

A global dataset for coal-to-renewables cost crossover analysis

Lilly Loghmani

Columbia University

Haoran Fu

Columbia University

Gernot Wagner

gwagner@columbia.edu

Columbia University

data-descriptor

Keywords:

Posted Date: October 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10989453/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

A global dataset for coal-to-renewables cost crossover analysis

Lilly Loghmani¹, Haoran Fu¹, and Gernot Wagner^{1,*}

¹Columbia University, New York, NY, USA

*Corresponding author: gwagner@columbia.edu

Abstract

Renewable electricity is cheap and getting cheaper, while coal is still the single largest power source globally. Plant-level cost comparisons are difficult because coal fleets, renewable resources, technology costs, and financing conditions are reported in separate datasets with often incompatible conventions. We here combine these sources for 6,580 operating coal-fired units at 2,384 plants in 73 countries, allowing for a direct comparison of ‘coal crossover’ costs, comparing unit-level expenses of continuing to operate coal plants (OpEx) with the complete levelized cost for capital and operating expenses (CapEx and OpEx) of nearby solar and onshore wind, given local resource and financing conditions. We report least-cost technologies, break-even financing rates, and required renewable capacity and investment, in addition to four-hour lithium-ion battery storage to complement intermittent solar and wind power generation. The dataset and corresponding dashboard, available at <https://coalcrossover.org> [*password: "GCT" (before publication)*], provides a common, transparent basis for studying how local resources, technology costs, and finance shape coal-to-renewable crossover estimates.

Background & Summary

Renewable electricity has never been this cheap. More than 90% of the utility-scale renewable projects commissioned in 2025 generated power at a lower cost than the cheapest new fossil-fuel alternative [1]. Yet deployment remains uneven, and coal is still deeply embedded in power systems worldwide. Existing coal plants could continue operations and emissions for decades [2]. To meet the Paris climate goals, many plants would need to retire years earlier than expected coal plant lifetimes [3–5].

Equipment prices alone do not determine whether new renewables can compete with an operating coal unit. The answer also depends on the quality of local wind and solar resources and the financing rate for the upfront capital expenditures (CapEx) of those projects. Because renewable projects are capital-intensive, financing costs account for a substantial share of delivered electricity costs [6, 7]. Applying a single cost of capital across countries therefore obscures real differences [8], especially where scarce or expensive financing limits investment [9, 10].

Coal crossover analysis, therefore, asks a narrower question: how do operating expenses (OpEx) of an existing coal unit compare with the levelized cost—the discounted lifetime cost of both CapEx plus OpEx, divided by the discounted lifetime electricity output (see Equation (5))—of new renewable generation nearby? Previous studies have made this comparison for the United States, including local wind, solar, and storage options [11], and for a global sample of nearly 2,500 coal plants with four-hour storage [12]. A related literature examines early retirement [13], coal-fleet lock-in [14], stranded power assets [15], and uncertainty in plant-level transition choices [16]. These studies establish the value of plant-level analysis. They do not, however, provide a

common unit-level record in which the coal comparator, renewable resource, technology cost, and financing assumptions can all be inspected and changed.

Relevant data are available, but are scattered across sources and differ in format, definitions, units, and level of aggregation. Coal databases describe generating units; resource atlases organize values on geographic grids; cost reports use country, regional, or global averages; financing datasets have their own technology and vintage conventions. Public data products now cover renewable assets, energy-system models, supply regions, siting constraints, technology costs, and the cost of capital [17–23]. Combining them requires explicit choices about identifiers, spatial matching, units, currency years, and fallbacks, together with enough provenance to reconstruct those choices [24].

This Data Descriptor supplies that structure for 6,580 operating coal-fired generating units at 2,384 plants in 73 countries and makes it available at <https://coalcrossover.org> [*password: "GCT" (before publication)*] for ready exploration. We match every plant to nearby, same-country solar and onshore-wind resources adjusted for capacity factors, then calculate renewable generation costs using harmonized technology and financing inputs. The main table reports results for each coal unit, including estimated coal OpEx, solar and wind Levelized Costs of Electricity (LCOEs), assigned and break-even Weighted Average Cost of Capital (WACC), and the renewable capacity and investment required for replacement to match modeled annual coal generation at the selected average renewable capacity factor. It also identifies the least-cost technology, economic classification, and source tier. A separate plant-level table supports mapping and regional analysis while preserving links to the underlying units.

The package also preserves the variation behind the central estimate. Resource records span empirical percentiles P0–P100 at search radii of 30–60 km, and coal-price records retain the full empirical quantile curves. Separate tables cover prespecified resource cases, wind losses, renewable CapEx, financing support, and four-hour battery assumptions. Researchers can reproduce the reported comparisons, examine where a classification changes, or substitute alternative inputs without rebuilding the geospatial match from the beginning. Versioned inputs, documented transformations, tests, and independent recalculations make the same workflow available for future source releases.

Methods

Study design and observation grain

We built the dataset around a direct generation-cost comparison for each operating coal-fired unit. For each coal unit i , crossover occurs when at least one matched renewable technology has a lower modeled levelized cost (LCOE) than the unit’s OpEx alone:

$$\min(\text{LCOE}_{i,\text{solar}}, \text{LCOE}_{i,\text{wind}}) < \text{OpEx}_i. \quad (1)$$

The generating unit i is the scientific observation. This approach preserves differences in capacity, heat rate, service, and modeled utilization, even when multiple units are located at the same plant. Plant tables aggregate the unit records for mapping and regional summaries. Scenario tables hold the unit fixed and change one or more assumptions, allowing direct comparison with the central record.

Coal-unit population and reporting samples

The unit registry comes from the January 2026 release of the Global Energy Monitor Global Coal Plant Tracker [25]. We use the GEM unit or phase identifier as the unit key and the GEM location

identifier as the plant key. Records designated as operating form the study population; all other status categories are excluded.

This expanded operating fleet contains 6,580 units at 2,384 plants in 73 countries, with 2,202,528 MW of combined capacity. GEM modeled capacity factors and heat rates remain in the records, together with fields identifying them as modeled values. For the `grid_power_comparable` comparison sample, we remove units identified as captive or combined heat and power because their services are not directly comparable with grid electricity alone. The resulting sample comprises 2,836 units at 1,056 plants in 60 countries. The companion dashboard opens with the full operating fleet; formal manuscript summaries use the comparison sample unless noted otherwise.

Modeled generation and coal operating cost

Annual generation is modeled from unit capacity and the GEM capacity factor:

$$G_i = C_i \times 8,760 \times CF_i, \quad (2)$$

where C_i is capacity in megawatt (MW), CF_i is the modeled capacity factor, and 8,760 is the number of hours in a 365-day year. Dividing heat rate in Btu/kWh^{-1} by 1,000 gives MMBtu/MWh^{-1} ; multiplying that value by the assigned coal price gives fuel cost per MWh.

The coal comparator is the modeled cost of continued operation:

$$\text{OpEx}_i = \text{FuelCost}_i + \text{O\&M}_i, \quad (3)$$

with $\text{O\&M}_i$ as the operations & maintenance (O&M) cost assigned from the IEA/NEA table, in USD per MWh. Sunk construction expenditure and historical financing are not part of an ongoing operating decision. Carbon prices, decommissioning, contracts, subsidies, sustaining capital, and reliability services are also left outside the central comparator so that they can be introduced based on local conditions.

Coal prices are drawn from the World Bank Commodity Price Data monthly Australian and South African thermal-coal series for January 2023–December 2025 (`Monthly Prices`, cells F763:G798) [26]. We convert USD per metric tonne to USD per MMBtu using a heating value of $6,000 \text{ kcal/kg}^{-1}$. Units in Asia and Oceania receive the Australian benchmark; those in Europe and Africa receive the South African benchmark. For the Americas, we use the unweighted mean of the two. The median monthly price is the central case.

The coal-price scenarios capture temporal variation in a small set of international benchmarks, not the potentially persistent difference between those benchmarks and delivered fuel costs at individual plants. The resulting unit-level classifications and financing gaps should therefore be interpreted as benchmark-scenario outputs rather than estimates conditioned on plant-specific coal prices, particularly where mine-mouth supply, transport costs, contracts, coal quality, or domestic pricing create substantial differences.

The same 36 monthly observations define the coal-price response surface. Linear interpolation yields P0–P100 in one-percentage-point increments, with the observed endpoints retained at P0 and P100. We report P10, P50, and P90. Because the data records include the assigned series, percentile, conversion, and resulting fuel cost, users with local delivered-coal prices can replace the international benchmark without altering the rest of the calculation.

The values used for $\text{O\&M}_i$ come from compatible coal technologies in Table 3.12 of the IEA and OECD Nuclear Energy Agency datatables (rows 6–24, column N) [27]. Matching first uses country and technology. If no such observation exists, the hierarchy falls back to the median among the closest technology matches, then to the median of all compatible coal records. We convert the

reported 2018 USD values to constant 2025 USD with the World Bank United States GDP deflator [28]; the selected tier is stored for every unit.

Local renewable-resource matching

Solar input is the annual long-term-average photovoltaic output raster from Global Solar Atlas version 2.13 [29]. Onshore-wind input is the Global Wind Atlas 4.0 IEC Class II gross capacity-factor raster at 100 m hub height and 136 m rotor diameter [30]. Natural Earth boundaries identify the country associated with each plant and raster cell [31].

For each plant and technology, the pipeline collects valid cells in the same country at every integer radius from 30 to 60 km. It then stores the empirical distribution of those cells from P0 to P100, again in one-percentage-point steps. P0 is the lowest observed value in the search area and P100 the highest. The central case uses P90 at 45 km, while the prespecified comparisons retain P50, P90, and P100 at 45 km and P90 at 30, 45, and 60 km.

Daily photovoltaic output divided by 24 gives the solar capacity factor. Global Wind Atlas reports gross wind capacity factors, so the central calculation applies a 15% gross-to-net loss. That value follows the electrical and wake-effect loss assumption reported in Appendix H of the NREL *Wind Vision* report (p. 78) [32]. A separate response surface varies the loss from 10% to 20% in one-percentage-point increments; 10%, 15%, and 20% are the formal cases.

Renewable cost and financing inputs

IRENA’s *Renewable Power Generation Costs in 2025* data package supplies installed cost, annual all-in operating expenditure (OpEx), lifetime, and real after-tax weighted average cost of capital (WACC) [1]. Installed costs come from workbook Tables 2.1 and 3.1 and the global series in Figures 2.3 and 3.1. We assign a 2025 country or regional weighted average where one is reported, otherwise the relevant global weighted average. Solar and wind OpEx values are taken from Tables A3 and A4, and both technologies use the 25-year lifetime reported in Annex Table A1.

Financing inputs are country- and technology-specific real after-tax WACC values from Table A2. Although this table accompanies the 2025 cost report, it describes 2024 financing conditions; the source and model vintages are recorded separately. Missing countries receive the median reported WACC for the same technology, with the fallback marked in the output. We use the global renewable cost-of-capital dataset of Steffen et al. [22] as an independent validation source.

For technology τ , the capital-recovery factor is:

$$\text{CRF}_\tau = \frac{r_\tau(1 + r_\tau)^{n_\tau}}{(1 + r_\tau)^{n_\tau} - 1}, \quad (4)$$

where r_τ is real after-tax WACC and n_τ is the asset’s expected lifetime. At zero WACC, we use the continuous limit $1/n_\tau$. Renewable LCOE then is:

$$\text{LCOE}_\tau = \frac{\text{CapEx}_\tau \text{CRF}_\tau + \text{OpEx}_\tau}{8.76 CF_\tau}, \quad (5)$$

with CapEx in USD kW⁻¹, annual all-in OpEx in USD kW⁻¹ year⁻¹, and the result in constant 2025 USD MWh⁻¹. The denominator converts capacity factors CF_τ into annual MWh per kW. Required renewable capacity is derived from modeled coal generation and the matched capacity factor; multiplying that capacity by installed cost gives the corresponding investment estimate. This method follows the standard capital-recovery approach to levelized electricity cost [7, 33].

Economic-screen classification and financing threshold

The central comparison produces two modeled economic classes. An `already_crossed` unit has renewable LCOE below coal operating cost at its assigned IRENA WACC. A `finance_sensitive` unit does not cross at the assigned WACC but reaches parity at a lower, non-negative renewable WACC.

For each renewable technology τ , we numerically solve for the real after-tax WACC that sets renewable LCOE_τ equal to coal operating cost. The financing gap is the assigned WACC minus this break-even value for the lowest-cost renewable technology, expressed in percentage points; it is zero for units already below the threshold. We also identify coal-cost-sensitive units, defined as those whose crossover classification changes between the P10 and P90 coal-price scenarios.

To isolate financing, the WACC-support analysis reduces the assigned rate by ρ percentage points:

$$r_\tau^\rho = \max\left(0, r_\tau - \frac{\rho}{100}\right). \quad (6)$$

The lower bound prevents negative financing rates. At each value of ρ , solar and wind LCOE are recalculated, the cheaper technology is selected, and its cost is compared with the unchanged coal OpEx.

Scenario analysis and selected sensitivity cases

Exploratory controls and cases presented here draw from the same response surfaces. Renewable-resource records contain every integer search radius from 30 to 60 km and every empirical percentile from P0 to P100. Coal-price records likewise retain P0–P100. The dashboard evaluates a common renewable CapEx multiplier from 0.60 to 1.40 in increments of 0.05, together with the wind-loss and WACC grids described above. Formal comparisons use 30, 45, and 60 km; resource values P50, P90, and P100; wind losses of 10%, 15%, and 20%; coal prices P10, P50, and P90; and unscaled renewable CapEx. Each scenario identifier records the radius, resource percentile, wind loss, coal-price percentile, financing support, and CapEx multiplier needed to reconstruct it.

Four-hour battery cost sensitivity

Storage is added as a separate cost layer. The default capital-cost input is 164.44 USD kWh⁻¹ for a four-hour grid battery, which is the nominal 2025 value used in a companion PyPSA-Earth analysis [34]. The underlying battery series draws on historical lithium-ion cost data from Ziegler and Trancik [35], their archived data release [36], and BloombergNEF [37].

The companion analysis maps pack costs to stationary storage with a balance-of-system adder and a separate power component. Its 2050 median is 99.30 USD kWh⁻¹, with a P10–P90 range of 61.20–201.70 USD kWh⁻¹. For comparison, the IEA *World Energy Outlook 2025* reports 150 USD kWh⁻¹ in 2024 and 2050 values of 90 USD kWh⁻¹ under Current Policies and Stated Policies and 80 USD kWh⁻¹ under Net Zero Emissions, all in 2024 USD [38].

NREL supplies the technical assumptions used to turn capital cost into the levelized cost of storage (LCOS): a 15-year life, 85% round-trip efficiency, one full cycle per day, zero variable O&M, and fixed O&M equal to 4% of capital cost per kW-year [39]. Financing follows the paired renewable WACC. Storage LCOS is annualized battery capital plus fixed O&M divided by annual discharged energy, consistent with the general levelized-cost treatment of electricity storage [40, 41]. Charging cost is renewable LCOE divided by round-trip efficiency.

The prespecified tables route 0%, 25%, 50%, or 100% of delivered renewable energy through the battery. The resulting records report the storage cost added to each generation comparison.

200 Lazard’s 2026 LCOE+ report provides an external industry benchmark for generation and storage
201 costs [42].

202 **Provenance and software environment**

203 Original source files are preserved unchanged. The source manifest records release information,
204 retrieval date, byte size, integrity check, and SHA-256 hash; the build manifest adds method and
205 source-manifest hashes, software versions, output dimensions, and output hashes. The central build
206 used Python 3.12.13 with NumPy 2.2.6 [43], pandas 2.2.3 [44], PyArrow 20.0.0, Rasterio 1.4.3, SciPy
207 1.15.3 [45], and Shapely 2.1.1.

208 **Data Records**

209 The data package is centered on one record per operating coal unit. Plant summaries support
210 maps and regional comparisons. Resource surfaces, scenario tables, and validation files retain the
211 calculations behind each central estimate. Figure 1 shows the geographic coverage and the links
212 among these records.

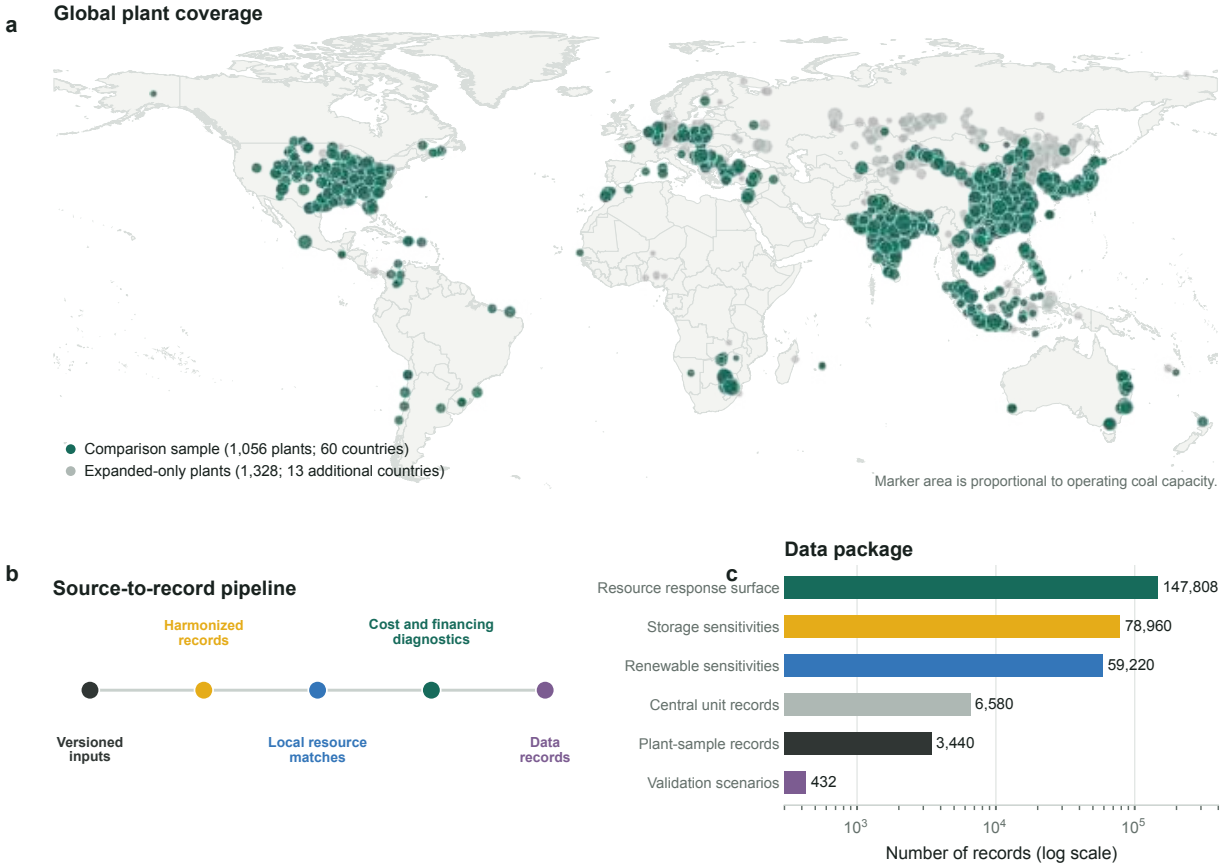


Figure 1: Coverage and organization of the dataset. **a**, Plants in the expanded operating fleet and the grid-power-comparable comparison sample; marker area scales with operating coal capacity. **b**, Data flow from registered source files through identifier and unit harmonization, same-country resource matching, cost calculation, and data records. **c**, Number of rows in the central, response-surface, sensitivity, and validation records. A logarithmic scale accommodates the larger scenario tables.

Central records

unit_results.parquet is the main scientific record. It contains 6,580 rows and 94 fields. A row identifies an operating unit and carries its plant location, service and sample flags, capacity, modeled generation and emissions, coal-cost inputs, matched solar and wind values, LCOE components, assigned and break-even WACC, renewable capacity and investment estimates, least-cost technology, economic class, and source tiers. The inputs needed to reconstruct the central comparison are therefore available on the same row as its result.

Two smaller tables support plant-level use. **plant_results.parquet** has 3,440 plant-sample rows and 64 fields: 1,056 for the comparison sample and 2,384 for the expanded fleet. It aggregates unit capacity, modeled generation and emissions, and economic classes. **plant_renewable_inputs.parquet** contains one solar and one wind record per plant (4,768 rows, 21 fields), including resource, capacity factor, installed cost, O&M, WACC, lifetime, LCOE, and source-assignment tier.

Response surfaces and prespecified sensitivity records

Nearby renewable resources are stored in **resource_matches.parquet**. Its 147,808 rows cover 2,384 plants, two technologies, and every integer search radius from 30 to 60 km. The 112 fields include the count of valid same-country raster cells and their empirical distribution from P0 to P100. The highest observed cell appears both as P100 and as **resource_max**.

Coal-price variation is much more compact. **coal_price_quantiles.csv** has 303 rows and seven fields: 101 quantiles for each of the Australian, South African, and Americas-average benchmarks. These curves come directly from the 36 monthly prices used in the coal model. Together, the resource and coal-price files allow continuous scenario analysis without changing the central unit table.

Formal renewable cases are written to **renewable_sensitivity_results.parquet**. Nine cases are retained for each operating unit, giving 59,220 rows and 34 fields: P90 at 30, 45, and 60 km; P50, P90, and P100 at 45 km; and wind gross-to-net losses of 10%, 15%, and 20%. The corresponding sample-level totals appear in **renewable_sensitivity_summary.csv** (18 rows, 13 fields).

Storage records cross three technology selections with the nominal 2025 battery capital cost and four stored-energy shares. **storage_sensitivity_results.parquet** contains all 12 combinations for every operating unit (78,960 rows, 44 fields). It reports battery duration, round-trip efficiency, storage and charging LCOS, renewable-plus-storage cost, power and energy capacity, and the adjusted crossover indicator. **storage_sensitivity_summary.csv** reduces these cases to 24 scenario-sample rows.

Table 1: Principal records in the data package.

File	Observation grain	Rows	Fields
unit_results.parquet	operating unit	6,580	94
plant_results.parquet	plant by reporting sample	3,440	64
plant_renewable_inputs.parquet	plant by technology	4,768	21
resource_matches.parquet	plant by technology and radius	147,808	112
coal_price_quantiles.csv	benchmark by percentile	303	7
renewable_sensitivity_results.parquet	unit by prespecified resource case	59,220	34
renewable_sensitivity_summary.csv	scenario by reporting sample	18	13
storage_sensitivity_results.parquet	unit by storage case	78,960	44
storage_sensitivity_summary.csv	scenario by reporting sample	24	20

Validation and provenance records

Three files document the independent numerical checks. `phase3_validation_sample.csv` identifies the deterministic 48-unit sample. `phase3_recalculation_results.csv` compares stored and recalculated central values, and `phase3_scenario_recalculation.csv` extends the check to 432 unit-scenario rows. Machine-readable manifests record the software environment, input versions, methods, file dimensions, output hashes, and validation coverage. For sources that cannot be redistributed, the manifest provides the citation, retrieval details, integrity check, and ingestion path for an authorized local copy.

Identifiers, missing values, and joins

Files join on `unit_id` or `plant_id`, which correspond to the GEM unit or phase identifier and GEM location identifier. Scenario and resource tables add technology, radius, sample, or scenario keys where needed. Missing numerical inputs remain null, and records that cannot be classified are labeled explicitly. Monetary fields are reported in constant 2025 USD, with units included in field names where practical.

Data Overview

The expanded data cover 2,384 operating coal plants in 73 countries. Of these, 1,056 plants in 60 countries meet the grid-power-comparable criteria used for the comparison sample. Solar is the lower-cost modeled renewable option at 822 comparison-sample plants; wind is lower at 234 (Fig. 2).

At unit level, 2,702 of 2,836 comparison-sample records have renewable LCOE below coal operating cost at the assigned WACC. The other 134 are finance-sensitive: they reach parity at a lower, non-negative WACC. Their required WACC reduction has a median of 2.44 percentage points and a 90th percentile of 4.89 percentage points. The largest modeled gap is 8.245 percentage points (Fig. 3).

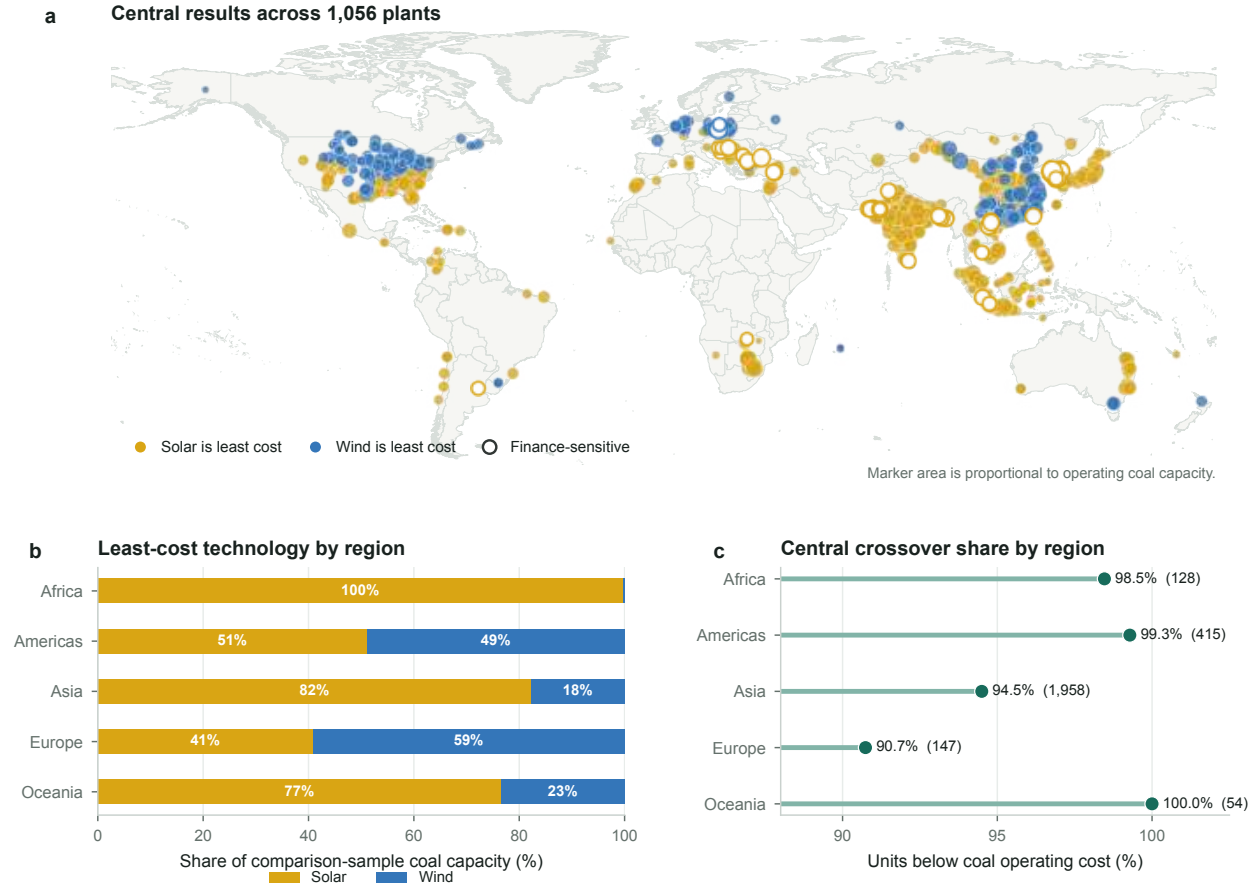


Figure 2: Central crossover patterns by location and region. **a**, Comparison-sample plants colored by the cheaper of local solar and onshore wind. Filled markers denote plants where renewable LCOE is below coal operating cost at the assigned WACC; open markers denote finance-sensitive plants. Marker area scales with operating coal capacity. **b**, Regional shares of comparison-sample plant capacity by least-cost renewable technology. **c**, Regional shares and counts of units below the coal operating-cost threshold.

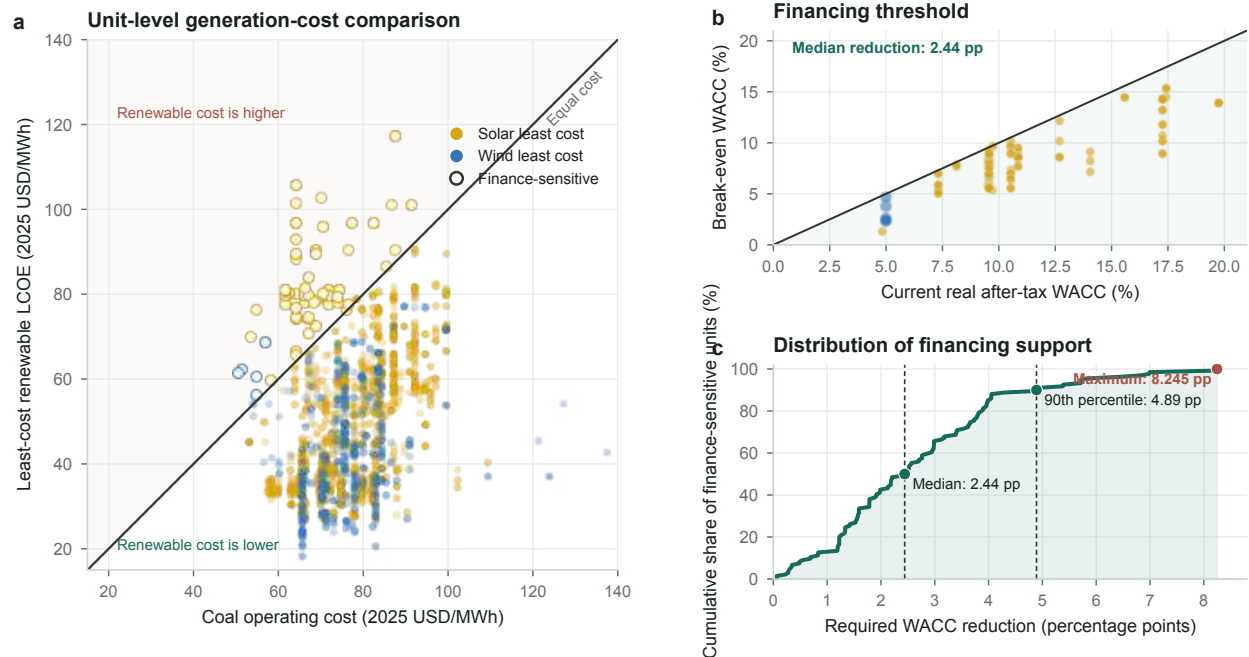


Figure 3: **Cost margins and financing thresholds at the unit level.** **a**, Coal operating cost plotted against the lower of solar and wind LCOE for 2,836 comparison-sample units. Filled points below the diagonal are already crossed at the assigned WACC; open points above it are finance-sensitive. Color indicates the least-cost renewable technology. **b**, Assigned and break-even real after-tax WACC for the 134 finance-sensitive units. **c**, Empirical cumulative distribution of the WACC reduction required to reach parity.

Technical Validation

We tested four aspects of the data separately: provenance, input coverage, numerical accuracy, and reproducibility. The checks combine source-file inspection, assignment diagnostics, automated tests, an isolated rebuild, independent recalculation, and sensitivity analysis.

Source integrity and provenance

Every primary input appears in the source manifest with its release or product description, retrieval date, byte size, integrity check, and SHA-256 hash. Archives and workbooks were opened or tested before normalization. For the selected Global Wind Atlas raster, both the official byte size and MD5 checksum matched. The manifest therefore identifies the precise input vintage used by the scientific build. Historical processed tables are kept only as comparison artifacts.

Financing-response and boundary checks

The WACC-support test changes financing and holds all other inputs fixed. In the comparison sample, the share of units below coal operating cost rises from 95.28% at assigned WACC to 97.28%, 99.29%, 99.79%, 99.93%, and 100% after reductions of 2, 4, 6, 8, and 8.5 percentage points (Fig. 4). For the expanded fleet, the corresponding sequence is 95.85%, 97.99%, 99.12%, 99.57%, 99.92%, and 100%.

The final comparison-sample crossover occurs at a reduction of approximately 8.245 percentage points. At the 8.5-percentage-point stress test, adjusted WACC has reached the 0% floor for 2,554 of 2,836 comparison-sample units (90.1%) and 5,826 of 6,580 expanded-fleet units (88.5%). Reporting this floor separately distinguishes a modeled crossover from an unbounded extrapolation of financing rates.

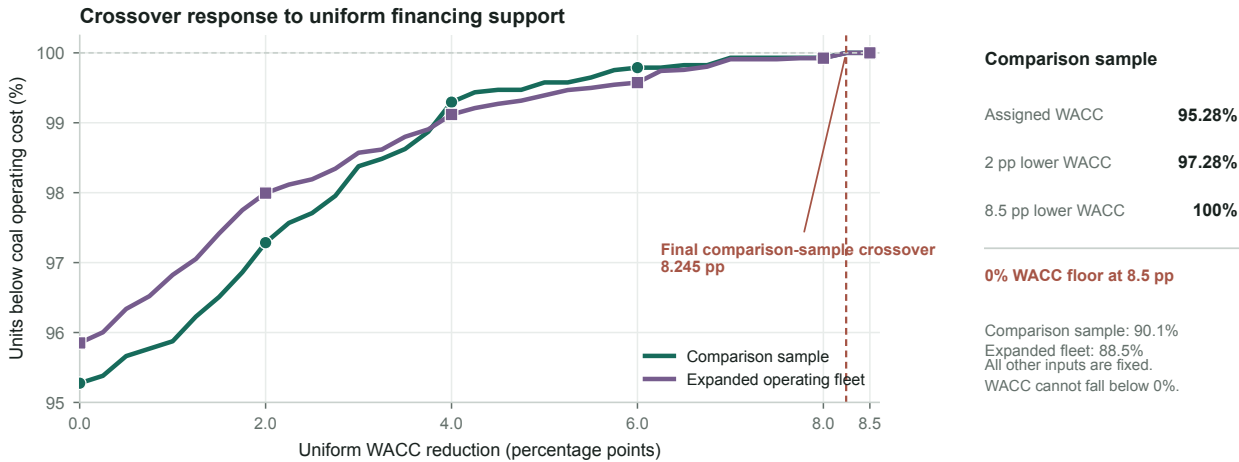


Figure 4: **Modeled crossover response to lower renewable weighted average cost of capital (WACC).** Comparison-sample and expanded-fleet shares below coal operating cost after a uniform WACC reduction, with every non-financing input held fixed. Markers identify reductions of 0, 2, 4, 6, 8, and 8.5 percentage points. Adjusted WACC cannot fall below 0%; at 8.5 percentage points, that floor binds for 90.1% of comparison-sample units and 88.5% of expanded-fleet units for the selected renewable technology.

Population and ingestion checks

The January 2026 GEM ingest produced 6,580 unique operating units totaling 2,202,528 MW. Applying the captive and CHP exclusions yielded the expected 2,836 comparison-sample units and 1,319,505 MW. Tests cover operating-status selection, unit-key uniqueness, plant and country counts, capacity totals, currency conversion, coal-technology matching, and each cost-source fallback. Assignment coverage is shown in Fig. 5. A country–technology coal O&M observation is available for 70.1% of comparison-sample units and 73.4% of the expanded fleet; the remaining records use a technology-specific median. Country-specific WACC covers 97.3% of plant–technology records for both solar and wind. Valid same-country cells are present for every wind match and for 2,379 of the 2,384 solar matches at each retained radius. The five unmatched solar plants retain null resource fields. Because the assignment tier travels with each result, analyses can be restricted to direct matches when desired.

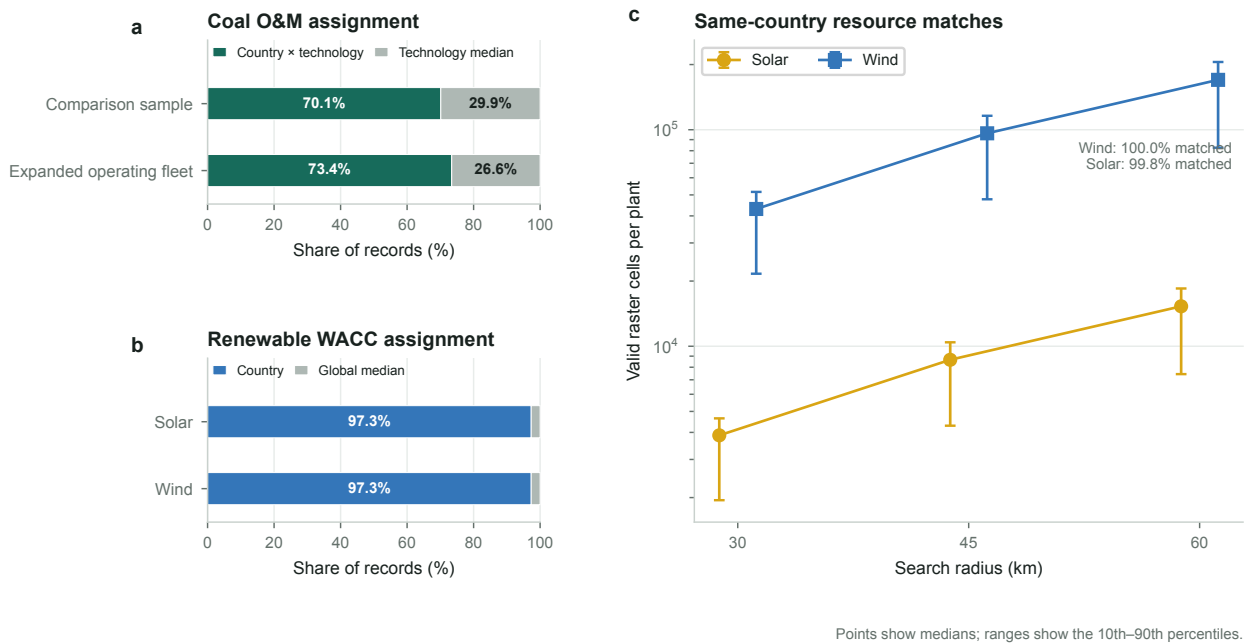


Figure 5: **Coverage of cost, financing, and resource inputs.** **a**, Coal O&M assignment tiers in the comparison sample ($n = 2,836$) and expanded operating fleet ($n = 6,580$). **b**, Country-specific and technology-median real after-tax WACC assignments for solar and wind ($n = 2,384$ plants per technology). **c**, Valid same-country raster-cell counts within 30, 45, and 60 km. Points show plant medians; intervals span the 10th–90th percentiles, and annotations report the share of 2,384 plants with at least one match.

Formula and unit tests

The release suite contains 51 automated tests, all of which passed. They cover the zero-WACC limit of the capital-recovery factor, renewable LCOE units, heat-rate and coal-price conversion, numerical break-even WACC, economic classification, same-country raster matching, storage energy accounting, strict JSON output, and agreement between the scientific and website records. Additional checks reproduce the formal resource and coal-price anchors from the continuous response surfaces and verify the public download package.

Isolated source-to-output rebuild

We reran the release-candidate pipeline in a temporary workspace containing the registered raw inputs, source code, method specifications, and historical comparison artifacts, but no intermediate files or prior outputs. A fresh Python 3.12.13 environment was created from the pinned requirements. The run repeated raster extraction and resource matching before rebuilding the central, sensitivity, storage, and validation tables.

Eighteen normalized-source, scientific-output, and validation files were byte-identical to the reference build. Four JSON records also agreed after generation timestamps, absolute root paths, and the invocation string were removed. Phase 2 validation, Phase 3 independent recalculation, and all 44 tests available at the release-candidate stage passed in the clean workspace. Seven tests were added subsequently, bringing the release suite to 51. These additions extend coverage of the empirical coal-price curve, continuous resource controls, public-source metadata, agreement between browser-facing response surfaces and the scientific records, and deterministic scenario-data compression. Build logs and file-hash comparisons accompany the validation records.

Independent recalculation

The deterministic validation sample contains 48 units at 48 plants in 23 countries across all five represented regions. Selection deliberately covers both economic classes, observations close to the crossover boundary, large units, high- and low-WACC cases, coal-cost-sensitive records, every formal renewable scenario family, storage-sensitive cases, missing solar resources, captive units, CHP units, and expanded-fleet records.

Separate validation functions reimplemented the capital-recovery factor, renewable LCOE, coal operating cost, classification, break-even WACC, required capacity and investment, modeled emissions, and the nominal 2025 battery case. They reproduced all 48 central records and all 432 unit-scenario records (Fig. 6). Source-trace checks also matched the associated GEM, IRENA, central renewable, and 45-km resource inputs. The largest residual at a numerical break-even WACC was 2.84×10^{-14} USD MWh⁻¹, compared with the prespecified tolerance of 10^{-6} USD MWh⁻¹.

Sensitivity behavior

Under the central 45-km P90 resource rule and 15% wind loss, 2,702 of 2,836 comparison-sample units are already crossed and 134 are finance-sensitive. The prespecified resource and wind-loss cases place the crossed share between 93.86% (45-km P50) and 98.27% (45-km P100). At P90, changing the search radius to 30 or 60 km gives 95.13% and 94.92%, respectively. Wind losses of 10% and 20% produce 95.31% and 95.03% (Fig. 7a). Automated parity checks confirm that P50, P90, and P100 in the continuous resource files reproduce these formal cases.

The WACC response is monotonic by construction: rates decline toward a 0% floor while every other input stays constant. A global reduction of 8.5 percentage points brings both reporting samples to 100% crossover in this cost screen, but the floor has already bound for 90.1% of comparison-sample units and 88.5% of the expanded fleet. Those two quantities are stored and reported separately.

Storage changes the comparison more sharply. At the nominal 2025 capital cost of 164.44 USD kWh⁻¹ and using each unit's least-cost renewable technology, the comparison-sample crossover share is 95.28% when no energy passes through storage. It falls to 80.75% at a 25% stored-energy share, 36.60% at 50%, and 0.42% at 100% (Fig. 7b). The storage table preserves each cost and sizing component rather than folding a single storage requirement into the central record.

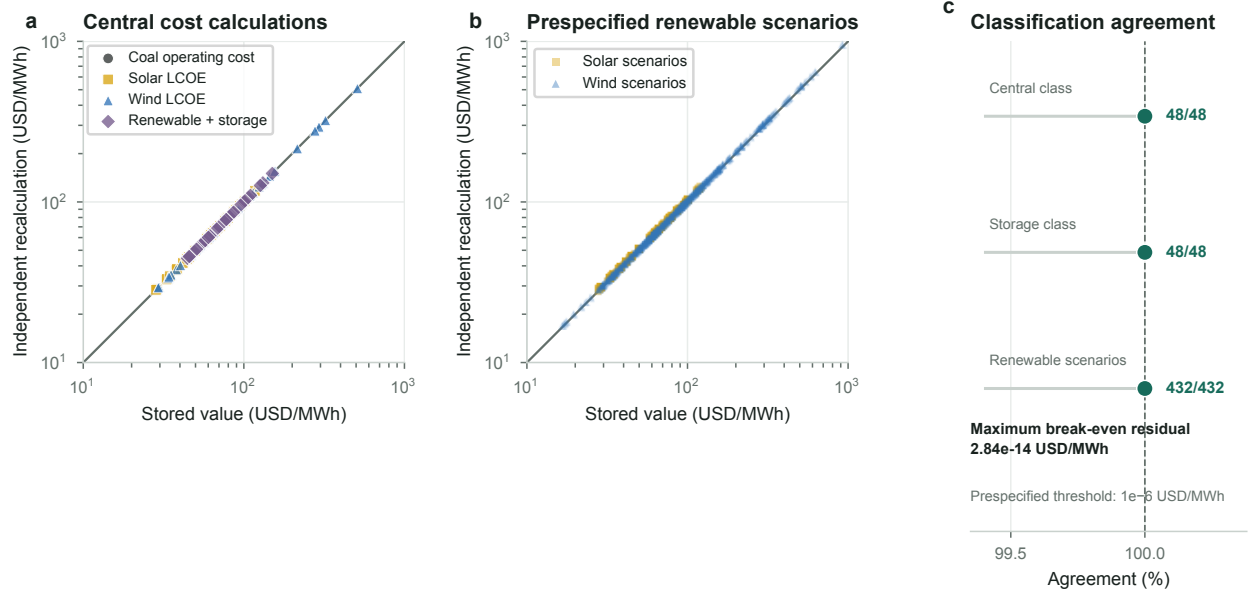


Figure 6: **Independent recalculation of stored outputs.** **a**, Stored and recalculated coal operating cost, solar LCOE, wind LCOE, and renewable-plus-storage cost for the 48-unit deterministic sample. Two units have no valid solar match, leaving 46 solar observations; the other metrics contain 48. **b**, Stored and recalculated solar and wind LCOE for 432 unit-scenario records (414 valid solar and 432 valid wind comparisons). Diagonal lines in **a** and **b** denote equality. **c**, Agreement in central, storage-adjusted, and renewable-scenario classes, with the maximum residual at numerical break-even WACC.

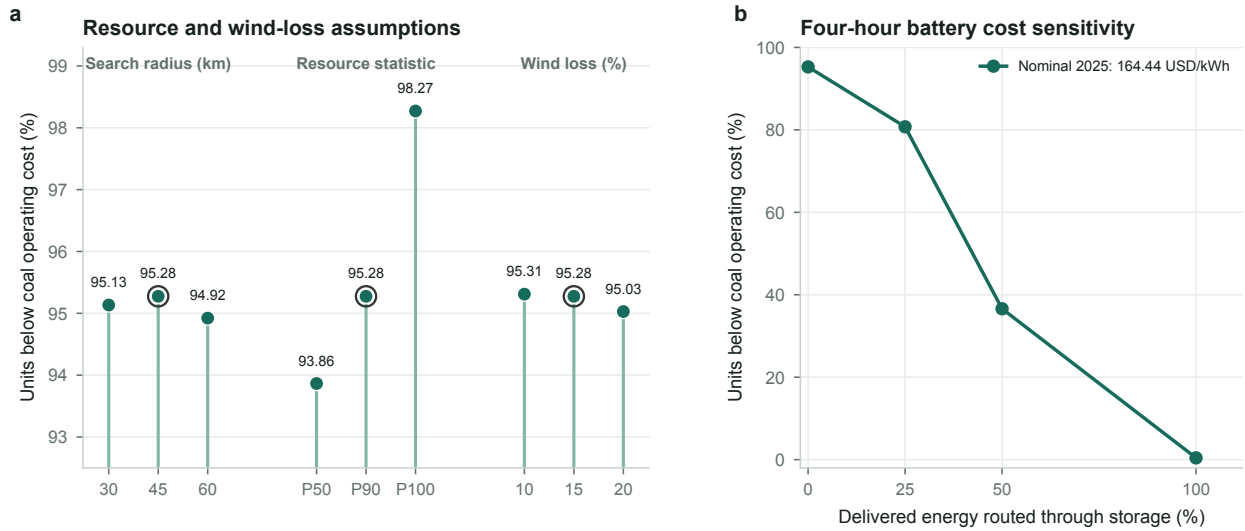


Figure 7: **Sensitivity of comparison-sample crossover estimates.** **a**, Share of 2,836 units below coal operating cost under the prespecified search-radius, resource-percentile, and wind-loss cases. Open rings mark the central 45-km P90 resource rule with a 15% wind loss. **b**, Crossover share when 0, 25, 50, or 100% of delivered renewable energy passes through a four-hour battery at the nominal 2025 capital cost of 164.44 USD kWh⁻¹. Storage calculations use each unit's least-cost renewable technology.

Usage Notes

For unit-level work, the natural starting point is `unit_results.parquet`. Selecting `grid_power_comparable` in `analysis_sample` returns the 2,836 units used in formal manuscript summaries. Keeping every row gives the expanded fleet of 6,580 operating units shown by default in the dashboard. Maps and country summaries can use the plant table, with `unit_id` providing the route back to individual generating units.

The central record uses P90 solar and wind resources within 45 km, a 15% wind gross-to-net loss, the median benchmark coal price, assigned renewable WACC, unscaled renewable CapEx, and no stored energy. Near the crossover boundary, the continuous cost margin and break-even WACC are more informative than the categorical label by itself.

Prespecified sensitivity files reproduce the cases reported here. The resource and coal-price surfaces allow finer choices based on the same raster extraction and monthly price observations. In the dashboard, WACC support can be scoped to a plant, city, country, or the EU27; rates never fall below 0%, and units outside the selected geography keep their assigned WACC. Renewable CapEx, stored-energy share, and four-hour battery cost can be varied independently.

Captive and combined heat and power units remain in the expanded file with their service flags intact. Users can therefore define a sample appropriate to grid electricity, captive supply, or heat-producing service. The source-tier fields are equally important: they distinguish direct country–technology assignments from fallbacks and make it possible to substitute local coal-price or O&M evidence.

These records compare modeled generation costs. Their stable plant and unit identifiers allow them to be joined with land, transmission, interconnection, dispatch, contract, retirement, labor, or heat-service data when a broader project or power-system analysis is required.

Data Availability

The dataset is archived on Zenodo at <https://doi.org/10.5281/zenodo.22092469>. For access during peer review, use the reviewer link. No login, password, or access request is required. In the Files section, select Download next to `coal_crossover_dataset_v1.0.0-rc1_review_2026-09-06.zip`.

The archive includes the main results, sensitivity and validation records, a data dictionary, and source and build manifests. This is version 1.0.0-rc1 (the 6 September 2026 data snapshot). Original third-party files are not included. Their citations, retrieval instructions, and checksums accompany the data; the code archive below includes the routines needed to import authorized local copies.

Code Availability

Development code is maintained at <https://github.com/Davidfu777/coal-crossover>. The repository is private during manuscript preparation. Version 1.0.0-rc1 is archived on Zenodo at <https://doi.org/10.5281/zenodo.22091620>. For access during peer review, use the reviewer link. No login, password, or access request is required. In the Files section, select Download next to `coal_crossover_code_v1.0.0-rc1.zip`.

The archive includes code for source ingestion, normalization, geospatial matching, central cost calculations, response surfaces, sensitivity tables, validation, data records, and all manuscript figures, together with the pinned software environment. Credentials, machine-specific paths, restricted inputs, deployment records, and website operations are excluded.

References

- [1] International Renewable Energy Agency. Renewable power generation costs in 2025. International Renewable Energy Agency, Abu Dhabi, 2026. URL <https://www.irena.org/Publications/2026/Jul/Renewable-Power-Generation-Costs-in-2025>. Report and accompanying data workbook retrieved 3 August 2026.
- [2] Dan Tong, Qiang Zhang, Yixuan Zheng, Ken Caldeira, Christine Shearer, Chaopeng Hong, Yue Qin, and Steven J. Davis. Committed emissions from existing energy infrastructure jeopardize 1.5 °C climate target. *Nature*, 572(7769):373–377, 2019. doi: 10.1038/s41586-019-1364-3.
- [3] Ryna Yiyun Cui, Nathan Hultman, Morgan R. Edwards, Linlang He, Arijit Sen, Kavita Surana, Haewon McJeon, Gokul Iyer, Pralit Patel, Sha Yu, Ted Nace, and Christine Shearer. Quantifying operational lifetimes for coal power plants under the Paris goals. *Nature Communications*, 10(1):4759, 2019. doi: 10.1038/s41467-019-12618-3.
- [4] Alexander Pfeiffer, Cameron Hepburn, Adrien Vogt-Schilb, and Ben Caldecott. Committed emissions from existing and planned power plants and asset stranding required to meet the paris agreement. *Environmental Research Letters*, 13(5):054019, 2018.
- [5] Ottmar Edenhofer, Jan Christoph Steckel, Michael Jakob, and Christoph Bertram. Reports of coal’s terminal decline may be exaggerated. *Environmental Research Letters*, 13(2):024019, 2018.
- [6] International Energy Agency. Financing clean energy transitions in emerging and developing economies. International Energy Agency, Paris, 2021. URL <https://www.iea.org/reports/financing-clean-energy-transitions-in-emerging-and-developing-economies>.
- [7] Janosch Ondraczek, Nadejda Komendantova, and Anthony Patt. WACC the dog: The effect of financing costs on the levelized cost of solar PV power. *Renewable Energy*, 75:888–898, 2015. doi: 10.1016/j.renene.2014.10.053.
- [8] Florian Egli, Bjarne Steffen, and Tobias S. Schmidt. Bias in energy system models with uniform cost of capital assumption. *Nature Communications*, 10:4588, 2019. doi: 10.1038/s41467-019-12468-z.
- [9] Nadia Ameli, Olivier Dessens, Matthew Winning, Jennifer Cronin, Hugues Chenet, Paul Drummond, Alvaro Calzadilla, Gabriel Anandarajah, and Michael Grubb. Higher cost of finance exacerbates a climate investment trap in developing economies. *Nature Communications*, 12:4046, 2021. doi: 10.1038/s41467-021-24305-3.
- [10] Matteo Calcaterra, Lara Aleluia Reis, Panagiotis Fragkos, Thomas Briera, Harmen Sytze de Boer, Florian Egli, Johannes Emmerling, Gokul Iyer, Shivika Mittal, Friedemann H. J. Polzin, Mark W. J. L. Sanders, Tobias S. Schmidt, Aleksandra Serebriakova, Bjarne Steffen, Dirk-Jan van de Ven, Detlef P. van Vuuren, Paul Waidelich, and Massimo Tavoni. Reducing the cost of capital to finance the energy transition in developing countries. *Nature Energy*, 9: 1241–1251, 2024. doi: 10.1038/s41560-024-01606-7.
- [11] Eric Gimon, Michelle Solomon, and Mike O’Boyle. Coal cost crossover 3.0. Energy Innovation report, 2023. URL <https://energyinnovation.org/report/the-coal-cost-crossover-3-0/>.

- [12] Paul Bodnar, Matthew Gray, Tamara Grbusic, Steve Herz, Amanda Lonsdale, Sam Mardell, Caroline Ott, Sriya Sundaresan, and Uday Varadarajan. How to retire early: Making accelerated coal phaseout feasible and just. Rocky Mountain Institute report, 2020. URL <https://rmi.org/insight/how-to-retire-early>.
- [13] Benjamin M. Kefford, Benjamin Ballinger, Diego R. Schmeda-Lopez, Chris Greig, and Simon Smart. The early retirement challenge for fossil fuel power plants in deep decarbonisation scenarios. *Energy Policy*, 119:294–306, 2018. doi: 10.1016/j.enpol.2018.04.018.
- [14] Achmed Edianto, Gregory Trencher, Niccolò Manych, and Kazuyo Matsubae. Forecasting coal power plant retirement ages and lock-in with random forest regression. *Patterns*, 4(7):100776, 2023. doi: 10.1016/j.patter.2023.100776.
- [15] Yangsiyu Lu, Francois Cohen, Stephen M. Smith, and Alexander Pfeiffer. Plant conversions and abatement technologies cannot prevent stranding of power plant assets in 2 °C scenarios. *Nature Communications*, 13(1):806, 2022. doi: 10.1038/s41467-022-28458-7.
- [16] Xizhe Yan, Dan Tong, Yixuan Zheng, Yang Liu, Shaoqing Chen, Xinying Qin, Chuchu Chen, Ruochong Xu, Jing Cheng, Qinren Shi, Dongsheng Zheng, Kebin He, Qiang Zhang, and Yu Lei. Cost-effectiveness uncertainty may bias the decision of coal power transitions in China. *Nature Communications*, 15(1):2272, 2024. doi: 10.1038/s41467-024-46549-5.
- [17] Sebastian Dunnett, Alessandro Sorichetta, Gail Taylor, and Felix Eigenbrod. Harmonised global datasets of wind and solar farm locations and power. *Scientific Data*, 7(1), 2020. doi: 10.1038/s41597-020-0469-8.
- [18] Trevor Barnes, Abhishek Shivakumar, Maarten Brinkerink, and Taco Niet. OSeMOSYS Global, an open-source, open data global electricity system model generator. *Scientific Data*, 9(1), 2022. doi: 10.1038/s41597-022-01737-0.
- [19] Rebecca Peters, Jürgen Berlekamp, Klement Tockner, and Christiane Zarfl. RePP Africa: a georeferenced and curated database on existing and proposed wind, solar, and hydropower plants. *Scientific Data*, 10(1), 2023. doi: 10.1038/s41597-022-01922-1.
- [20] Sebastian Sterl, Bilal Hussain, Asami Miketa, Yunshu Li, Bruno Merven, Mohammed Bassam Ben Ticha, Mohamed A. Eltahir Elabbas, Wim Thiery, and Daniel Russo. An all-Africa dataset of energy model “supply regions” for solar photovoltaic and wind power. *Scientific Data*, 9(1): 664, 2022. doi: 10.1038/s41597-022-01786-5.
- [21] Chris R. Vernon, Kendall Mongird, Kristian D. Nelson, and Jennie S. Rice. Harmonized geospatial data to support infrastructure siting feasibility planning for energy system transitions. *Scientific Data*, 10(1), 2023. doi: 10.1038/s41597-023-02694-y.
- [22] Bjarne Steffen, Florian Egli, Anurag Gumber, Mak Đukan, and Paul Waidelich. A global dataset of the cost of capital for renewable energy projects. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-05912-x.
- [23] Hadi Vatankhah Ghadim, Jannik Haas, Christian Breyer, Hans Christian Gils, Philip Odonkor, E. Grant Read, Mengzhu Xiao, and Rebecca Peer. Clean technology cost projections: investment and levelized costs of solar, wind, battery, and hydrogen. *Scientific Data*, 12(1), 2025. doi: 10.1038/s41597-025-05951-4.

- [24] Mark D. Wilkinson et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, 2016. doi: 10.1038/sdata.2016.18.
- [25] Global Energy Monitor. Global coal plant tracker, january 2026 release. Dataset and documentation, 2026. URL <https://globalenergymonitor.org/projects/global-coal-plant-tracker/>. Retrieved 3 August 2026.
- [26] World Bank. Commodity price data (the pink sheet). Monthly historical workbook, 2026. URL <https://www.worldbank.org/en/research/commodity-markets>. Retrieved 3 August 2026.
- [27] International Energy Agency and OECD Nuclear Energy Agency. Projected costs of generating electricity 2020. International Energy Agency, Paris, 2020. URL <https://www.iea.org/reports/projected-costs-of-generating-electricity-2020>.
- [28] World Bank. World development indicators: Gdp deflator (ny.gdp.defl.zs). Application programming interface dataset, 2026. URL <https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS>. Retrieved 3 August 2026.
- [29] World Bank Group and Solargis. Global solar atlas version 2.13: photovoltaic power output long-term average raster. Geospatial dataset, 2026. URL <https://globalsolaratlas.info/download/world>. Annual world layer retrieved 3 August 2026; data licensed under CC BY 4.0 with the attribution terms specified by the provider.
- [30] Technical University of Denmark and World Bank Group. Global wind atlas 4.0, 2025. IEC Class II gross capacity-factor raster at 100 m hub height.
- [31] Natural Earth. Admin 0 countries, 1:10m cultural vectors, version 5.1.1. Geospatial dataset, 2022. URL <https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-admin-0-countries/>.
- [32] National Renewable Energy Laboratory. Wind vision: A new era for wind power in the united states, appendices. National Renewable Energy Laboratory, 2015. URL <https://docs.nrel.gov/docs/fy15osti/63197-3.pdf>.
- [33] K. Branker, M. J. M. Pathak, and J. M. Pearce. A review of solar photovoltaic levelized cost of electricity. *Renewable and Sustainable Energy Reviews*, 15(9):4470–4482, 2011. doi: 10.1016/j.rser.2011.07.104.
- [34] Theo Moers, Davide Fioriti, and Gernot Wagner. Innovation uncertainty shapes divergent pathways toward decarbonization. Manuscript in preparation, 2026.
- [35] Micah S. Ziegler and Jessika E. Trancik. Re-examining rates of lithium-ion battery technology improvement and cost decline. *Energy & Environmental Science*, 14(4):1635–1651, 2021. doi: 10.1039/D0EE02681F.
- [36] Micah S. Ziegler and Jessika E. Trancik. Data series for lithium-ion battery technologies, 2021.
- [37] BloombergNEF. Lithium-ion battery pack prices fall to \$108 per kilowatt-hour, despite rising metal prices, 2025. URL <https://about.bnef.com/insights/clean-transport/lithium-ion-battery-pack-prices-fall-to-108-per-kilowatt-hour-despite-rising-metal-prices-bloombergnef/>.

- [38] International Energy Agency. World energy outlook 2025. International Energy Agency, 2025. URL <https://www.iea.org/reports/world-energy-outlook-2025>.
- [39] National Renewable Energy Laboratory. Cost projections for utility-scale battery storage: 2025 update. National Renewable Energy Laboratory, 2025.
- [40] Verena Jülch. Comparison of electricity storage options using levelized cost of storage (LCOS) method. *Applied Energy*, 183:1594–1606, 2016. doi: 10.1016/j.apenergy.2016.08.165.
- [41] Oliver Schmidt, Sylvain Melchior, Adam Hawkes, and Iain Staffell. Projecting the future levelized cost of electricity storage technologies. *Joule*, 3(1):81–100, 2019. doi: 10.1016/j.joule.2018.12.008.
- [42] Lazard. Levelized cost of energy+. Lazard report, 2026. URL <https://www.lazard.com/research-insights/levelized-cost-of-energyplus-lcoeplus/>.
- [43] Charles R. Harris et al. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.
- [44] Wes McKinney. Data structures for statistical computing in Python. In *Proceedings of the Python in Science Conference*, pages 56–61, 2010. doi: 10.25080/Majora-92bf1922-00a.
- [45] Pauli Virtanen et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3):261–272, 2020. doi: 10.1038/s41592-019-0686-2.

Author contributions

LL and GW conceived of the study. LL designed and performed the initial analysis for her Columbia University undergraduate project; HF designed and performed the present analysis. All authors helped interpret the results. HF and GW drafted the manuscript. GW acquired funding and supervised the project.

Acknowledgements

LL thanks the Columbia Undergraduate Business Scholars (CUBS) Program for financial support; LL and HF thank Columbia Business School’s Climate Knowledge Initiative (CKI) for financial support. We thank Connor Butterfield for excellent research assistance, and Sam Fankhauser for valuable feedback.

Competing interests

The authors declare no competing interests.