bin on sensor 2, for intervals with time-based system availability equal to 1 and power above zero. (d) Step in the  $8 \text{ m s}^{-1}$  bin from 2017 to 2018 to 2022 per turbine: observed on the primary channel, predicted from the local power-curve slope times half the sensor offset, and observed on sensor 2; bars are bootstrap 95 % intervals (1000 replicates).

shortfall of the prediction against the observed step on turbines 2 to 6 (9 to 22 kW) into the three contributions: the choice of the full year 2017 rather than June to December as the reference window accounts for 7 to 10 kW on every turbine; the sensor 2 step adds 4 to 6 kW on turbines 4 and 5 and reduces the shortfall on the other turbines; the linearisation adds 1 to 5 kW except on turbine 3, where it is the largest term at 11 kW. Most of the shortfall therefore comes from the reference window written into the plan; with the June to December window the prediction would still fall outside the interval on turbines 1, 3 and 4, so the decision rule would disconfirm it either way.

### 4.3 The 2023 and 2024 files are paginated exports (H3)

Each of the twelve 2023 and 2024 Turbine\_Data files fails check 1 by a wide margin (the six 2016 files fail it by the 288 missing steps of Sect. 2.1, with an otherwise complete grid): a 2023 file has 2,174,760 rows against 52,560 ten-minute steps, and a 2024 file 2,174,904 rows against 52,704. Check 2 explains the excess (Table 5, Fig. 3). Each file consists of 81 blocks. Block  $k$  ( $k = 1$  to 80) contains the first  $655 \cdot k$  rows of the year, starting again at 1 January; block 81 contains the whole year (52,560 or 52,704 rows). The structure is arithmetic:  $655 \times (1 + 2 + \dots + 80) + 52,560 = 2,174,760$ . In each of the first 80 blocks only the last 655 rows are new, and in block 81 only the last 160 or 304; for each timestamp, the first occurrence in



**Table 3.** Identity tests for the primary wind speed channel per turbine. For each test window, the number of complete months, the largest absolute difference between the monthly mean of the primary channel and that of the comparison channel (the mean of sensors 1 and 2 in the first window, sensor 2 alone in the second), and the number of months in which that difference exceeds the  $0.02 \text{ m s}^{-1}$  tolerance; then the first day on which the primary channel equals sensor 2 and the last day on which it equals the sensor mean in the day-level classification.

| Turbine | Complete months, Jun to Oct 2017 | Largest  primary – sensor mean  ( $\text{m s}^{-1}$ ) | Months above $0.02 \text{ m s}^{-1}$ | Complete months, May 2018 to Dec 2024 | Largest  primary – sensor 2  ( $\text{m s}^{-1}$ ) | Months above $0.02 \text{ m s}^{-1}$ | First day primary = sensor 2 | Last day primary = sensor mean |
|---------|----------------------------------|-------------------------------------------------------|--------------------------------------|---------------------------------------|----------------------------------------------------|--------------------------------------|------------------------------|--------------------------------|
| 1       | 4                                | 0.0026                                                | 0                                    | 79                                    | 0.0050                                             | 0                                    | 2017-12-08                   | 2018-04-10                     |
| 2       | 4                                | 0.0043                                                | 0                                    | 79                                    | 0.0064                                             | 0                                    | 2017-11-25                   | 2018-04-10                     |
| 3       | 4                                | 0.0010                                                | 0                                    | 79                                    | 0.0059                                             | 0                                    | 2017-11-25                   | 2018-04-10                     |
| 4       | 4                                | 0.0025                                                | 0                                    | 78                                    | 0.0052                                             | 0                                    | 2017-11-28                   | 2018-04-10                     |
| 5       | 4                                | 0.0031                                                | 0                                    | 79                                    | 0.0071                                             | 0                                    | 2017-11-13                   | 2018-04-10                     |
| 6       | 4                                | 0.0014                                                | 0                                    | 79                                    | 0.0062                                             | 0                                    | 2017-11-25                   | 2018-04-10                     |

**Table 4.** Consequences for downstream analyses.

(a) Annual capacity factor per turbine for 2023 and 2024 from the reconstructed files, from keep-last de-duplication and from averaging all occurrences of each timestamp, with the share of rows in the raw file that carry no power value.

| Year | Turbine | Capacity factor, reconstructed (%) | Capacity factor, keep-last (%) | Capacity factor, mean of occurrences (%) | Share of raw rows without power (%) |
|------|---------|------------------------------------|--------------------------------|------------------------------------------|-------------------------------------|
| 2023 | 1       | 31.1                               | 59.8                           | 31.1                                     | 97.6                                |
| 2023 | 2       | 33.9                               | 64.7                           | 33.9                                     | 97.6                                |
| 2023 | 3       | 29.2                               | 55.4                           | 29.2                                     | 97.6                                |
| 2023 | 4       | 28.2                               | 59.7                           | 28.2                                     | 97.6                                |
| 2023 | 5       | 27.9                               | 59.3                           | 27.9                                     | 97.6                                |
| 2023 | 6       | 25.5                               | 50.1                           | 25.5                                     | 97.6                                |
| 2024 | 1       | 29.6                               | 79.6                           | 29.6                                     | 97.6                                |
| 2024 | 2       | 32.9                               | 81.9                           | 32.9                                     | 97.6                                |
| 2024 | 3       | 27.5                               | 77.0                           | 27.5                                     | 97.6                                |
| 2024 | 4       | 28.3                               | 81.0                           | 28.3                                     | 97.6                                |
| 2024 | 5       | 27.8                               | 79.9                           | 27.8                                     | 97.6                                |
| 2024 | 6       | 23.4                               | 67.8                           | 23.4                                     | 97.6                                |

215 the file carries the measurements and every later occurrence is NaN in all measured columns. Nine derived columns (eight  
 production-based availability ratios and Equivalent Full Load Hours) are the exception: their later copies carry the  
 first occurrence's value (or its NaN), as verified row by row for the 2023 file of turbine 1 and by value counts for all twelve files,  
 so they survive any reading strategy; a column-by-column pass over all 2,122,200 later occurrences in each file finds values  
 in exactly these nine columns and in none of the other 302. This holds for all twelve files, including the five that were held  
 220 out when the hypothesis was written: in each, the number of later occurrences with a power value is zero, the number of first



**Table 4.** Continued.

(b) Missing-data share in the onset year per onset group, computed over all rows and from the onset date, and the overstatement.

| Onset group (date, UTC)  | Columns | Missing share over the whole year (%) | Missing share from the onset date (%) | Overstatement (percentage points) |
|--------------------------|---------|---------------------------------------|---------------------------------------|-----------------------------------|
| 2016-05-03 09:40         | 32      | 34.8                                  | 2.1                                   | 32.6                              |
| 2016-08-22 07:50         | 1       | 64.1                                  | 0.9                                   | 63.3                              |
| 2017-06-12 09:00         | 15      | 44.7                                  | 0.4                                   | 44.3                              |
| 2017-09-25 00:00         | 39      | 73.2                                  | 0.2                                   | 73.0                              |
| 2018-02-07 16:50         | 1       | 13.8                                  | 3.9                                   | 9.9                               |
| 2018-03-06/08            | 7       | 21.6                                  | 4.2                                   | 17.4                              |
| 2018-05-09 10:00         | 3       | 38.6                                  | 5.3                                   | 33.3                              |
| 2019-03-12/14            | 14      | 20.2                                  | 1.1                                   | 19.2                              |
| 2019-06-27 to 07-04      | 22      | 50.4                                  | 0.2                                   | 50.2                              |
| 2020-01-24/26            | 81      | 8.3                                   | 1.5                                   | 6.8                               |
| 2021-01-01 (file schema) | 9       | 2.3                                   | 2.3                                   | 0.0                               |
| 2023-01-01 (file schema) | 21      | 9.9                                   | 9.9                                   | 0.0                               |

occurrences with a power value is 51,868 to 52,604, and the set of unique timestamps is exactly the 10-minute grid of the year. The signature is that of a paginated export in which every page was written cumulatively from the start of the year, with values only for the rows new to that page; the export dates in the file headers (11 August 2025 for both years, against 2022 and 2023 for the earlier files) are consistent with a change in the export tool or its use, but the files themselves do not say which.

225 The consequence depends on how the file is read (Fig. 3c). Reading the file as it is, 97.6 % of rows have no wind speed or power value. De-duplicating on the timestamp and keeping the last occurrence, a choice made when duplicates are assumed to be later corrections and the opposite of the default of the duplicate-removal function in pandas (McKinney, 2010), retains values for only the 160 (2023) or 304 (2024) timestamps that occur once because they lie beyond the last page boundary at row 52,400: 0.30 % and 0.58 % of the year, all in late December. Sorting by timestamp before keeping the last occurrence gives  
 230 the same result. Averaging the occurrences of each timestamp recovers the year, because the mean ignores NaN. Keeping the first occurrence recovers the year exactly: 52,560 or 52,704 unique timestamps, no duplicates, no gaps, with 98.4 to 99.8 % of intervals carrying a power value, within the 88.9 to 99.9 % range of the earlier years. The reconstruction rule is therefore "keep the first occurrence of each timestamp"; it does not depend on the page length and it preserves the file header (Appendix B).

The same structure exists in the Penmanshiel v3 release (Table 5b). The monthly February 2023 files for nine turbines have  
 235 20,076 rows for 4032 ten-minute steps, in eight cumulative blocks with a 573-row increment; the full-year 2023 and 2024 files for turbines WT11 and WT15 have 2,451,138 and 2,451,282 rows in 92 blocks with the same 573-row increment, an inflation of 46.5 to 46.6. In all 13 files the later occurrences carry no power value and keep-last de-duplication retains 21 (monthly) or 416 to 561 (annual) rows. The page length differs between the two sites (655 against 573 rows) and is constant within a site across monthly and annual exports; we infer that it is a parameter of the export, but the files do not say. The same structure



**Table 4.** Continued.

(c) Step in the  $8 \text{ m s}^{-1}$  bin from 2017 to 2018 to 2022 per turbine on the primary channel and on sensor 2, the delta-method prediction, and the within-2017 redefinition term, with bootstrap 95 % intervals.

| Turbine | Observed step, primary (kW) | Step, primary (%) | Predicted from offset (kW) | Prediction inside observed 95 % interval | Observed step, sensor 2 (kW) | Step, sensor 2 (%) | Redefinition term within 2017 (kW) |
|---------|-----------------------------|-------------------|----------------------------|------------------------------------------|------------------------------|--------------------|------------------------------------|
| 1       | 89.4 [85.2, 93.4]           | 9.1               | 85.5                       | yes                                      | 15.8 [10.8, 21.1]            | 1.5                | 88.0 [80.5, 95.3]                  |
| 2       | 33.2 [29.1, 37.2]           | 3.2               | 43.0                       | no                                       | 1.3 [−3.8, 6.3]              | 0.1                | 40.6 [33.9, 47.4]                  |
| 3       | 34.6 [29.9, 39.0]           | 3.4               | 49.8                       | no                                       | 3.7 [−1.9, 9.6]              | 0.4                | 38.5 [31.2, 46.2]                  |
| 4       | 12.1 [7.9, 15.6]            | 1.2               | 33.6                       | no                                       | −6.1 [−11.5, −1.0]           | −0.6               | 28.4 [21.8, 35.2]                  |
| 5       | 27.0 [22.5, 31.5]           | 2.6               | 41.6                       | no                                       | −3.7 [−9.7, 1.9]             | −0.3               | 40.5 [33.1, 47.7]                  |
| 6       | 44.3 [39.7, 48.8]           | 4.3               | 53.7                       | no                                       | 3.9 [−2.4, 10.1]             | 0.4                | 50.9 [42.8, 59.2]                  |

(d) Energy export statistics per year: number of 10-minute intervals with an energy export value, share of those intervals in which the energy export exceeds the rated 10-minute energy, the smallest and largest standard deviation among turbines of the residual between energy export and power/6, and the annual difference between summed energy export and summed power  $\times 1/6$  h, with 95 % moving-block bootstrap intervals (7-day blocks).

| Year | Intervals with energy export | Intervals above rated energy (%) | 95 % interval of the share, 7-day blocks (%) | Smallest residual SD among turbines (kWh) | Largest residual SD among turbines (kWh) | Annual energy export minus power $\times 1/6$ h (%) | 95 % interval of the difference, 7-day blocks (%) |
|------|------------------------------|----------------------------------|----------------------------------------------|-------------------------------------------|------------------------------------------|-----------------------------------------------------|---------------------------------------------------|
| 2016 | 290,139                      | 1.29                             | [0.80, 1.88]                                 | 16.3                                      | 21.1                                     | 0.055                                               | [0.030, 0.077]                                    |
| 2017 | 315,360                      | 2.93                             | [1.94, 3.73]                                 | 19.1                                      | 23.8                                     | 0.018                                               | [−0.001, 0.044]                                   |
| 2018 | 315,360                      | 3.12                             | [2.08, 4.03]                                 | 19.1                                      | 22.8                                     | 0.048                                               | [0.023, 0.079]                                    |
| 2019 | 315,359                      | 2.92                             | [2.05, 4.12]                                 | 18.6                                      | 22.7                                     | 0.048                                               | [0.030, 0.071]                                    |
| 2020 | 316,097                      | 4.75                             | [3.48, 6.27]                                 | 21.9                                      | 26.0                                     | 0.037                                               | [0.018, 0.056]                                    |
| 2021 | 315,082                      | 3.17                             | [2.06, 4.29]                                 | 17.1                                      | 22.3                                     | 0.073                                               | [0.047, 0.111]                                    |
| 2022 | 314,563                      | 2.60                             | [1.51, 3.46]                                 | 18.4                                      | 22.4                                     | 0.075                                               | [0.050, 0.118]                                    |
| 2023 | 315,360                      | 3.20                             | [1.96, 4.42]                                 | 19.4                                      | 23.4                                     | 0.052                                               | [0.034, 0.079]                                    |
| 2024 | 315,958                      | 3.03                             | [1.92, 4.12]                                 | 18.5                                      | 23.2                                     | 0.057                                               | [0.037, 0.084]                                    |

(e) Decomposition of the shortfall of the H2c prediction against the observed step into reference-window, sensor 2, linearisation and residual terms, in kW.

| Turbine | Shortfall, prediction minus observed (kW) | Reference-window term (kW) | Sensor 2 term (kW) | Linearisation term (kW) | Residual (kW) |
|---------|-------------------------------------------|----------------------------|--------------------|-------------------------|---------------|
| 1       | −3.8                                      | 10.2                       | −15.8              | −2.5                    | 4.2           |
| 2       | 9.8                                       | 7.7                        | −1.3               | 2.3                     | 0.9           |
| 3       | 15.3                                      | 7.1                        | −3.7               | 11.3                    | 0.6           |
| 4       | 21.5                                      | 9.5                        | 6.1                | 5.2                     | 0.8           |
| 5       | 14.5                                      | 9.6                        | 3.7                | 1.0                     | 0.2           |
| 6       | 9.5                                       | 8.9                        | −3.9               | 2.9                     | 1.6           |

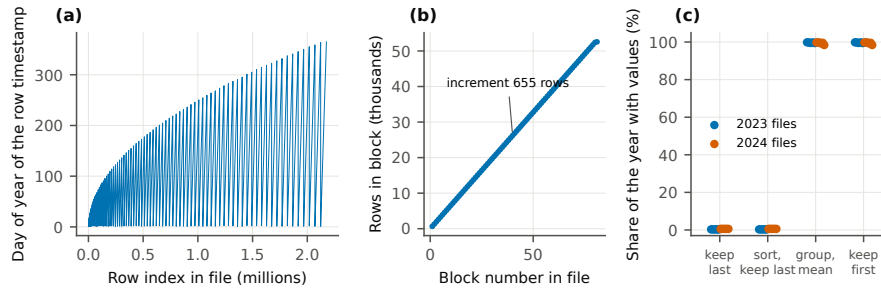


**Figure 3.** The paginated export. (a) Day of year of each row against row index for the 2023 file of turbine 1 (every 200th row), showing 81 cumulative blocks. (b) Rows per block against block number for the same file; the increment is 655 rows. (c) Share of the year with a power value under four reading strategies for the twelve 2023 and 2024 files.

240 would arise if each file were assembled from 81 cumulative partial exports; either way it belongs to the export and not to the measurements, and the reconstruction rule does not depend on which.

#### 4.4 Energy-export jitter, a scale switch and sentinel values (H4, H5)

**Energy export.** The 10-minute `Energy Export` column exceeds the rated interval energy of 341.7 kWh (2050 kW for one sixth of an hour) in 1.3 % of intervals in 2016 and 2.6 to 4.8 % in each year from 2017 to 2024 (per turbine-year 0.6 to 6.0 %; Table 4d, Fig. 4b), with values up to 615 kWh. Part of this is physical, because the 10-minute mean power itself exceeds 2050 kW in 1.1 % of intervals (maximum 2089 kW); the exceedance while the mean power is at or below rated is 2.7 % of intervals, and the residual against power exceeds +60 kWh in 3.0 %. The residual against mean power times one sixth of an hour has a standard deviation of 16 to 26 kWh per turbine-year (Fig. 4a). The residual sums out: daily totals agree with power times 24 h within  $\pm 3$  % on 93.7 % of complete days (99.4 % of days with mean power above 50 kW; Fig. 4c) and annual totals agree within 0.075 % per year and 0.171 % per turbine-year, with all 7-day block bootstrap intervals lying between  $-0.001$  and 0.119 %. That the residual cancels over a day suggests a timing difference between the energy and power values rather than a loss of energy; we did not test the mechanism. All four pre-specified predictions of H4 hold. The practical rule is to use `Power` for 10-minute work and `Energy Export` only for sums over a day or longer.

**Grid voltage scale.** Before 2016-05-03 09:40 the `Grid voltage` channel equals  $\sqrt{3}$  times the mean of the three phase voltages in all 80,532 intervals where both are present (ratio between 1.68 and 1.78, line-to-line against line-to-neutral); from that timestamp, which is also the first post-commissioning onset group of Sect. 4.1, the ratio lies between 0.97 and 1.03 (median 1.000) in 2,681,619 of 2,681,627 intervals (99.9997 %; Table 6a). The ratio test cannot see a second artefact: from 2019 to 2024 there are 66 episodes (53 of at least one hour), 21,921 intervals in total, in which `Grid voltage` returns to about 699 V while the phase-voltage channels are NaN. The episode levels (median 698.6 V, range 694.4 to 702.5 V) are 1.74 times the median phase L1 voltage after the switch (402.3 V), close to  $\sqrt{3}$ , which is consistent with the line-to-line scale of the period before 2016-05-03 reappearing; because the phase channels are empty in these intervals, the ratio cannot be tested interval by



**Table 5.** Paginated-export statistics.

(a) One row per Kelmarsh 2023 and 2024 file: rows, blocks, block increment, unique timestamps, first and later occurrences with a power value, rows with a power value after keep-last de-duplication, capacity factor under keep-first and keep-last reading, and the number of rows per unique timestamp.

| File             | Rows      | Blocks | Block increment (rows) | Unique times-tamps | First occurrences with power | Later occurrences with power | Rows with power after keep-last | Capacity factor, keep-first (%) | Capacity factor, keep-last (%) | Rows per unique timestamp |
|------------------|-----------|--------|------------------------|--------------------|------------------------------|------------------------------|---------------------------------|---------------------------------|--------------------------------|---------------------------|
| Kelmarsh T1 2023 | 2,174,760 | 81     | 655                    | 52,560             | 52,477                       | 0                            | 160                             | 31.1                            | 59.8                           | 41.38                     |
| Kelmarsh T2 2023 | 2,174,760 | 81     | 655                    | 52,560             | 52,479                       | 0                            | 160                             | 33.9                            | 64.7                           | 41.38                     |
| Kelmarsh T3 2023 | 2,174,760 | 81     | 655                    | 52,560             | 52,306                       | 0                            | 160                             | 29.2                            | 55.4                           | 41.38                     |
| Kelmarsh T4 2023 | 2,174,760 | 81     | 655                    | 52,560             | 52,331                       | 0                            | 160                             | 28.2                            | 59.7                           | 41.38                     |
| Kelmarsh T5 2023 | 2,174,760 | 81     | 655                    | 52,560             | 52,454                       | 0                            | 160                             | 27.9                            | 59.3                           | 41.38                     |
| Kelmarsh T6 2023 | 2,174,760 | 81     | 655                    | 52,560             | 52,261                       | 0                            | 160                             | 25.5                            | 50.1                           | 41.38                     |
| Kelmarsh T1 2024 | 2,174,904 | 81     | 655                    | 52,704             | 52,574                       | 0                            | 304                             | 29.6                            | 79.6                           | 41.27                     |
| Kelmarsh T2 2024 | 2,174,904 | 81     | 655                    | 52,704             | 52,604                       | 0                            | 304                             | 32.9                            | 81.9                           | 41.27                     |
| Kelmarsh T3 2024 | 2,174,904 | 81     | 655                    | 52,704             | 52,575                       | 0                            | 304                             | 27.5                            | 77.0                           | 41.27                     |
| Kelmarsh T4 2024 | 2,174,904 | 81     | 655                    | 52,704             | 52,293                       | 0                            | 304                             | 28.3                            | 81.0                           | 41.27                     |
| Kelmarsh T5 2024 | 2,174,904 | 81     | 655                    | 52,704             | 52,493                       | 0                            | 304                             | 27.8                            | 79.9                           | 41.27                     |
| Kelmarsh T6 2024 | 2,174,904 | 81     | 655                    | 52,704             | 51,868                       | 0                            | 304                             | 23.4                            | 67.8                           | 41.27                     |

(b) The 13 Penmanshiel files summarised by export type: rows, blocks, block increment, unique timestamps, later occurrences with a power value, rows with a power value after keep-last de-duplication, and rows per unique timestamp (ranges over the files). Per-file Penmanshiel values are in the Supplement (Table S3).

| Export type                            | Files | Rows                   | Blocks | Block increment (rows) | Unique timestamps | Later occurrences with power | Rows with power after keep-last | Rows per unique timestamp |
|----------------------------------------|-------|------------------------|--------|------------------------|-------------------|------------------------------|---------------------------------|---------------------------|
| Full year 2023 and 2024, WT11 and WT15 | 4     | 2,451,138 to 2,451,282 | 92     | 573                    | 52,560 to 52,704  | 0                            | 416 to 561                      | 46.51 to 46.64            |
| Monthly, February 2023, nine turbines  | 9     | 20,076                 | 8      | 573                    | 4032              | 0                            | 21                              | 4.98                      |

interval. Most episodes start in the same interval on all six turbines; the largest, 2344 intervals in May 2019, is on turbine 1 alone. These were found by an absolute-level test (Grid voltage above 550 V) added after the ratio test was run (Appendix A). Users of the voltage channel need both tests.

265 **Sentinels.** Wind speed Sensor 2 takes the exact value  $-15.000 \text{ m s}^{-1}$  in 1973 intervals (1573 in 2023, 371 in 2018 and 29 in 2020 to 2022), and in addition 248 values between  $-15$  and  $-14.9$  and 116 other negative values, 2337 negative values in all, of which 1881 of the 1901 in 2023 are on turbine 4 (Table 6b). Transformer temperature and Transformer cell temperature both take the value  $-273.200012 \text{ }^{\circ}\text{C}$ , the single-precision representation of  $-273.2$ , in the same five

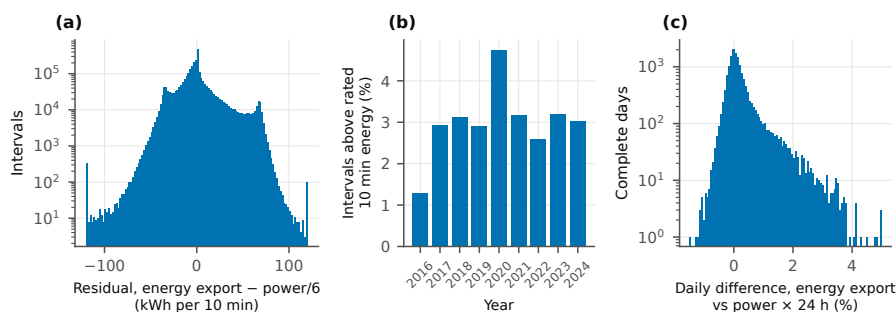


**Figure 4.** Energy export against power. (a) Distribution of Energy Export minus Power/6 per 10-minute interval, all turbine-years, logarithmic count axis; values beyond  $\pm 120$  kWh are accumulated in the end bins. (b) Share of intervals in which Energy Export exceeds the rated interval energy of 341.7 kWh, by year, six turbines pooled. (c) Distribution of the daily relative difference between the energy sum and power times 24 h over complete days with mean power above 50 kW.

intervals, and  $-101.2$  and  $-112.1$  °C in the one interval preceding them on turbine 1. The pre-specified prediction that sentinels occur only as exact values fails; the neighbouring values are consistent with 10-minute averaging of a sentinel with valid samples, an interpretation we did not test. Rotor speed above 20 RPM (up to 49.9 RPM) occurs in 24 intervals on turbine 1 in 2016 and 2017, and pitch angle A above  $180^\circ$  in 70 intervals in 2016 to 2018; wind speed above  $40 \text{ m s}^{-1}$ , power below  $-100$  kW and power above 2300 kW never occur. The flag table (Data availability) marks all of these by physical range rather than by exact value, so the interpretation of the neighbours does not affect it. The range rules of the flag table also surface values that the pre-specified sentinel list did not anticipate (Table 7): a plateau of  $323.7$  to  $324.4$  °C in four drivetrain temperature channels (front and rear bearing, gear oil inlet, gear oil) on turbine 4 for 1598 flagged intervals from 7 to 19 July 2023 (1597 at the plateau, one transitional value) while the turbine was stopped, the same level in the rear generator bearing temperature of turbine 2 for 134 flagged intervals (133 at the plateau) on 28 and 29 November 2022, a  $-113.3$  °C converter ambient temperature in the interval preceding the  $-273.2$  °C values on turbine 1, and a constant blade C pitch of  $105.1^\circ$  during stops (turbine 6 for 1013 intervals from 12 to 19 May 2022, turbine 4 for 152 intervals in May 2018). Episode details are listed in the archived outputs (see Code availability). We report these as flagged values without interpretation.

#### 4.5 Consequences for three ordinary analyses (H6)

**Annual capacity factor.** With the reconstructed files, the annual capacity factor per turbine is 20.2 to 38.4 % over 2016 to 2024 (farm mean 25.1 % in the partial year 2016, 26.2 to 33.9 % thereafter). Read with keep-last de-duplication, the 2023 files give 50.1 to 64.7 % and the 2024 files 67.8 to 81.9 %, that is 24.6 to 31.6 and 44.4 to 52.7 percentage points too high (Table 4a, Fig. 5a), because the few hundred surviving intervals fall in the last two days of December. The pre-specified prediction was that the naive value would be undefined or below 1 %; it is instead defined and too high by 25 to 53 percentage points; the plan's prediction of the direction of the error was wrong. Averaging the occurrences gives the reconstructed value.



**Table 6.** Scale and sentinel checks.

(a) Ratio of the grid voltage to the mean of the three phase voltages, before and from 2016-05-03 09:40: number of 10-minute intervals in which both channels are present, number of intervals in which the ratio lies near  $\sqrt{3}$  (1.68 to 1.78), near 1 (0.97 to 1.03) or elsewhere, the share in the band expected for the period, and the median ratio after the switch.

| Period                  | Intervals with both channels | Ratio 1.68 to 1.78 (near $\sqrt{3}$ ) | Ratio 0.97 to 1.03 (near 1) | Ratio elsewhere | Share in the expected band (%) | Median ratio   |
|-------------------------|------------------------------|---------------------------------------|-----------------------------|-----------------|--------------------------------|----------------|
| Before 2016-05-03 09:40 | 80,532                       | 80,532                                | 0                           | 0               | 100                            | not applicable |
| From 2016-05-03 09:40   | 2,681,627                    | 0                                     | 2,681,619                   | 8               | 99.9997                        | 1.000          |

(b) Absolute-level voltage test and sentinel and out-of-range counts: number of 10-minute intervals flagged, years in which they occur, and details of the flagged values.

| Check                                                                                             | Intervals flagged | Years affected     | Detail                                                                                  |
|---------------------------------------------------------------------------------------------------|-------------------|--------------------|-----------------------------------------------------------------------------------------|
| Grid voltage above 550 V after the switch, phase channels absent                                  | 21,921            | 2019 to 2024       | 66 episodes, 53 of at least 1 h; median level 699 V; longest 2344 intervals (turbine 1) |
| Sensor 2 wind speed exactly $-15.000 \text{ m s}^{-1}$                                            | 1973              | 2018, 2020 to 2023 | 1573 in 2023, 371 in 2018 and 29 in 2020 to 2022                                        |
| Sensor 2 wind speed between $-15$ and $-14.9 \text{ m s}^{-1}$                                    | 248               | 2023               | near-sentinel values not caught by an exact test                                        |
| Sensor 2 wind speed negative, other values                                                        | 116               | 2018, 2020 to 2023 | 2337 negative sensor 2 values in all                                                    |
| Transformer temperature at $-273.2$ °C (single-precision value)                                   | 5                 | 2019, 2022         | 6 intervals below $-40$ °C in total; exact $-273.2$ in double precision: 0              |
| Rotor speed above 20 RPM                                                                          | 24                | 2016 to 2017       | turbine 1 only; maximum 49.9 RPM                                                        |
| Blade A pitch angle above $180^\circ$                                                             | 70                | 2016 to 2018       | the flag table maps angles above $180^\circ$ to negative values before its range test   |
| Wind speed above $40 \text{ m s}^{-1}$ ; power below $-100 \text{ kW}$ or above $2300 \text{ kW}$ | 0                 | none               | none                                                                                    |

**Year-to-year change in the binned power curve.** On the primary channel the  $8 \text{ m s}^{-1}$  bin rises by 1.2 to 9.1 % from 2017 to 2018 to 2022 on all six turbines; on sensor 2 the change is between  $-0.6$  and  $+1.5$  % (Table 4c). A multi-year performance study that uses the primary channel, as a user might without the identity test of Sect. 4.2, would read the redefinition as a performance change of the order of the ageing effects that fleet studies set out to measure (Staffell and Green, 2014; Hamilton

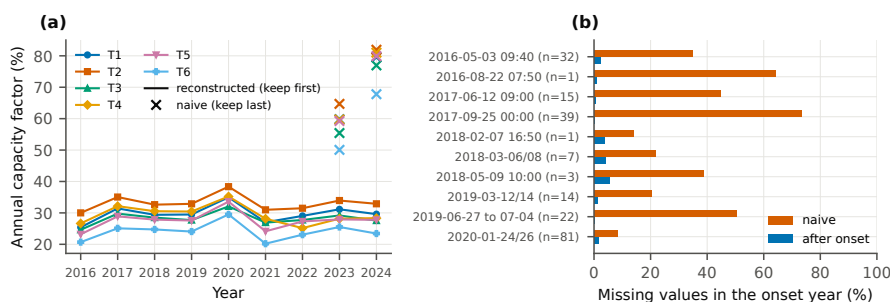


**Figure 5.** Consequences for two ordinary analyses. (a) Annual capacity factor per turbine from the reconstructed files (lines) and from keep-last de-duplication of the raw 2023 and 2024 files (crosses). (b) Missing-data share in the onset year for the ten monitoring-system onset groups, computed over all rows (naive) and from the onset date.

et al., 2020), and it would add to the turbine-specific history, such as control upgrades, that long-term single-site studies have to separate from performance trends (Astolfi et al., 2022).

295 **Missing-data rates.** For the ten monitoring-system onset groups of Table 2, the share of NaN in the onset year computed over all rows is 8.3 to 73.2 %, whereas the share computed from the onset date is 0.2 to 5.3 %; the overstatement is 6.8 to 73.0 percentage points (Table 4b, Fig. 5b). For the two file-schema groups the two rates coincide by construction (0 percentage points), which the plan's wording ("every signal group") failed to exclude. A user selecting signals by completeness would discard, for example, the 39 ten-minute statistics that begin on 2017-09-25 as 73 % missing in 2017, when they are 0.2 %

300 missing after their onset.

#### 4.6 Predictions that failed, and exploratory findings

Of the pre-specified predictions (Appendix A), one failed by its decision rule (the magnitude of the power step, H2c, outside the interval on five of six turbines); four failed as worded (the exactness of the sentinel values, H5; the direction of the capacity-factor error, H6; the statement that later occurrences are NaN in every column, H3, which holds for 302 columns and fails for

305 nine derived columns; and the statement that every onset group shows a missingness overstatement, H6, which fails for the two file-schema groups); and three fell outside a stated bound by a small margin on a single turbine or group (the 2021 file-schema onset group at 88.9 % against a stated 90 %, H1; the sensor 2 control at 15.8 kW against a stated  $\pm 15$  kW on turbine 1; and the 9.14 % step against a stated upper bound of 9 % on turbine 1). None of the eight changes the finding it was attached to, and each is reported above with its original wording. A ninth prediction, the H2b transition month, could not be evaluated as

310 specified: the plan's monthly criterion leaves the last month of the mean definition undefined where several months satisfy both identities, so we report that half as not evaluable and the other half (the first month of the sensor 2 definition) as confirmed; the day-level criterion used instead was chosen after the data were seen and is exploratory. Eight findings were not anticipated and are labelled exploratory: the five-month alternation of the wind speed definition before 10 April 2018; the  $\sim 700$  V episodes invisible to the ratio test; the concentration of the  $-15 \text{ m s}^{-1}$  sentinel on turbine 4 in 2023; the onset groups that the plan's list



**Table 7.** Number of 10-minute intervals flagged by each rule of the flag table, per year, summed over the six turbines; the first row gives the number of intervals in the year. All thresholds were fixed before the analysis. The voltage-scale flag uses the ratio of the grid voltage to the phase L1 voltage (the reduced data hold one phase), which flags 80,533 intervals, one more than the three-phase ratio test of Table 6a. Counts per turbine-year are given in the Supplement (Table S4).

| Flag                                                                                      | 2016    | 2017    | 2018    | 2019    | 2020    | 2021    | 2022    | 2023    | 2024    | All years |
|-------------------------------------------------------------------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|-----------|
| Intervals in the year (six turbines)                                                      | 314,496 | 315,360 | 315,360 | 315,360 | 316,224 | 315,360 | 315,360 | 315,360 | 316,224 | 2,839,104 |
| Sensor 2 wind speed exactly $-15 \text{ m s}^{-1}$                                        | 0       | 0       | 371     | 0       | 2       | 1       | 26      | 1573    | 0       | 1973      |
| Sensor 2 wind speed below $0 \text{ m s}^{-1}$                                            | 0       | 0       | 373     | 0       | 3       | 9       | 51      | 1901    | 0       | 2337      |
| Sensor 1 wind speed below $0 \text{ m s}^{-1}$                                            | 0       | 0       | 3       | 0       | 2       | 1       | 0       | 1       | 0       | 7         |
| Wind speed below 0 or above $40 \text{ m s}^{-1}$                                         | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0         |
| Power below $-100$ or above 2300 kW                                                       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0         |
| Rotor speed below 0 or above 20 RPM                                                       | 13      | 11      | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 24        |
| Blade pitch outside $-10^\circ$ to $100^\circ$                                            | 112     | 168     | 182     | 0       | 0       | 13      | 1028    | 5       | 0       | 1508      |
| Temperature outside $-40$ to $150^\circ \text{C}$<br>(transformer: $200^\circ \text{C}$ ) | 0       | 0       | 0       | 3       | 0       | 0       | 137     | 1598    | 0       | 1738      |
| Transformer temperature at $-273.2^\circ \text{C}$                                        | 0       | 0       | 0       | 3       | 0       | 0       | 2       | 0       | 0       | 5         |
| Grid voltage on the $\sqrt{3}$ scale (before<br>2016-05-03)                               | 80,533  | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 80,533    |
| Grid voltage above 550 V (from 2016-05-03)                                                | 0       | 0       | 0       | 2344    | 1524    | 14,061  | 258     | 0       | 3734    | 21,921    |
| Row from a reconstructed 2023 or 2024 file                                                | 0       | 0       | 0       | 0       | 0       | 0       | 0       | 315,360 | 316,224 | 631,584   |
| Any flag except the reconstructed-file flag                                               | 80,658  | 179     | 555     | 2347    | 1527    | 14,083  | 1474    | 1930    | 3734    | 106,487   |

315 did not name (Sect. 4.1); the renamed acceleration columns counted as 2023 onsets; the  $324^\circ \text{C}$  temperature plateau and the other range-flagged values of Sect. 4.4; the 288 steps absent from the 2016 files against their header; and the decomposition of the power step in Sect. 4.2, which was computed after the H2c prediction failed. We did not revise any hypothesis after seeing the data; the changes of operationalisation (cross-turbine synchrony for H1, the H2b transition criterion and the H6(ii) comparison window) are disclosed in Sect. 4.1 and Appendix A.

## 320 5 Discussion

### 5.1 What the audit adds to value-level cleaning

Three of the four defect classes pass through an outlier filter untouched. The onset staircase produces no values at all before the onset, and the redefinition produces plausible wind speeds throughout; the paginated export produces either complete rows or NaN rows. Of the fourth class, the sentinels are caught by a range filter, but the scale switch and the energy-export jitter  
 325 are within the range of plausible values except in the tails. The cleaning and imputation methods applied to this dataset in the works we read (Martín-Calzada et al., 2025; Hammer and Barber, 2025), and value-level quality-assurance methods such as that of Cui et al. (2026), operate on values and would not be expected to detect any of them, and diagnostics of yaw alignment and anemometer errors from SCADA data (Mittelmeier and Kühn, 2018; Astolfi et al., 2023; Chen et al., 2026) rely on, or diagnose errors in, the nacelle wind speed signal and would be affected by the redefinition if they used the primary channel



330 across 2018. The audit complements cleaning: after reconstruction the 2023 and 2024 files still contain the sentinels and jitter of the other years, and those still need a range filter.

## 5.2 Generality

The checks use nothing but the files. Checks 1, 2, 3 and 8 apply to any tabular time series export; checks 4, 5 and 7 apply wherever a quantity is stored redundantly (two anemometers, three phases, a counter and a rate), as in this release. The Penmanshiel  
335 replication indicates that the export defect belongs to the export path and is not specific to one site. Because both releases come from the same export system, we expect the paginated signature in any other release made with it in the same way until the export is corrected; whether other public releases are affected we have not tested. The onset staircase and the redefinition are properties of what was logged and would remain in a re-export of the same years.

## 5.3 Recommendations

340 For users of the Kelmarsh and Penmanshiel releases: read the 2023 and 2024 files with a first-occurrence rule or use the reconstructed files; use `Wind speed Sensor 2` as the wind speed basis for any comparison that spans 2018, or restrict the primary channel to before 13 November 2017 or after 10 April 2018; compute missing-data statistics from the onset date of each signal; use `Power` at 10-minute resolution and `Energy Export` for daily and longer sums; apply both the ratio and the absolute-level test to `Grid voltage`; flag negative wind speeds and temperatures below  $-40^{\circ}\text{C}$  by range. For maintainers  
345 of similar releases: publish a per-signal onset table and a changelog of channel definitions with the data, state the export tool and its version in the record, and add a row-count-against-grid check to the release process, which would have caught the 2023 and 2024 files immediately. Several of these are the provenance elements the FAIR principles call for (Wilkinson et al., 2016) and that Clifton et al. (2023) identify as missing in practice.

## 5.4 Limitations

350 The audit is of one release version of one dataset, with two checks replicated on 13 files of a second; the defects are dated to v4 and v3 and may be corrected by the owner. We have no access to the controller or monitoring configuration, so the attribution of the onsets and the redefinition to the monitoring system is an inference from synchrony across turbines. The reconstruction recovers what the export contains, which is a 98.4 to 99.8 % completeness comparable with the earlier years; it cannot recover intervals the source system never logged. The 2017 reference year of the power step combines months in which the definition of  
355 the primary channel cannot be checked (before the two sensors are logged from 12 June), months in which it equals the sensor mean, and the alternation from mid-November; the reference-window term of Table 4e (7 to 10 kW) measures part of this, and the June to December variant of Appendix C compares a partial year with full years without a correction for air density. The sensor 2 control in Sect. 4.2 assumes sensor 2 itself did not drift between 2017 and 2018; the sensor 2 steps are positive on four turbines and reach +15.8 kW on turbine 1 with an interval that excludes zero, which could be drift or a real change, and  
360 we cannot separate the two without an independent wind reference. The consequence estimates are for three analyses chosen



before the data were examined; other analyses will be affected differently. A user who keeps the first occurrence, as pandas does by default, or who averages the occurrences is unaffected by the export defect; the capacity-factor consequence applies to keep-last de-duplication and the 97.6 % of empty rows to any reading of the files as delivered. Finally, the failed magnitude prediction for the power step means that the size of the redefinition artefact is only approximately given by the sensor offset  
 365 (within 12 kW here); it is better measured directly, as in Sect. 4.2.

## 6 Conclusions

Nine reproducible checks on the Kelmarsh v4 release found four classes of structural defect, three of which value-level cleaning cannot detect: 245 columns (233 distinct signals) that begin on twelve system-wide dates or short date ranges, a primary wind speed channel that switched from the mean of two anemometers to one of them between November 2017 and April 2018,  
 370 twelve 2023 and 2024 files that hold their year in 81 cumulative blocks with values only in the first copy of each row, and a set of energy-export, scale and sentinel artefacts. The export defect is reproduced in 13 files of the companion Penmanshiel release and is fixed by keeping the first occurrence of each timestamp. Left uncorrected, the export defect leaves 97.6 % of the 2023 and 2024 rows without a power value and, under a keep-last de-duplication rule, overstates the capacity factors by 25 to 53 percentage points; the defects also present a 1.2 to 9.1 % step in binned power that shrinks to  $-0.6$  to  $+1.5$  % on sensor  
 375 2, and overstate missing-data rates by up to 73 percentage points in a signal's first year. One pre-specified prediction, that the power step could be predicted from the sensor offset, failed by its decision rule on five of six turbines, and seven others failed as worded without changing their findings; the transition-month prediction could not be evaluated as specified. The audit code, the reconstructed files and a flag table are released so that the checks can be rerun on future versions and on other releases from the same export system.

## 380 Appendix A: Pre-specified hypotheses and outcomes

Table A1 lists each pre-specified hypothesis with its prediction, the observed outcome and the resulting status.



**Table A1.** Pre-specified hypotheses, predictions and outcomes (Appendix A). WS, the primary wind speed channel; S1 and S2, the two anemometers; T1 to T6, the six turbines.

| Hypothesis (written before confirmatory analysis) | Pre-specified prediction                                                                                                                                                | Outcome                                                                                                                                                                                                                                                                                                                 | Status                                                                                                                                                                                                                         |
|---------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| H1 onset staircase                                | $\geq 90$ % of columns in each group share the onset to one 10-min step on all six turbines; disconfirmed if spread $> 1$ day on $\geq 2$ turbines                      | 242 of 245 columns synchronous across turbines; 100 % in every monitoring-system group; 19 of 21 in the 2023 file-schema group; 8 of 9 (88.9 %) in the 2021 file-schema group; two columns with 5.7-day spreads in the 2023 group (not counted as a disconfirmation; Sect. 4.1)                                         | Confirmed for eleven of twelve groups; failed as worded for the 2021 file-schema group (88.9 %); operationalised as cross-turbine synchrony                                                                                    |
| H2a identity winds                                | $ WS - \text{mean}(S1, S2)  \leq 0.02 \text{ m s}^{-1}$ every complete month June to October 2017; $ WS - S2  \leq 0.02$ every complete month May 2018 to December 2024 | 0 violations on all turbines; max $0.0043$ and $0.0071 \text{ m s}^{-1}$ ; 0 violations at $0.01 \text{ m s}^{-1}$                                                                                                                                                                                                      | Confirmed                                                                                                                                                                                                                      |
| H2b transition                                    | Same transition month on all turbines within one month                                                                                                                  | Monthly criterion: first month equal to sensor 2 only is 2018-04 (T4, T5) or 2018-05 (others), within one month; the last month equal to the mean only is undefined where up to ten months satisfy both identities. Day-level criterion: last "mean" day 2018-04-10 on all six, first "S2" day 2017-11-13 to 2017-12-08 | Not evaluable as specified for the last month of the mean definition (the monthly criterion is undefined); confirmed for the first month of the sensor 2 definition; the day-level criterion adopted afterwards is exploratory |
| H2c step sign                                     | Step at $8 \text{ m s}^{-1}$ positive on all six turbines                                                                                                               | +12.1 to +89.4 kW, all intervals above zero                                                                                                                                                                                                                                                                             | Confirmed                                                                                                                                                                                                                      |
| H2c step magnitude                                | slope $\times \Delta v$ inside the observed 95 % interval; disconfirmed if outside on $\geq 2$ turbines                                                                 | Inside on T1 only; observed 9 to 22 kW below prediction on T2 to T6                                                                                                                                                                                                                                                     | Disconfirmed                                                                                                                                                                                                                   |
| H2c sensor 2 control                              | Sensor 2 step within $\pm 15$ kW of zero                                                                                                                                | −6.1 to +15.8 kW; T1 at 15.8 kW outside the bound                                                                                                                                                                                                                                                                       | Failed as worded on T1; within the bound on 5 of 6                                                                                                                                                                             |
| H3 export structure                               | All 12 files: cumulative 655-k blocks, 81 blocks, later occurrences NaN in every column, complete grid after keep-first                                                 | Block structure and grid exactly as predicted in all 12, including the 5 held-out files; later occurrences NaN in 302 columns but carrying values in 9 derived columns                                                                                                                                                  | Confirmed for the structure; failed as worded for the nine derived columns                                                                                                                                                     |
| H3 reading strategies                             | keep-last 0.3 %, sort-then-keep-last 0.3 %, group-mean $\approx 100$ %, keep-first 100 %                                                                                | 0.30/0.58 %, same, 98.4 to 99.8 %, 98.4 to 99.8 %                                                                                                                                                                                                                                                                       | Confirmed                                                                                                                                                                                                                      |



**Table A1.** Continued.

| Hypothesis (written before confirmatory analysis) | Pre-specified prediction                                                                                                                                     | Outcome                                                                                                                                                                                                                                     | Status                                                                                          |
|---------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| H4 energy export                                  | > 1 % of intervals above rated in every year 2017 to 2024; residual SD > 15 kWh; daily within $\pm 3$ % on $\geq 90$ % of complete days; annual within 0.5 % | 2.6 to 4.8 %; 16 to 26 kWh; 93.7 %; $\leq 0.171$ %                                                                                                                                                                                          | Confirmed                                                                                       |
| H5 voltage ratio                                  | $\sqrt{3}$ before 2016-05-03 09:40, 1 in $\geq 99$ % after                                                                                                   | 100 % and 99.9997 %; plus 66 episodes at $\sim 700$ V with phase channels absent                                                                                                                                                            | Confirmed, with amendment                                                                       |
| H5 sentinels                                      | −15.000 and −273.2 occur only as exact values                                                                                                                | 248 near-values and 116 other negatives (2337 negative values in all); −273.2 stored as float32                                                                                                                                             | Failed as worded                                                                                |
| H6(i) capacity factor                             | Naive (keep-last) 2023 to 2024 value undefined or < 1 % where corrected is 25 to 35 %                                                                        | Naive 50 to 82 %, corrected 23 to 34 % (the plan's premise said 25 to 35 %)                                                                                                                                                                 | Effect confirmed; failed as worded (direction of the error)                                     |
| H6(ii) power curve                                | Primary step 1 to 9 %, sensor 2 lstepl < 2 %                                                                                                                 | 1.2 to 9.1 % (T1 9.14 %, above the stated upper bound); sensor 2 at most 1.5 %; evaluated on the H2c comparison (2018 to 2022 minus 2017, $8 \text{ m s}^{-1}$ bin), where the plan named a 2017 to 2018 step at 6 to $10 \text{ m s}^{-1}$ | Failed as worded on T1 (upper bound); confirmed otherwise                                       |
| H6(iii) missingness                               | Overstatement $\geq 30$ percentage points for 2016 and $\geq 5$ percentage points for every group in its onset year                                          | 32.6 percentage points for the 2016-05-03 group; 6.8 to 73.0 percentage points over the ten monitoring-system groups; 0 percentage points for the two file-schema groups                                                                    | Confirmed for the ten monitoring-system groups; failed as worded for the two file-schema groups |



## Appendix B: Reconstruction algorithm

The input is a 2023 or 2024 `Turbine_Data` file. Read the ten header lines and keep them. For each data row, take the 19-character timestamp prefix; if the timestamp has not been seen, store the row and record the timestamp; otherwise discard  
 385 the row. Count a block boundary whenever a timestamp is smaller than its predecessor. Write the header and the stored rows in timestamp order. The reconstruction script reports the row count, the block count, the number of unique timestamps, the first and last timestamp and the number of later occurrences that carried a wind speed value, for comparison with the expected values (52,560 or 52,704 unique timestamps, 81 blocks, zero later copies with a value); the constant 655-row increment and the comparison with the 10-minute grid are checked separately by the audit scripts. The rule "first occurrence" is independent  
 390 of the page length; the checks are what tie the output to this release. The script processes a 2.9 GB file in a few minutes and produces a file of 214 to 218 MB.

## Appendix C: Robustness variants for the power step

With an operating-state filter (time-based system availability = 1 or data availability = 1, each with power > 0), at bin centres of 7, 8 and 9 m s<sup>-1</sup> and widths of 0.5 and 1.0 m s<sup>-1</sup>, with and without the anemometer-defect turbine-years (T1 2021 to 2022,  
 395 T5 2022 to 2023, T4 2023), the primary-channel step from 2017 to 2018 to 2022 is positive on all six turbines in all 24 variants (range +7.3 to +116.4 kW, +0.7 to +10.5 %) and the sensor 2 step lies between -9.4 and +18.1 kW, with the bootstrap interval including zero on two to four turbines per variant. Without an operating-state filter the primary step is negative on turbine 4 (down to -39.5 kW) and the sensor 2 step is negative in every variant on turbines 4 to 6 (down to -63.1 kW) and in some variants on turbines 1 to 3, because intervals of curtailment and downtime enter the bins in proportions that differ between  
 400 years; this variant does not compare like with like and is reported for completeness. With 12 June to 31 December 2017 instead of the whole of 2017 as the reference window (the plan named June to October), the primary-channel step is +21.6 [16.1, 26.9] kW on turbine 4 to +99.6 [94.0, 105.1] kW on turbine 1, larger than the main estimates on every turbine; Table 4e gives the resulting decomposition of the H2c shortfall. The plan's consequence hypothesis H6(ii) speaks of 6 to 10 m s<sup>-1</sup> while its robustness section names centres of 7, 8 and 9 m s<sup>-1</sup>; the latter were run. The identity tests of Sect. 4.2 have zero violations at  
 405 tolerances of 0.01, 0.02 and 0.05 m s<sup>-1</sup>; the day-level transition criterion gives the same last "mean" day (2018-04-10) at 0.03 m s<sup>-1</sup> on all turbines and at 0.01 m s<sup>-1</sup> on five, with turbine 5 moving to 2018-05-05, a day on which the two sensors differ by 0.03 m s<sup>-1</sup>. Bootstrap intervals for the exceedance share widen by at most 1.3 percentage points when 7-day blocks replace 1-day blocks; the 7-day intervals are reported in Table 4d, computed on the same intervals as the point estimates. The full set of variants, the tolerance sensitivity of the day-level transition criterion and the block-length comparison are tabulated in the  
 410 archived outputs (see Code availability).

*Code availability.* The audit code is archived on Zenodo at <https://doi.org/10.5281/zenodo.23171462> (Park et al., 2026a) under CC BY 4.0. It contains the per-file profiling script (`02_profile_scada.py`), the reconstruction script for the 2023 and 2024 files with a batch



driver (10\_reconstruct\_2023\_2024.py, 11\_batch\_reconstruct.sh, reconstruct\_all.py), the audit scripts in run or-  
der (p02\_fullpass.py to p16\_header\_interval.py, p19\_derived\_outputs.py and p22\_review\_checks.py), the  
415 scripts that build the figures and tables, the outputs from which every number, figure and table in this paper is generated, a README that  
maps each figure and table to the script that produces it and the output file it is built from, the log of deviations from the pre-specified plan,  
the plan itself and three independent verification reports.

*Data availability.* The Kelmarsh v4 release is available at <https://doi.org/10.5281/zenodo.16807551> (Plumley and Takeuchi, 2025a) and the  
Penmanshiel v3 release at <https://doi.org/10.5281/zenodo.16807304> (Plumley and Takeuchi, 2025b), both under CC BY 4.0. The derived data  
420 are archived on Zenodo at <https://doi.org/10.5281/zenodo.23171808> (Park et al., 2026b) under CC BY 4.0, with attribution to the original  
release. They comprise the reconstructed 2023 and 2024 `Turbine_Data` files for Kelmarsh (twelve files, 214 to 218 MB each), the per-  
signal onset manifest (first and last valid timestamp and count per column and turbine) and the flag table (54 turbine-year files with twelve  
boolean flags and a summary flag per 10-minute interval, produced by `p12_flags.py`; counts in Table 7; ranges: wind speed 0 to 40  
 $\text{m s}^{-1}$ , power  $-100$  to 2300 kW, rotor speed 0 to 20 RPM, pitch  $-10$  to  $100^\circ$  after wrapping angles above  $180^\circ$  to negative, temperatures  
425  $-40$  to  $150^\circ\text{C}$  and  $-40$  to  $200^\circ\text{C}$  for the transformer, grid voltage above 550 V after the scale switch), with a README describing every  
file and flag.

*Author contributions.* AP: conceptualisation, methodology, software, writing (original draft), writing (review and editing), project adminis-  
tration. LY: writing (review and editing). CL: formal analysis, data curation. AW: methodology. YC: supervision.

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

430 *Acknowledgements.* We thank Cubico Sustainable Investments, and Charlie Plumley and Roberta Takeuchi in particular, for releasing the  
Kelmarsh and Penmanshiel data under an open licence. An AI system (Claude, Anthropic) assisted with writing the analysis scripts and  
validating the text for correct formatting and writing conventions, under the authors' direction; the authors reviewed and verified the content  
of the text and the generated scripts and take full responsibility for the content.

*Financial support.* This research has not received external funding.



## 435 References

- Astolfi, D., Pandit, R., Celesti, L., Lombardi, A., and Terzi, L.: SCADA data analysis for long-term wind turbine performance assessment: A case study, *Sustain. Energy Technol. Assess.*, 52, 102 357, <https://doi.org/10.1016/j.seta.2022.102357>, 2022.
- Astolfi, D., Pandit, R., Lombardi, A., and Terzi, L.: Diagnosis of wind turbine systematic yaw error through nacelle anemometer measurement analysis, *Sustain. Energy Grids Netw.*, 34, 101 071, <https://doi.org/10.1016/j.segan.2023.101071>, 2023.
- 440 Barber, S. and Ding, Y.: Improving data sharing in practice – power curve benchmarking case study, *J. Phys.: Conf. Ser.*, 2745, 012 002, <https://doi.org/10.1088/1742-6596/2745/1/012002>, 2024.
- Chen, D., Li, Q., Lin, H., Cheng, S., and Yang, X.: A method using SCADA data for diagnosing wind turbine’s anemometer systematic errors, *Renew. Energy*, 259, 125 081, <https://doi.org/10.1016/j.renene.2025.125081>, 2026.
- Clifton, A., Barber, S., Bray, A., Enevoldsen, P., Fields, J., Sempreviva, A. M., Williams, L., Quick, J., Purdue, M., Totaro, P., and Ding, Y.:  
 445 Grand challenges in the digitalisation of wind energy, *Wind Energ. Sci.*, 8, 947–974, <https://doi.org/10.5194/wes-8-947-2023>, 2023.
- Cui, H., Wang, X., Liu, B., and Liu, Y.: Physics-Guided Adaptive DBSCAN for Sensor Data Quality Assurance in Wind Turbine SCADA Systems, *IEEE Sens. J.*, 26, 24 675–24 685, <https://doi.org/10.1109/jsen.2026.3710027>, 2026.
- Deignan, C., González Sopena, J. M., Ghosh, B., and Pakrashi, V.: Portability of short term wind power forecasting: investigating model calibration using wind power data from Ireland and UK, *Sci. Rep.*, 15, 33 891, <https://doi.org/10.1038/s41598-025-08751-3>, 2025.
- 450 Effenberger, N. and Ludwig, N.: A collection and categorization of open-source wind and wind power datasets, *Wind Energy*, 25, 1659–1683, <https://doi.org/10.1002/we.2766>, 2022.
- Fiocchi, F., Ladopoulou, D., and Dellaportas, P.: Probabilistic Multilayer Perceptrons for Wind Farm Condition Monitoring, *Wind Energy*, 28, e70 012, <https://doi.org/10.1002/we.70012>, 2025.
- Hamilton, S. D., Millstein, D., Bolinger, M., Wiser, R., and Jeong, S.: How Does Wind Project Performance Change with Age in the United  
 455 States?, *Joule*, 4, 1004–1020, <https://doi.org/10.1016/j.joule.2020.04.005>, 2020.
- Hammer, F. and Barber, S.: Data imputation for SCADA data using Graph Neural Networks, *J. Phys.: Conf. Ser.*, 3025, 012 014, <https://doi.org/10.1088/1742-6596/3025/1/012014>, 2025.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J.,  
 Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant,  
 460 P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, *Nature*, 585, 357–362, <https://doi.org/10.1038/s41586-020-2649-2>, 2020.
- International Electrotechnical Commission: IEC 61400-12-1:2022, Tech. rep., IEC, Geneva, wind energy generation systems – Part 12-1: Power performance measurements of electricity producing wind turbines, 3rd edn., International Electrotechnical Commission, Geneva, 2022.
- 465 Jacquet, C., Hermoso, N. M., and Pulikollu, R.: A SCADA-Informed Framework Combining Machine Learning and Wind Speed Correction for Yaw Misalignment Detection, *J. Phys.: Conf. Ser.*, 3224, 062 049, <https://doi.org/10.1088/1742-6596/3224/6/062049>, 2026.
- Leahy, K., Gallagher, C., O’Donovan, P., and O’Sullivan, D. T. J.: Issues with Data Quality for Wind Turbine Condition Monitoring and Reliability Analyses, *Energies*, 12, 201, <https://doi.org/10.3390/en12020201>, 2019.
- Martín-Calzada, P., Martín Sánchez, P., Rodríguez-Sánchez, F. J., Santos-Pérez, C., and Ballesteros, J.: Optimized Sensor Data Preprocessing  
 470 Using Parameter-Transfer Learning for Wind Turbine Power Curve Modeling, *Sensors*, 25, 5329, <https://doi.org/10.3390/s25175329>, 2025.



- McKinney, W.: Data Structures for Statistical Computing in Python, in: Proceedings of the 9th Python in Science Conference, 56–61, <https://doi.org/10.25080/Majora-92bf1922-00a>, 2010.
- Mittelmeier, N. and Kühn, M.: Determination of optimal wind turbine alignment into the wind and detection of alignment changes with SCADA data, *Wind Energ. Sci.*, 3, 395–408, <https://doi.org/10.5194/wes-3-395-2018>, 2018.
- Morrison, R., Liu, X., and Lin, Z.: Anomaly detection in wind turbine SCADA data for power curve cleaning, *Renew. Energy*, 184, 473–486, <https://doi.org/10.1016/j.renene.2021.11.118>, 2022.
- Park, A., Yin, L., Lee, C., Wong, A., and Choi, Y.: Code and outputs for “Structural defects in the Kelmarsh open SCADA dataset: a provenance audit with file reconstruction and downstream consequences”, Zenodo [code], <https://doi.org/10.5281/zenodo.23171462>, 2026a.
- 480 Park, A., Yin, L., Lee, C., Wong, A., and Choi, Y.: Kelmarsh wind farm SCADA data (v4): reconstructed 2023–2024 files, signal onset manifest and flag table, Zenodo [data set], <https://doi.org/10.5281/zenodo.23171808>, 2026b.
- Plumley, C. and Takeuchi, R.: Kelmarsh wind farm data (v4), Zenodo [data set], <https://doi.org/10.5281/zenodo.16807551>, 2025a.
- Plumley, C. and Takeuchi, R.: Penmanshiel wind farm data (v3), Zenodo [data set], <https://doi.org/10.5281/zenodo.16807304>, 2025b.
- Pozzobon, E. P., Pinheiro, H., and Andrade, T. A. N. d.: An Ensemble of Neighbor-Informed Models for Wind Turbine Power Prediction with Uncertainty Quantification, *J. Phys.: Conf. Ser.*, 3224, 062 068, <https://doi.org/10.1088/1742-6596/3224/6/062068>, 2026.
- 485 Staffell, I. and Green, R.: How does wind farm performance decline with age?, *Renew. Energy*, 66, 775–786, <https://doi.org/10.1016/j.renene.2013.10.041>, 2014.
- Tartt, K., Kazemi-Amiri, A. M., Nejad, A. R., Carroll, J., and McDonald, A.: Life extension of wind turbine drivetrains by means of SCADA data: Case study of generator bearings in an onshore wind farm, *Results Eng.*, 24, 102 921, <https://doi.org/10.1016/j.rineng.2024.102921>, 2024.
- 490 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., ’t Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*, 3, 160 018, <https://doi.org/10.1038/sdata.2016.18>, 2016.