

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

Adam Park, Luke Yin, Cooper Lee, Andrew Wong, Yoo Choi. Correspondence: Adam Park (apark1@imsa.edu).

This supplement contains Tables S1 to S4 and describes the supplementary archive (code, outputs and tables) from which every number in the paper is generated.

Table S1. Header-stated interval against unique timestamps, all 54 files (2023 and 2024 after reconstruction).

Year	Files	Header start	Header end	Header steps	Unique timestamps	First timestamp	Missing steps
2016	6	2016-01-01 00:00:00	2017-01-01 00:00:00	52,704	52,416	2016-01-03 00:00:00	288
2017	6	2017-01-01 00:00:00	2018-01-01 00:00:00	52,560	52,560	2017-01-01 00:00:00	0
2018	6	2018-01-01 00:00:00	2019-01-01 00:00:00	52,560	52,560	2018-01-01 00:00:00	0
2019	6	2019-01-01 00:00:00	2020-01-01 00:00:00	52,560	52,560	2019-01-01 00:00:00	0
2020	6	2020-01-01 00:00:00	2021-01-01 00:00:00	52,704	52,704	2020-01-01 00:00:00	0
2021	6	2021-01-01 00:00:00	2022-01-01 00:00:00	52,560	52,560	2021-01-01 00:00:00	0
2022	6	2022-01-01 00:00:00	2023-01-01 00:00:00	52,560	52,560	2022-01-01 00:00:00	0
2023	6	2023-01-01 00:00:00	2024-01-01 00:00:00	52,560	52,560	2023-01-01 00:00:00	0
2024	6	2024-01-01 00:00:00	2025-01-01 00:00:00	52,704	52,704	2024-01-01 00:00:00	0

Table S2. Constant-column classification (check 8) of the 319 columns with values.

Classes: constant: 10; flag-like (2-3 values): 4; varying: 305. The 14 non-varying columns are listed below; the full classification of all 319 columns is included in the archived outputs (Code availability).

Column	Files present	Max distinct values	Value if constant	Classification
Compensated Lost Production (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Bats) (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Birds) (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Boat Action) (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Grid) (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Ice) (kWh)	54	1	0.0	constant

Column	Files present	Max distinct values	Value if constant	Classification
Lost Production to Curtailment (Marketing) (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Noise) (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Sector Management) (kWh)	54	1	0.0	constant
Lost Production to Curtailment (Shadow) (kWh)	54	1	0.0	constant
Available Capacity for Production (Planned) (kW)	54	3		flag-like (2-3 values)
Data Availability	54	2		flag-like (2-3 values)
Night Time	12	2		flag-like (2-3 values)
Time-based System Avail. (Planned)	54	3		flag-like (2-3 values)

Table S3. Per-file paginated-export statistics for the 13 Penmanshiel v3 files: rows as delivered, cumulative blocks, block increment, unique timestamps, timestamps expected from the header interval, first and later occurrences with a power value, rows with a power value after keep-last de-duplication, and rows per unique timestamp. The monthly files are the February 2023 exports of nine turbines; the full-year files cover 2023 and 2024 for WT11 and WT15.

File (turbine, period)	Rows	Blocks	Block increment (rows)	Unique timestamps	Timestamps expected from the header interval	First occurrences with power	Later occurrences with power	Rows with power after keep-last	Rows per unique timestamp
WT01, Feb 2023	20,076	8	573	4032	4032	4032	0	21	4.98
WT02, Feb 2023	20,076	8	573	4032	4032	4032	0	21	4.98
WT04, Feb 2023	20,076	8	573	4032	4032	4032	0	21	4.98
WT05, Feb 2023	20,076	8	573	4032	4032	4032	0	21	4.98
WT06, Feb 2023	20,076	8	573	4032	4032	4030	0	21	4.98
WT07, Feb 2023	20,076	8	573	4032	4032	4030	0	21	4.98
WT08, Feb 2023	20,076	8	573	4032	4032	4030	0	21	4.98
WT09, Feb 2023	20,076	8	573	4032	4032	4030	0	21	4.98
WT10, Feb 2023	20,076	8	573	4032	4032	4030	0	21	4.98
WT11, 2023	2,451,138	92	573	52,560	52,560	52,395	0	417	46.64
WT15, 2023	2,451,138	92	573	52,560	52,560	52,393	0	416	46.64

File (turbine, period)	Rows	Blocks	Block increment (rows)	Unique timestamps	Timestamps expected from the header interval	First occurrences with power	Later occurrences with power	Rows with power after keep-last	Rows per unique timestamp
WT11, 2024	2,451,282	92	573	52,704	52,704	52,288	0	561	46.51
WT15, 2024	2,451,282	92	573	52,704	52,704	52,580	0	561	46.51

Table S4. Number of 10-minute intervals flagged by each rule of the flag table, per turbine-year. Thresholds as in Table 7 of the paper; the voltage-scale flag uses the ratio of the grid voltage to the phase L1 voltage.

Year	Turbine	Intervals	Sensor 2 wind speed = -15 m/s-1	Sensor 2 wind speed < 0	Sensor 1 wind speed < 0	Wind speed out of range	Power out of range	Rotor speed out of range	Pitch out of range	Temp. out of range	Transf. temp. = -273.2 °C	Voltage on √3 scale	Grid voltage > 550 V	Reconstructed file	Any flag
2016	1	52,416	0	0	0	0	0	13	20	0	0	14,175	0	0	14,208
2016	2	52,416	0	0	0	0	0	0	18	0	0	14,462	0	0	14,480
2016	3	52,416	0	0	0	0	0	0	17	0	0	13,307	0	0	13,324
2016	4	52,416	0	0	0	0	0	0	18	0	0	12,570	0	0	12,588
2016	5	52,416	0	0	0	0	0	0	17	0	0	13,569	0	0	13,586
2016	6	52,416	0	0	0	0	0	0	22	0	0	12,450	0	0	12,472
2017	1	52,560	0	0	0	0	0	11	17	0	0	0	0	0	28
2017	2	52,560	0	0	0	0	0	0	33	0	0	0	0	0	33
2017	3	52,560	0	0	0	0	0	0	30	0	0	0	0	0	30
2017	4	52,560	0	0	0	0	0	0	32	0	0	0	0	0	32
2017	5	52,560	0	0	0	0	0	0	26	0	0	0	0	0	26
2017	6	52,560	0	0	0	0	0	0	30	0	0	0	0	0	30
2018	1	52,560	0	0	0	0	0	0	8	0	0	0	0	0	8
2018	2	52,560	0	0	0	0	0	0	3	0	0	0	0	0	3
2018	3	52,560	187	188	1	0	0	0	4	0	0	0	0	0	192
2018	4	52,560	0	0	0	0	0	0	157	0	0	0	0	0	157
2018	5	52,560	184	185	2	0	0	0	5	0	0	0	0	0	190
2018	6	52,560	0	0	0	0	0	0	5	0	0	0	0	0	5
2019	1	52,560	0	0	0	0	0	0	0	0	0	0	2344	0	2344
2019	2	52,560	0	0	0	0	0	0	0	0	0	0	0	0	0
2019	3	52,560	0	0	0	0	0	0	0	0	0	0	0	0	0
2019	4	52,560	0	0	0	0	0	0	0	3	3	0	0	0	3
2019	5	52,560	0	0	0	0	0	0	0	0	0	0	0	0	0
2019	6	52,560	0	0	0	0	0	0	0	0	0	0	0	0	0
2020	1	52,704	0	0	0	0	0	0	0	0	0	0	254	0	254
2020	2	52,704	0	0	0	0	0	0	0	0	0	0	254	0	254
2020	3	52,704	2	3	2	0	0	0	0	0	0	0	254	0	257
2020	4	52,704	0	0	0	0	0	0	0	0	0	0	254	0	254
2020	5	52,704	0	0	0	0	0	0	0	0	0	0	254	0	254
2020	6	52,704	0	0	0	0	0	0	0	0	0	0	254	0	254
2021	1	52,560	0	7	0	0	0	0	7	0	0	0	2350	0	2364
2021	2	52,560	0	0	0	0	0	0	6	0	0	0	2344	0	2350

Year	Tur- bine	Inter- vals	Sensor 2 wind speed = -15 m s-1	Sensor 2 wind speed < 0	Sensor 1 wind speed < 0	Wind speed out of range	Power out of range	Rotor speed out of range	Pitch out of range	Temp. out of range	Transf. temp. = -273.2 °C	Voltage on √3 scale	Grid voltage > 550 V	Recon- structed file	Any flag
2021	3	52,560	0	0	0	0	0	0	0	0	0	0	2342	0	2342
2021	4	52,560	0	0	0	0	0	0	0	0	0	0	2343	0	2343
2021	5	52,560	1	2	1	0	0	0	0	0	0	0	2344	0	2346
2021	6	52,560	0	0	0	0	0	0	0	0	0	0	2338	0	2338
2022	1	52,560	1	19	0	0	0	0	9	3	2	0	43	0	74
2022	2	52,560	1	2	0	0	0	0	0	134	0	0	43	0	179
2022	3	52,560	0	0	0	0	0	0	0	0	0	0	43	0	43
2022	4	52,560	0	0	0	0	0	0	0	0	0	0	43	0	43
2022	5	52,560	0	5	0	0	0	0	6	0	0	0	43	0	54
2022	6	52,560	24	25	0	0	0	0	1013	0	0	0	43	0	1081
2023	1	52,560	0	0	0	0	0	0	0	0	0	0	0	52,560	0
2023	2	52,560	0	0	0	0	0	0	0	0	0	0	0	52,560	0
2023	3	52,560	0	0	0	0	0	0	0	0	0	0	0	52,560	0
2023	4	52,560	1573	1881	1	0	0	0	0	1598	0	0	0	52,560	1905
2023	5	52,560	0	20	0	0	0	0	5	0	0	0	0	52,560	25
2023	6	52,560	0	0	0	0	0	0	0	0	0	0	0	52,560	0
2024	1	52,704	0	0	0	0	0	0	0	0	0	0	625	52,704	625
2024	2	52,704	0	0	0	0	0	0	0	0	0	0	625	52,704	625
2024	3	52,704	0	0	0	0	0	0	0	0	0	0	625	52,704	625
2024	4	52,704	0	0	0	0	0	0	0	0	0	0	625	52,704	625
2024	5	52,704	0	0	0	0	0	0	0	0	0	0	625	52,704	625
2024	6	52,704	0	0	0	0	0	0	0	0	0	0	609	52,704	609

Supplementary archive

The code archive (see Code availability in the article) contains the profiling, reconstruction and audit scripts; the outputs from which every number in the paper is generated; Tables 1 to 7 and S1 to S4 as CSV and Markdown; Figures 1 to 5; a README mapping every figure and table to its script and output file; the pre-specified analysis plan; the log of every departure from the plan; and the three independent verification reports. The reconstructed 2023 and 2024 files, the onset manifest and the flag table are deposited separately (see Data availability).