COMPOUNDING ENERGY LTD

Working Paper Series • CE-NOTE-2026-03

The Graded Scorecard

What CompoundVision Publishes About Its Own Wind and Solar Forecast Accuracy, in Full

Dr Christopher T. M. Clack
Founder, Compounding Energy Ltd
christopher@compoundingenergy.com

September 2026

Abstract. A forecast vendor's accuracy page is usually a curated number. CompoundVision publishes three scorecards on open endpoints instead, recomputed on a schedule, with the farms it declines to score and the regions where it loses to persistence printed alongside the wins. This note freezes those scorecards on a stated date and reports them in full. On the 6 September 2026 snapshot, site calibration against Elexon metered output improves the out-of-sample RMSE of 64 of the 65 scored GB wind farms relative to the scorecard's IEC reference curve, by a median of 42% and a capacity-weighted 45% across 10.5 GW; one 36 MW farm scores worse after calibration and five farms are not scored, each with its published reason. Across 38 regional fleets the product beats its stated persistence baseline in 24 of 32 wind regions and 29 of 30 solar regions at up to 48 hours ahead, and the eight wind regions where it does not are named. Five forward-check bands were registered against that snapshot before any later one existed and are graded here, exactly as registered, on a second snapshot taken on 22 September: the calibration result holds on all

three of its bands, while the set of wind regions beating persistence moves by more than its band allowed, and every region that changed sides had sat near zero. The note explains what each scorecard does and does not measure, and what a buyer should take from a record published in full.

Keywords: forecast verification, wind power, solar power, site calibration, persistence skill, validation record, CompoundVision

Contents

1 Why publish the scorecard

The previous note in this series (Clack, 2026) published the CECadence battery revenue record in one place, failures included, on the argument that a model sold to lenders is judged by what happened on data its authors had not seen. This note does the same for CompoundVision, the wind and solar power forecasting product, with one difference: CompoundVision already publishes its record continuously, on three open endpoints that anyone can fetch without an account (Compounding Energy Ltd, 2026b,a,c). What a regularly recomputed record lacks is a fixed point to cite. A number read from a live page on Monday is not the number on the page on Friday, and a claim frozen into marketing copy drifts away from the endpoint that was supposed to back it. Between the two snapshots this note reports, about two weeks apart, the list of GB farms that scored worse after calibration changed completely; a sentence frozen around either list would already be wrong.

The discipline this note adopts is therefore the one the series uses for models: freeze the record on a stated date with a content hash, report all of it, and register a forward check before any later recompute exists, so that at least one claim in the note is a prediction rather than a description. The descriptive sections were written after their snapshot was inspected and say so; the five bands of Section 5 are the one part registered blind.

2 What is published, and how it is graded

Three scorecards are reported. The public payloads carry their own method lines, and the description below quotes them rather than adding to them.

GB per-farm site calibration. The scorecard's cohort is 70 GB wind farms, all metered in Elexon B1610 (Elexon, 2026); the payload does not state how the cohort is selected. For each farm it scores, it compares two power curves on the same held-out hours: what the payload calls the IEC reference curve, and the site-calibrated curve the product fits to the farm's own metered record. The published method line states the basis: out-of-sample RMSE of forecast farm power in MW against metered generation, with air density pinned, a single deterministic weather member, and no ensemble or statistical post-processing on either side, so that the comparison isolates the curve. Curves are re-fit on a day-block holdout with every third calendar day held out and scored on those days only. Hours flagged by the payload's curtailment filter, `elexon_windfor_boalf`, are excluded. Farms the scorecard declines to score are listed with a suppression label.

EU per-farm site calibration. The same design against ENTSO-E per-unit metered generation (ENTSO-E, 2026), wind units only, with hours whose bidding-zone day-ahead price cleared negative excluded from fit and scoring on the published reasoning that a negative price is a market decision to curtail rather than a weather miss; hours without price evidence are kept, and each farm's evidence coverage is published (0.87 to 0.88 on snapshot A). On the snapshot the cohort is three farms; it is reported as such.

Regional fleet skill. For each of 38 regional fleets, 28 in Europe and 10 in the United States, the scorecard grades the fleet-level wind and solar forecast at up to 48 hours ahead over a fourteen-day lookback against the product's stated persistence baseline, reporting a relative skill score in which zero equals the naive forecast and positive values beat it, of the kind standard in forecast verification (Murphy, 1988); persistence is the conventional floor for short-horizon power forecasting (Giebel and Kariniotakis, 2017). Where the modelled fleet and the available actuals cover materially different capacity the scorecard suppresses the cell rather than score it, and prints the coverage ratio.

What makes the record citable is that every payload carries the timestamp at which it was computed alongside its method line and suppression labels, so a snapshot can be tied to a specific recompute and read against the definition it was computed under. The thresholds behind the suppression labels are the product's internals and are not described here.

3 The GB calibration record

Table 1 reports the two calibration scorecards as published on the snapshot, and Table 2 the distribution across the scored GB farms. Figure 1 plots every scored farm against its capacity.

Table 1: Site-calibration scorecards as published (snapshot 2026-09-06_A, fetched 2026-09-06 17:03 UTC). Out-of-sample RMSE of forecast farm power against metered generation on a day-block holdout, IEC reference curve before versus site-calibrated curve after; the product publishes the method line and the suppression labels alongside the numbers.

	GB (Elexon B1610)	EU (ENTSO-E A73)
Scorecard computed (UTC)	2026-09-06 03:04	2026-09-06 04:26
Window (days)	60	60
Curtailment filter	elexon_windfor_boalf	negative_da_price
Farms scored	65	3
Farms suppressed (not scored)	5	0
Capacity scored (MW)	10,497	1,363
Farms improved / regressed	64 / 1	3 / 0
Median RMSE improvement	42.0%	57.2%
Mean RMSE improvement	38.5%	60.0%
Capacity-weighted RMSE improvement	45.5%	59.3%
Summed RMSE reduction (MW)	1,156	249

Table 2: Distribution of per-farm out-of-sample RMSE improvement across the 65 scored GB farms (snapshot 2026-09-06_A). The regressed farm and the suppressed farms are listed in Table 4.

Statistic	Value
Minimum	-10.7%
10th percentile	10.8%
25th percentile	29.1%
Median	42.0%
75th percentile	51.7%
90th percentile	60.0%
Maximum	73.8%
Share of farms improving by $\geq 25\%$	80%
Share of farms improving by $\geq 50\%$	29%
Farms improving by less than 10%	5
Farms scoring worse after calibration	1

Read as a distribution rather than a slogan, the record says this. Of 70 GB farms in the metered cohort, 65 are scored and 64 of those improve. Four in five scored farms improve by a quarter or more and three in ten by half or more; five improve by less than ten percent, and one scores worse. The median is 42%, the mean 38% and the capacity-weighted figure 45%, the last above the median because the largest farms improve most on average: the ten largest improve by 49.8% capacity-weighted against 37.4% for the rest (Tables 3 and 7), with Moray East at 67%, Gwynt y Môr at 62%, Greater Gabbard at 57%, and the two Hornsea sites at 49% and 38%. Every scored farm is listed in the appendix (Table 7). Summed across the scored fleet the reduction in RMSE is 1,156 MW on 10,497 MW of capacity. The three-farm EU cohort, Fécamp,

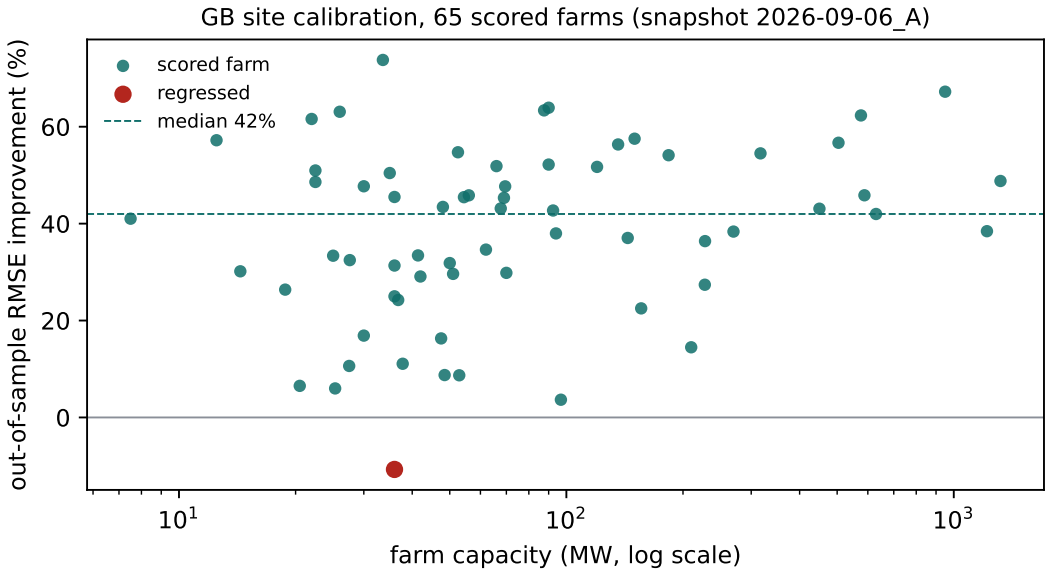

Figure 1: Out-of-sample RMSE improvement from site calibration for every scored GB farm on the 6 September 2026 snapshot, against farm capacity. The dashed line is the published median; the red point is the one farm that scored worse after calibration.

Yeu-Noirmoutier and Norther, improves by a median 57%; with three farms that is a report, not a statistic.

Table 3: The ten largest scored GB farms (snapshot 2026-09-06_A): out-of-sample RMSE in MW before and after site calibration and the held-out test pairs.

Farm	MW	Before	After	Improvement	Pairs
Hornsea 2 - Optimus and Breesea	1,320	340.5	174.3	48.8%	268
Hornsea 1 - Heron & Njord	1,218	297.9	183.4	38.4%	268
Moray East	950	274.7	90.0	67.2%	203
London Array Phase 1	630	131.6	76.3	42.0%	265
Beatrice	588	156.9	84.9	45.9%	267
Gwynt y Mor	576	163.0	61.4	62.3%	268
Greater Gabbard Wind Farm	504	131.4	56.9	56.7%	266
Neart na Gaoithe	450	95.7	54.5	43.1%	246
Sheringham Shoal	317	75.2	34.2	54.5%	268
Centrica (Lincs)	270	49.8	30.7	38.4%	268

The exceptions are Table 4. Andershaw, a 36 MW onshore farm, scores 11% worse after calibration on 268 held-out pairs. Five farms are not scored: Viking (443 MW) and West of Duddon Sands (389 MW), the two largest of them, together with Windy Rig and Coire Na Cloiche, carry the label `low_wind_output_corr`, and Sandy Knowe carries `holdout_fit_failed`. The note does not diagnose these. What matters for a reader is that 832 MW of capacity, in two of the cohort's larger farms, is withheld from the headline under a published label, and its scores on this snapshot are unknown.

Two things the GB scorecard does *not* say. It does not measure the accuracy of the served forecast, which uses an ensemble and post-processing that the curve-isolation basis deliberately switches off; it measures what site calibration does to one deterministic member, and the gain in the served forecast is not measured by it. And it does not promise any individual farm an improvement: on this snapshot one farm regressed, and snapshot B, two weeks later, shows two others (Section 5). The claim the record supports is the distribution in Table 2, dated.

Table 4: The exceptions, printed as published: the GB farm whose calibrated curve scored worse out of sample, and the farms the scorecard declined to score, with the product's own suppression label.

Farm	MW	RMSE before	RMSE after	Status
Andershaw	36	3.13	3.46	regressed (-10.7%, 268 test pairs)
Viking Wind Farm	443	—	—	suppressed: low_wind_output_corr
West of Duddon Sands	389	—	—	suppressed: low_wind_output_corr
Sandy Knowe Wind Farm	49	—	—	suppressed: holdout_fit_failed
Windy Rig Wind Farm	43	—	—	suppressed: low_wind_output_corr
Coire Na Cloiche Wind Farm	30	—	—	suppressed: low_wind_output_corr

4 The regional skill record

Table 5 prints all 38 regional fleets and Figure 2 sorts them. The product beats its persistence baseline in 24 of 32 scored wind regions with a median skill of 0.26, and in 29 of 30 scored solar regions with a median skill of 0.56. The strongest wind lines are several large US markets and northern European fleets: MISO, Denmark, SPP, PJM, ERCOT, Poland, Great Britain and Belgium all sit above 0.55. The eight wind regions below persistence are the Netherlands, CAISO, Switzerland, Czechia, Greece, Norway, Luxembourg and ISO New England, with the Netherlands and CAISO the worst at −0.82 and −0.57. Solar is below persistence only in Norway. France's wind cell and the solar cells of Austria, Belgium, Switzerland, Luxembourg, the Netherlands and Slovenia are suppressed under the label `fleet_basis_mismatch`: the published coverage ratios show a modelled solar fleet of a tenth or less of the capacity the actuals cover in Belgium, Switzerland and Slovenia, a mismatch in the other direction in the Netherlands and for French wind, and the product declines to score a fleet whose basis it cannot match.

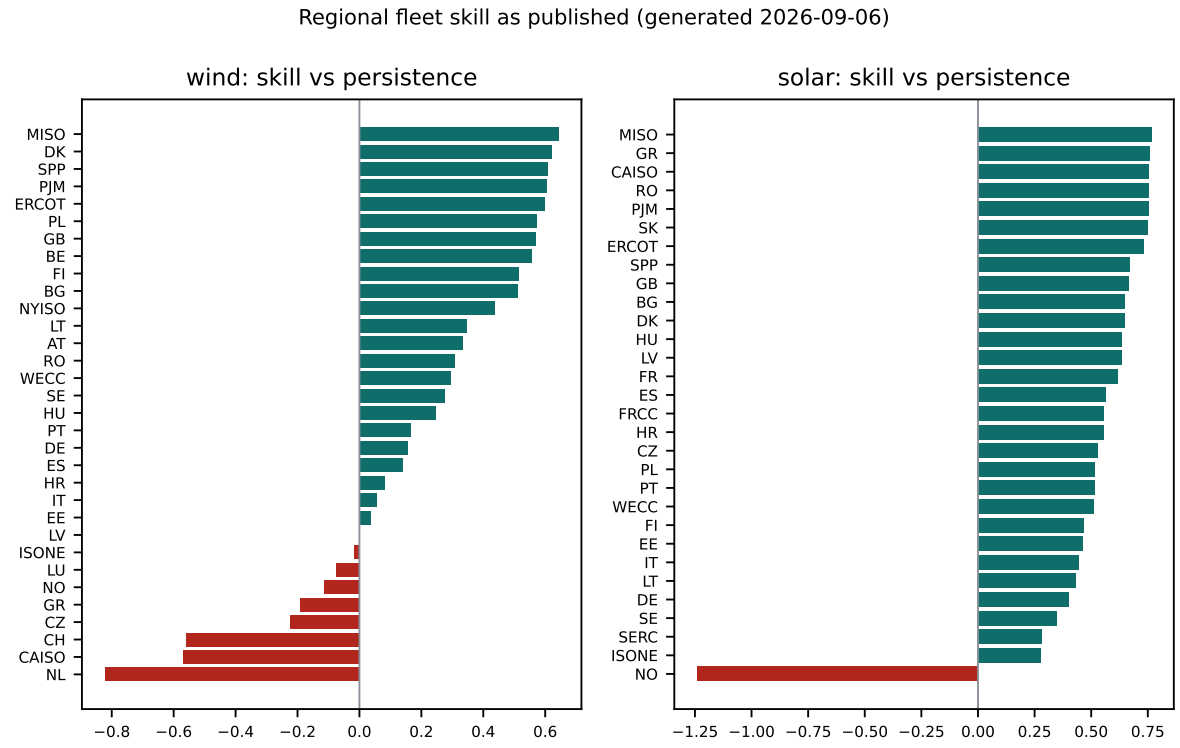

Figure 2: Regional fleet skill against the product's stated persistence baseline, wind (left) and solar (right), every scored region on the 6 September 2026 snapshot. Red bars are regions where the forecast loses to persistence.

Table 5: Regional fleet skill against persistence, all 38 regions as published (generated 2026-09-06 17:06 UTC; 14-day lookback, up to 48 hours ahead). Skill above zero beats the naive forecast. nRMSE is the published normalised RMSE (field rmse_pct; the payload does not state the normaliser). Dashes carry the product's own labels: [a] fleet_basis_mismatch, with the published coverage ratio in brackets; [b] insufficient_samples; [c] persistence_rmse_below_threshold; an unmarked dash means the payload has no entry for that fuel.

Region	Family	Wind skill	Wind nRMSE	n	Solar skill	Solar nRMSE	n
AT	Europe	0.336	0.166	201	—[a] (0.13)	2.769	201
BE	Europe	0.556	0.116	197	—[a] (0.07)	4.782	197
BG	Europe	0.513	0.043	120	0.647	0.352	120
CAISO	US	-0.571	0.383	335	0.754	0.137	335
CH	Europe	-0.559	0.187	197	—[a] (0.00)	122.476	198
CZ	Europe	-0.224	0.407	221	0.527	0.550	221
DE	Europe	0.158	0.118	203	0.398	0.670	185
DK	Europe	0.623	0.106	220	0.646	0.169	220
EE	Europe	0.038	0.107	217	0.463	0.159	217
ERCOT	US	0.599	0.069	309	0.732	0.155	309
ES	Europe	0.141	0.055	222	0.563	0.165	222
FI	Europe	0.514	0.060	219	0.467	0.644	211
FR	Europe	—[a] (4.39)	0.099	185	0.616	0.180	185
FRCC	US	—	—	—	0.555	0.144	308
GB	Europe	0.570	0.060	321	0.663	0.116	321
GR	Europe	-0.193	0.107	223	0.758	0.150	223
HR	Europe	0.082	0.191	222	0.555	0.590	222
HU	Europe	0.247	0.110	196	0.636	0.332	222
IE	Europe	—[b]	0.049	21	—[b]	0.355	88
ISONE	US	-0.018	0.267	308	0.276	0.167	308
IT	Europe	0.057	0.117	225	0.443	0.910	225
LT	Europe	0.348	0.211	222	0.431	0.632	222
LU	Europe	-0.077	0.320	185	—[a] (0.19)	1.428	185
LV	Europe	0.003	0.317	219	0.633	0.277	219
MISO	US	0.644	0.041	309	0.765	0.110	309
NL	Europe	-0.823	0.264	202	—[a] (25.76)	0.379	202
NO	Europe	-0.116	0.077	217	-1.241	0.277	217
NYISO	US	0.437	0.061	334	—[c]	0.373	308
PJM	US	0.604	0.060	308	0.753	0.105	308
PL	Europe	0.574	0.079	212	0.516	0.577	212
PT	Europe	0.166	0.137	201	0.515	0.161	201
RO	Europe	0.308	0.060	222	0.753	0.151	222
SE	Europe	0.277	0.047	213	0.347	1.069	213
SERC	US	—[c]	0.004	309	0.281	0.202	309
SI	Europe	—	—	—	—[a] (0.01)	60.475	222
SK	Europe	—	—	—	0.748	0.146	217
SPP	US	0.610	0.058	309	0.670	0.236	309
WECC	US	0.297	0.046	311	0.511	0.164	311
Regions scored / beating persistence		32 / 24 (median skill 0.26)			30 / 29 (median skill 0.56)		

A fourteen-day lookback is short, and the note treats these numbers accordingly: as the record on a date, not as a rating. The forward check in Section 5 asks whether the sets of regions above and below persistence are stable enough between recomputes to be quoted at all. What the snapshot already establishes is the shape of the disclosure. A buyer in CAISO is shown a wind line that loses to persistence and a solar line at 0.75 on the same page, and can ask for the one they operate in.

5 The forward check

Five bands were written into `PREREGISTRATION.md` on 6 September 2026 against the snapshot reported above, snapshot A, before the second snapshot, snapshot B, existed: (F1) at least 90% of scored GB farms improve; (F2) the GB median improvement stays within ten points of 42%; (F3) the GB capacity-weighted improvement stays at or above 30%; (F4) the set of wind regions beating persistence changes by at most four regions; (F5) the set of solar regions

beating persistence changes by at most three. Snapshot B was fixed by a dated addendum to the registration, written before it was fetched: the same three endpoints fetched on 22 September 2026, carrying the GB and EU calibration scorecards as recomputed on 20 September, two weeks after snapshot A, and the regional scorecard as computed on 22 September. The bands were chosen so that F1 to F3 test whether the calibration result is a property of the fleet rather than of one recompute's draw, and F4 and F5 whether the regional record is stable enough, over the gap between snapshots, to be quoted. Table 6 grades them exactly as registered.

Table 6: Forward check registered before the second snapshot (2026-09-22_B, computed 2026-09-20 03:09 UTC for GB) was fetched. Each band was written down in `PREREGISTRATION.md` against snapshot A and graded on snapshot B unmodified.

Band	Snapshot B	Detail	Result
F1: GB share of scored farms improving $\geq 90\%$	0.97	59/61	PASS
F2: GB median improvement within ± 10 points of snapshot A	45.6	42.0% $\rightarrow$ 45.6%	PASS
F3: GB capacity-weighted improvement $\geq 30\%$	46.6	46.6%	PASS
F4: wind regions beating persistence changes by ≤ 4 (symmetric difference)	6	DE, EE, HR, LV, NO, PT	FAIL
F5: solar regions beating persistence changes by ≤ 3	0	none	PASS

The calibration result held. On snapshot B, 59 of 61 scored GB farms improve, the median improvement is 45.6% against 42.0% on snapshot A, and the capacity-weighted figure is 46.6%. The farms at the edges moved: Andershaw, the regressed farm of snapshot A, now improves by 14%, while two others score worse after calibration: Viking, suppressed on snapshot A, from 42.4 to 44.3 MW of RMSE (4%), and Twentyshilling Hill, a 38 MW farm, from 5.20 to 5.80 MW (11%). Nine farms are suppressed rather than five: West of Duddon Sands, Windy Rig, Coire Na Cloiche and Sandy Knowe again, and Westermost Rough A, Fallago Rig, Arecleoch, Corriemoillie and Glen App newly, all under `low_wind_output_corr` except Sandy Knowe, still `holdout_fit_failed`. On the 60 farms scored in both snapshots the median moves from 41.5% to 45.9%, so the rise is not an artefact of which farms were scored. That is the behaviour the distribution statement in Section 3 anticipates and a per-farm promise would not survive. The solar regional record held exactly: the same 29 of 30 regions beat persistence.

The wind regional band failed. Six regions changed sides against a band of four: Germany, Portugal, Croatia, Estonia and Latvia fell below persistence and Norway rose above it. France and Ireland were newly scored, both below persistence, at -1.50 and -0.65, so the count beating persistence went from 24 of 32 to 20 of 34 even as the median wind skill rose from 0.26 to 0.34. Between the snapshots the median region's wind skill moved by 0.17, and all six regions that changed sides had been within 0.17 of zero on snapshot A, Latvia at 0.003 and Estonia at 0.04. That pattern was not registered, and it is offered as description rather than as a tested result: a regional wind line that sits closer to zero than lines typically move between snapshots should be read as undecided rather than as a win or a loss.

6 What a buyer should take from it

Three things. First, site calibration against metered output is a fleet property with a wide distribution, not a per-farm guarantee: on this snapshot the interquartile range runs from 29% to 52%, one farm regresses and five are withheld. A procurement that asks "how much better" should ask for the distribution and the exceptions, and the product's own page provides both. Second, the calibration scorecards grade the curve, not the served forecast, and say so in their method line; the regional scorecard grades the fleet-level forecast against persistence, and a buyer should look at the line for their own fleet. Third, the regional record is uneven, and the

unevenness is printed; and where a region's wind skill sits near zero, one fortnight's sign is not a finding, as the two snapshots show. Ask for a region's line across several recomputes before relying on it. Where the product loses to persistence it says so; where it cannot score a fleet it withholds the cell and prints why. A vendor page that shows only its median cannot be checked; one that shows the regions where it loses can.

7 Limitations

The calibration windows are 60 days and the regional lookback is fourteen; both are short, and the note's only defence against a lucky draw is the forward check. The curve-isolation basis of the calibration scorecards is deliberately not the served configuration, so the improvement figures neither bound nor summarise served accuracy. The persistence baseline is the product's, and its definition is not on the public payload; the note describes it as stated rather than asserting one. The EU per-farm cohort is three farms. The suppression labels are printed as published and not diagnosed. And the descriptive sections were written after their snapshot was inspected, which is why only the forward-check bands are offered as a test; the pattern behind the failed wind band was observed, not registered.

8 How to check it

The three endpoints are open. `fetch_snapshot.py` in the note's bundle fetches them and writes a dated snapshot with sha256 digests; `make_tables.py` builds every table and figure in this note from the two snapshots and nothing else. Snapshot A is `2026-09-06_A`, fetched 2026-09-06T17:03Z, with GB digest `c7c785fb3924`, EU digest `120de3342bf2` and regional digest `bc2ec74bf450` (first twelve characters; full digests in `manifest.json`). Snapshot B is `2026-09-22_B`, fetched 2026-09-22T21:19Z, with digests `0c5790ae0307`, `6bd019105bcb` and `d91dcfa4055a`. A reader who fetches today will get today's record, which will differ; the differences are the point, and the bands of Section 5 are the note's statement of how much difference the gap between two snapshots should make.

A Every scored farm

Table 7: Every scored farm on snapshot 2026-09-06_A, as published: out-of-sample RMSE in MW for the reference curve (before) and the site-calibrated curve (after), the improvement, and the held-out test pairs. GB farms against Elexon B1610, largest first; EU farms against ENTSO-E A73. The suppressed farms are listed in Table 4.

Farm	MW	Before	After	Improvement	Pairs
GB (Elexon B1610)					
Hornsea 2 - Optimus and Breesea	1320	340.5	174.3	48.8%	268
Hornsea 1 - Heron & Njord	1218	297.9	183.4	38.4%	268
Moray East	950	274.7	90.0	67.2%	203
London Array Phase 1	630	131.6	76.3	42.0%	265
Beatrice	588	156.9	84.9	45.9%	267
Gwynt y Mor	576	163.0	61.4	62.3%	268
Greater Gabbard Wind Farm	504	131.4	56.9	56.7%	266
Neart na Gaoithe	450	95.7	54.5	43.1%	246
Sheringham Shoal	317	75.2	34.2	54.5%	268
Centrica (Lincs)	270	49.8	30.7	38.4%	268
Pen Y Cymoedd	228	35.1	22.3	36.4%	268
Stronelairg Wind Farm	228	38.6	28.0	27.4%	268
Westermost Rough A	210	42.9	36.7	14.5%	268
Walney 1	184	48.6	22.3	54.1%	268
Griffin Wind farm	156	19.0	14.7	22.5%	267

Farm	MW	Before	After	Improvement	Pairs
Ormonde Offshore	150	43.5	18.5	57.5%	268
Fallago Rig	144	32.6	20.5	37.0%	259
Harestanes	136	21.4	9.4	56.3%	268
Arecleoch Windfarm	120	25.7	12.4	51.7%	268
European Offshore Wind Deployment Centre (EOWDC) (Aberdeen Bay - Demonstration site)	97	11.4	11.0	3.7%	268
Dunmaglass Wind Farm	94	18.0	11.2	38.0%	268
Creag Riabhach Wind Farm	92	16.6	9.5	42.7%	268
Burbo Bank	90	20.2	9.7	52.2%	268
Barrow	90	26.7	9.6	63.9%	268
Windy Standard III Wind Farm (Brockloch Rig Phase 2)	88	15.9	5.8	63.4%	265
Enoch Hill	70	13.6	9.6	29.8%	268
Corriegarth	70	18.8	9.8	47.7%	268
Dersalloch	69	14.6	8.0	45.3%	266
Strathy North	68	13.0	7.4	43.1%	268
Galawhistle	66	13.1	6.3	51.9%	268
Kennoxhead Wind Farm	62	11.2	7.3	34.6%	264
Mark Hill	56	10.4	5.6	45.8%	268
Corriemoillie	54	9.1	5.0	45.5%	268
Blackcraig	53	6.8	6.2	8.7%	268
Baillie Wind Farm	53	12.9	5.8	54.7%	268
Middle Muir Wind Farm	51	6.1	4.3	29.6%	266
Afton Wind Farm	50	11.6	7.9	31.9%	268
Glen Kyllachy Wind Farm	49	8.0	7.3	8.8%	267
Hill of Towie (Drummuir)	48	6.9	3.9	43.5%	268
Kilbraur Windfarm	48	6.6	5.5	16.3%	268
Solwaybank	42	6.2	4.4	29.1%	268
Edinbane Wind Farm	41	6.9	4.6	33.4%	253
Twentyshilling Hill Wind Farm	38	4.3	3.8	11.1%	264
Minsca	37	4.8	3.7	24.2%	268
Camster II	36	6.7	4.6	31.4%	268
Strathrory Wind Farm	36	6.9	5.2	25.0%	268
Andershaw	36	3.1	3.5	-10.7%	268
Auchrobert Wind Farm	36	7.2	3.9	45.5%	268
Airies Farm	35	9.5	4.7	50.4%	268
Pines Burn Wind Farm	34	12.6	3.3	73.8%	268
Minnygap	30	3.9	2.1	47.7%	268
Hywind Scotland Pilot Park (Hywind 2) Demonstrator	30	5.4	4.5	16.9%	268
Toddleburn	28	4.8	3.2	32.5%	268
Glenchamber Wind Farm	28	4.1	3.7	10.6%	268
Glens of Foudland	26	5.4	2.0	63.1%	268
Whiteside Hill	25	3.8	3.6	6.0%	268
Assel Valley Wind Farm	25	3.9	2.6	33.4%	268
Freasdail Wind Farm	23	4.7	2.3	51.0%	268
Burn of Whilk	23	5.2	2.7	48.6%	268
Glen App	22	6.9	2.7	61.6%	268
Cour Wind Farm	21	5.7	5.3	6.5%	268
Kirk Hill Wind Farm	19	2.9	2.2	26.4%	268
Tullymurdoch Wind Farm	14	2.0	1.4	30.2%	268
Gordonstown Hill Wind Farm	13	2.1	0.9	57.2%	268
BROWNIELEYS - 3 WIND TURBINES	8	1.1	0.7	41.0%	268
EU (ENTSO-E A73)					
Fecamp Wind Farm	497	113.7	50.7	55.5%	416
Yeu Noirmoutier Wind Farm	496	181.8	77.8	57.2%	433
Norther Offshore Wind	370	121.4	39.8	67.2%	432

References

Christopher T. M. Clack. The graded record: pre-registered backtests of the CECadence GB BESS revenue model, including the failures. Technical Note CE-NOTE-2026-02, Compounding Energy Ltd, 2026. URL https://www.compoundingenergy.com/library/CE-NOTE-2026-02_The_Graded_Record.pdf.

Compounding Energy Ltd. CompoundVision — EU per-farm accuracy scorecard (ENTSO-E A73 calibration), public endpoint `/v1/scorecard/eu`, 2026a. URL https://compoundvisi

on.compoundingenergy.com/v1/scorecard/eu. No authentication; recomputed weekly; snapshot of 6 September 2026 frozen in the note's bundle.

Compounding Energy Ltd. CompoundVision — GB per-farm accuracy scorecard (B1610 calibration), public endpoint `/v1/scorecard/gb`, 2026b. URL https://compoundvision.compoundingenergy.com/v1/scorecard/gb. No authentication; recomputed weekly; snapshot of 6 September 2026 frozen in the note's bundle.

Compounding Energy Ltd. CompoundVision — regional forecast accuracy scorecard, public endpoint `/v1/scorecard`, 2026c. URL https://compoundvision.compoundingenergy.com/v1/scorecard. No authentication; fourteen-day lookback, up to 48 hours ahead; snapshot of 6 September 2026 frozen in the note's bundle.

Elexon. B1610: Actual generation output per generation unit — balancing mechanism reporting service, 2026. URL https://bmrs.elexon.co.uk/. Half-hourly metered output by BM unit; retrieved September 2026.

ENTSO-E. Transparency platform: Actual generation per generation unit (A73), 2026. URL https://transparency.entsoe.eu/. Per-unit metered generation for units of 100 MW and above; retrieved September 2026.

Gregor Giebel and George Kariniotakis. Wind power forecasting — a review of the state of the art. In George Kariniotakis, editor, *Renewable Energy Forecasting: From Models to Applications*, pages 59–109. Woodhead Publishing, 2017. doi: 10.1016/B978-0-08-100504-0.00003-2.

Allan H. Murphy. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review*, 116(12):2417–2424, 1988. doi: 10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.