# Joint uncertainty analysis of the life-cycle carbon footprint of gas, wind and solar electricity in Italy: an open-data assessment with a stress-test of machine-learning surrogates

Kaleb Kassa[a]

[a] *University of Padova, M.Sc. Energy Engineering, Padova, Italy*

**Abstract**

We estimate the life-cycle greenhouse-gas emissions of 1 kWh delivered to the Italian grid by a natural-gas combined-cycle plant, onshore wind and utility-scale photovoltaics (PV), using only open data and a transparent parametric model built in Brightway. The three systems are sampled *jointly*: each of 10,000 Monte Carlo runs draws shared background factors (steel, aluminium, electricity grids) once for all systems. Median results are 473 (95 % interval 435–516), 25.2 (18.1–38.4) and 14.8 (9.7–21.1) g $CO_2$-eq/kWh for gas, PV and wind, all inside the IPCC AR5 literature ranges. Wind is below PV in 99.5 % of runs, but this depends on a dated PV mounting-structure inventory: with a structure four times lighter the figure falls to 81.9 %. Sobol indices computed for the *difference* between PV and wind show that a shared uncertainty can dominate each result yet cancel in the comparison: the steel emission factor has a total-order index of 0.41 for wind and 0.19 for PV, but only 0.03 for their difference. A first-order sharing factor $R = (a - b)^2/(a^2 + b^2)$ explains this ($R = 0.001$ for steel) and predicts the exact first-order shares ($\rho = 0.92$). We then use the exact model as ground truth for machine learning. Five surrogates of the PV model reach at best $R^2 = 0.81$ with 1,000 training runs and lose accuracy outside their training range, most for the flexible models. SHAP and permutation importance identify the right top drivers (Spearman $\rho = 0.86$–$1.00$) but misstate their variance shares by up to 0.16. With correlated inputs, graded against exact Shapley effects, SHAP variance shares stay within a total-variation distance of about 0.12, whereas permutation importance and Sobol indices that ignore the dependence degrade to 0.20 and 0.15. An interactive explorer accompanies the paper as supplementary material.

*Keywords:* life cycle assessment, electricity generation, uncertainty analysis, global sensitivity analysis, machine-learning surrogate, explainable machine learning

## 1. Introduction

Comparisons of electricity technologies are among the most common uses of life-cycle assessment (LCA). Two methodological issues limit how far such comparisons can be trusted. First, the technologies share background processes, and sampling them independently misrepresents the uncertainty of the comparison, as established in the LCA literature [1, 2]. Second, machine-learning (ML) surrogates and their explanations, for example SHAP [3], are increasingly used to explore LCA uncertainty, but are rarely checked against an exact answer.

Dependent sampling has been shown to matter in comparative LCA [4, 5, 6], and sensitivity analyses aimed at the comparison rather than at each alternative have been proposed [5, 7, 8]. Parameterised models with Sobol analysis are available as open tools [9, 10, 11], and Kim et al. [12] recently combined correlated sampling with gradient-boosted trees and SHAP to rank the inputs of large LCA models. To our knowledge, none of these studies checks how faithfully ML-based importance measures reproduce exact variance-based indices, or how far ML surrogates can be trusted outside their training range.

This study addresses these gaps with a deliberately compact, fully open model of three technologies for Italy in which the exact answer is always available. Its contributions are: (i) a joint-sampling case study of gas, wind and PV in Italy with open data only, including a Sobol decomposition of the *difference* between technologies that quantifies, parameter by parameter, how much of the importance of a shared uncertainty disappears in the comparison, together with a first-order shar-

ing factor that predicts which parameters cancel; (ii) a controlled test of how faithfully SHAP and permutation importance reproduce exact Sobol shares (and, for correlated inputs, exact Shapley effects), motivated by the recent use of SHAP for LCA sensitivity analysis [12]; and (iii) a stress-test of five ML surrogates inside and outside their training range. These claims rest on a documented but single-index literature search (protocol in the project repository). All results are reproducible from the project repository, and an interactive explorer allows readers to vary the key assumptions themselves.

## 2. Goal, scope and data

The study follows the structure of ISO 14040/44 [13, 14]; no critical review was performed. The functional unit is 1 kWh delivered to the Italian grid (plant gate; transmission losses excluded). The systems are a natural-gas combined-cycle plant (CCGT) representative of Italian main-activity electricity-only plants, an onshore wind plant, and a ground-mounted crystalline-silicon PV plant. The impact category is climate change, GWP100 with IPCC AR6 factors (fossil methane for fugitive emissions, non-fossil methane for combustion methane) [15]. End of life follows the cut-off approach (no recycling credits). Other impact categories, grid integration (storage, back-up) and prospective scenarios are out of scope.

No licensed database (e.g. ecoinvent) was used. Each of the 46 model parameters is recorded with its distribution and source (Appendix A), and every raw file is logged with URL, retrieval date, checksum and licence. Licences were checked before any automated processing; one candidate source (the 2026 IEA PVPS inventories) was rejected because its licence prohibits text and data mining. Table 1 lists the main sources.

## 3. Methods

Fig. 1 summarises the pipeline, from open data to verified and explained results.

### 3.1. Parametric models

Each system is one equation (all quantities per kWh delivered):

$$\mathrm{GWP}_{\mathrm{gas}} = \frac{3.6}{\eta}\,(EF_{\mathrm{comb}} + EF_{\mathrm{supply}}), \tag{1}$$

$$\mathrm{GWP}_{\mathrm{wind}} = \frac{(1+u)\sum_m M_m\,EF_m}{CF \cdot 8760 \cdot L \cdot 1000}, \tag{2}$$

$$\mathrm{GWP}_{\mathrm{PV}} = \frac{B_{\mathrm{chain}} + B_{\mathrm{mat}} + B_{\mathrm{BOS}}}{P\,H\,PR\,(L - dL^2/2)}, \tag{3}$$

$$B_{\mathrm{chain}} = E\,I_{\mathrm{CN}} + Q\,EF_{\mathrm{gas}} + C_{\mathrm{MG}}, \quad B_{\mathrm{mat}} = \sum_m M_m EF_m, \tag{4}$$

where $\eta$ is plant efficiency; $M_m$ material masses per MW (wind) or per $m^2$ of panel (PV); $CF$ capacity factor; $L$ lifetime; $E$ and $Q$ electricity and heat along the silicon chain (metallurgical silicon → solar-grade silicon → ingot → wafer → cell → panel); $I_{\mathrm{CN}}$ the Chinese grid intensity; $C_{\mathrm{MG}}$ the process $CO_2$ of silicon reduction; $B_{\mathrm{BOS}}$ mounting and balance of system; $P$ module power per $m^2$; $H$ in-plane irradiation; $PR$ performance ratio; $d$ linear degradation, integrated over the lifetime. Background factors are cradle-to-gate $CO_2$-eq per unit.

The models are implemented twice: in Brightway [31] with a project-owned biosphere and GWP100 method, used for the baseline and contribution analysis, and in vectorised NumPy for the uncertainty analyses. The two agree to machine precision at the baseline and on random parameter sets.

### 3.2. Wind calibration

Open material factors miss manufacturing energy, transport, installation and electronics. Run with each Vestas plant's own capacity factor and lifetime, our materials-only wind model gives 59 % of the manufacturer's result without end-of-life credit. We therefore calibrate an uplift $u$ = Vestas/ours − 1 (0.64–0.77 across the three plants) and sample it empirically. Calibrating on two plants predicts the third within 6 % (leave-one-out).

### 3.3. Joint Monte Carlo and global sensitivity

Parameters are sampled once per run and shared by every system that uses them (10,000 runs). Variance-based sensitivity uses Saltelli sampling with the first-order estimator of Saltelli et al. [32] and the total-order estimator of Jansen [33] (8,192 base samples, bootstrap confidence intervals, outputs centred). The implementation reproduces the analytic indices of the Ishigami function [34] within 0.02. Indices are computed for each system and for the difference PV−wind.

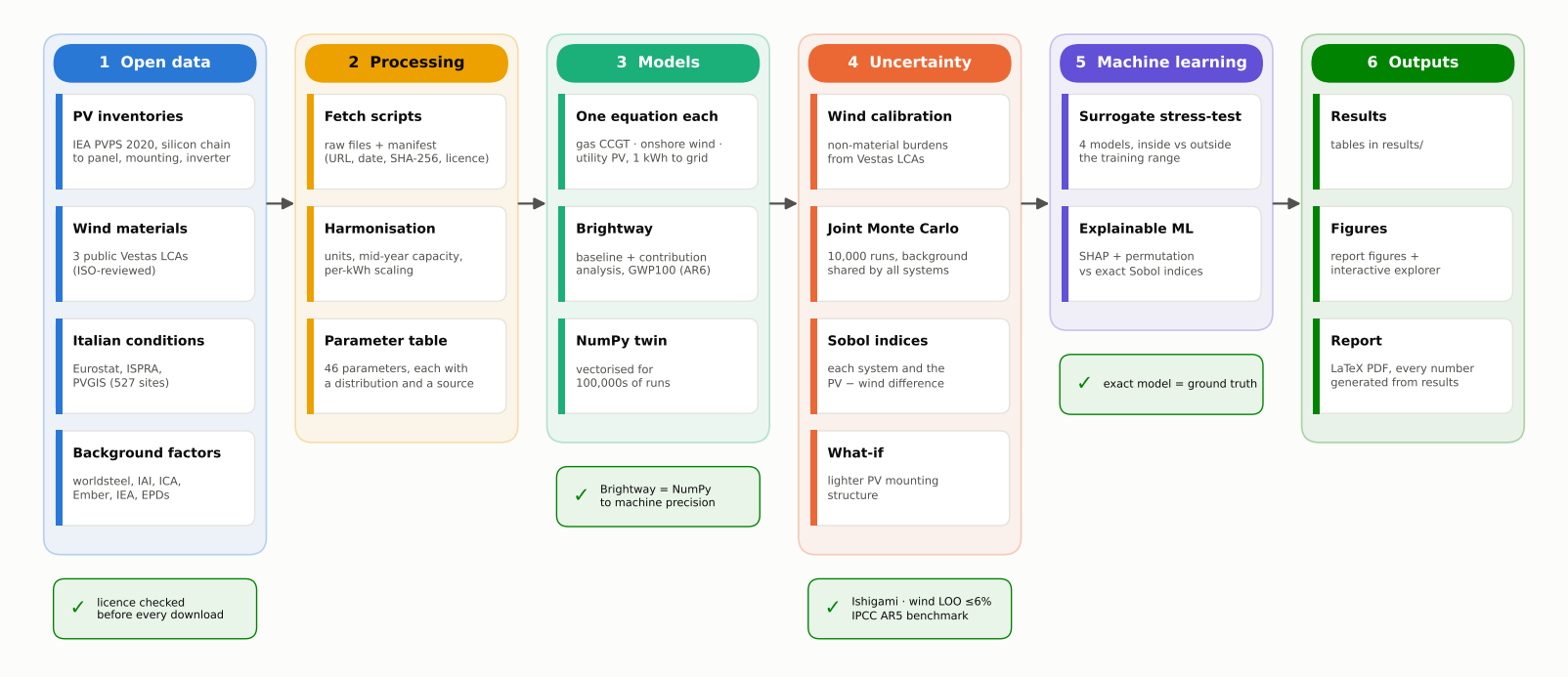



Figure 1: Study pipeline. Green badges mark the verification and validation checks.

Table 1: Main data sources. All are open or freely available; licences are recorded in the project's raw-data manifest.

| Item | Source and processing |
|---|---|
| PV silicon chain, panel, BOS | IEA PVPS 2020 unit-process inventories, Chinese supply chain [16]; pedigree uncertainty (SD95) used as lognormal spread |
| PV lifetime, degradation, PR | IEA PVPS methodology guidelines [17] |
| PV site irradiation | PVGIS 5.3, 527-point grid over Italy, optimal tilt, in-plane irradiation 1,157–2,151 kWh/m$^2$/yr [18] |
| Wind materials | Three public, ISO-reviewed Vestas LCAs (V162-6.2, V150-6.0, V163-4.5) [19, 20, 21] |
| Wind capacity factor | Eurostat, Italian fleet 2015–2024 (0.19–0.22), computed on mid-year capacity [22] |
| Gas plant efficiency | Eurostat energy balances: gross output over fuel input of main-activity electricity-only plants, 2015–2024 (0.53–0.55) [22] |
| Gas combustion and supply | ISPRA national $CO_2$ factor [23]; IEA supply-chain intensities [24] |
| Grid for PV manufacturing | Ember life-cycle grid intensity, China, 2015–2024 [25] |
| Materials | worldsteel [26], IAI [27], ICA [28], Taha et al. [29]; concrete, glass and PET from EPD and eco-profile listings |
| Benchmark | IPCC AR5 Annex III, Table A.III.2 [30] |

### 3.4. Sharing factor, tail sensitivity and robustness checks

Consider two systems $A$ and $B$ that depend on the same parameter $\theta_i$, with local sensitivities $a_i = \partial A/\partial\theta_i$ and $b_i = \partial B/\partial\theta_i$ and standard deviation $\sigma_i$. To first order and for independent inputs, the variance that $\theta_i$ adds to the difference $A - B$ is $(a_i - b_i)^2\sigma_i^2$ if both systems share one draw and $(a_i^2 + b_i^2)\sigma_i^2$ if they draw independently. Their ratio, the *sharing factor* $R_i = (a_i - b_i)^2/(a_i^2 + b_i^2)$, is 0 for perfect cancellation, 1 when sharing changes nothing (an unshared parameter) and up to 2 when the sensitivities have opposite signs. Sensitivities are central finite differences at the baseline; the linearised shares of the variance of PV−wind are compared with the exact first-order Sobol indices, and the Monte Carlo is repeated with independent draws for the two systems.

*Tail sensitivity.* Sobol total-order indices are computed for threshold-excess outputs, $\max(y - q_{90}, 0)$ for wind and PV, and $\max(q_{10}(D) - D, 0)$ with $D$ = PV − wind, using the same parameters as the whole-distribution analysis (16,384 base samples). *Robustness.* The spreads of single-source factors (assumption A1: ±20 % triangular) are narrowed, widened or made uniform, and all parametric spreads are halved and doubled (scaled lower bounds are floored at 10 % of the central value), and the medians, P(wind<PV) and the steel index are recomputed.

### 3.5. Surrogate stress-test and explainable ML

Five surrogates (linear regression, a sparse degree-2 polynomial-chaos expansion in an orthonormal Legendre basis fitted by Lasso, random forest [35], gradient boosting and Gaussian process; scikit-learn [36]) are trained on 100–1,000 runs and tested (a) inside the training range, (b) outside it (trained on the central 80 % of every parameter's quantile range, tested where at least one parameter lies outside), and (c) on the Sobol indices they imply. Separately, gradient-boosting and random-forest surrogates are explained with TreeSHAP [37] (mean |SHAP| and SHAP variance share) and permutation importance, and each explanation is graded against the exact Sobol total-order shares.

### 3.6. Explanations under correlated inputs

To test the explanations where they are used in practice, the PV−wind model is reduced to its eight most influential parameters (the others at central values) and three pairs are made dependent through a Gaussian copula with $\rho \in \{0, 0.3, 0.6, 0.9\}$: the steel and aluminium factors ($+\rho$, shared energy mix of materials production), site irradiation and performance ratio ($-\rho$, hot sunny sites derate) and the PV and wind lifetimes ($+\rho$, common durability assumptions). The pairs are assumptions, not estimates from data; marginals are unchanged. The reference answer is the Shapley effects [38], which are defined for dependent inputs, are non-negative and sum to one. They are estimated by nested Monte Carlo with exact Gaussian conditional draws (2,000 outer and 200 inner samples for each of the 254 subsets). The estimator reproduces the analytic effects of a dependent linear-Gaussian model (largest error 0.01) and, at $\rho = 0$, lies within the first- and total-order Sobol indices up to 0.03; the seed-to-seed noise is 0.02, so differences below about 0.03 are not resolved. Gradient boosting

(3,000 draws; $R^2$ 0.86–0.88 on 1,500 test draws; three repetitions) is explained with TreeSHAP in the interventional mode and in the path-dependent mode (the default of `shap.TreeExplainer`), each summarised as mean $|\phi|$ shares and as covariance shares $\mathrm{Cov}(\phi_i, f)/\mathrm{Var}(f)$, and with permutation importance. The baseline ignores the dependence (total-order Sobol indices of the independent model, normalised). Explanations are graded by total-variation distance to the exact effects, Spearman correlation and top-three agreement. Shapley effects are one definition of the reference answer: interventional SHAP answers a related, marginal question, so gaps are read as differences between questions as well as errors.

### *3.7. Interactive explorer*

Results are also provided as a self-contained interactive page (Supplementary material S1, Fig. 10). It contains a clickable map of the 527 PVGIS sites, live what-if sliders for ten key parameters (each spanning the parameter's own uncertainty range), and the Monte Carlo, Sobol and explainable-ML views. The model equations are re-implemented in JavaScript; on load the page checks its results against the Python/Brightway values.

## 4. Results and discussion

### *4.1. Baseline and contributions*

At central parameter values the model gives 472.1, 25.1 and 17.6 g $CO_2$-eq/kWh for gas, PV and wind (Fig. 2). Combustion is 84 % of the gas total. Steel is the largest source for both renewables: 63 % for wind and 42 % for PV, followed for PV by Chinese grid electricity in the silicon chain (27 %).

### *4.2. Uncertainty and comparison*

Table 2 and Fig. 3 summarise the joint Monte Carlo. Gas emits about 19 times more than PV and 32 times more than wind (medians). Wind is below PV in 99.5 % of runs, and the PV-to-wind ratio has a median of 1.7. Shared background factors correlate the PV and wind results ($r = 0.28$). All three medians fall inside the IPCC AR5 ranges; PV is below the IPCC median, consistent with Italy's high irradiation, and wind is above it, consistent with the low capacity factor of the Italian fleet.

*Mounting-structure what-if.*. The only open ground-mount inventory describes a heavy, dated structure. Scaling its steel and concrete to 25 % lowers the PV median to 17.9 g $CO_2$-eq/kWh and the probability that wind beats PV to 81.9 % (93.9 % at 50 %). The ranking holds, but its certainty depends on this dataset.

### *4.3. What drives the uncertainty*

Gas uncertainty is dominated by supply-chain emissions (total-order index 0.92). For wind, plant lifetime (0.44) and the steel factor (0.41) dominate; for PV, mounting steel (0.41), site irradiation (0.30) and the steel factor (0.19). Fig. 4 shows the key comparative result: the steel factor, important for each renewable on its own, explains only 0.03 of the variance of their difference, because a single draw moves both systems in the same direction. The comparison is instead driven by PV mounting steel (0.39), PV irradiation (0.30) and wind lifetime (0.15). Independent per-technology studies cannot show this.

### *4.4. Why shared parameters cancel*

The sharing factor (Fig. 5a) is 0.001 for steel, because PV and wind have almost the same absolute sensitivity to the steel factor and a common draw moves both together; it is 0.08 for polymers, 0.19 for copper, 0.28 for aluminium and 0.35 for concrete. Linearised variance shares of PV−wind predict the exact first-order Sobol indices with Spearman $\rho = 0.92$ and a largest error of 0.10; the main deviation is wind lifetime, whose effect is non-linear (Fig. 5b). Ignoring the sharing would widen the PV−wind difference by 15 % (standard deviation 6.0 instead of 5.2 g $CO_2$-eq/kWh) and lower P(wind<PV) from 99.5 % to 98.3 % (Fig. 5c). The rule is elementary, but it turns the cancellation from an observation into a prediction that can be made before any Sobol analysis: shared parameters matter for a decision only where the absolute sensitivities of the two systems differ.

### *4.5. Tail sensitivity*

The parameters that matter in the tail are the same as for the whole distribution (Spearman ≥ 0.99), but their weights shift (Fig. 6). The clearest change is for the runs in which PV comes closest to wind: the total-order index of wind lifetime rises from 0.15 to 0.49, because a long-lived wind plant is what narrows the gap. An analysis of the average therefore understates the parameter that matters most for the closest calls.

Table 2: Life-cycle GHG intensity (g $CO_2$-eq/kWh): this study (joint Monte Carlo, 10,000 runs) and IPCC AR5.

| System | Median | 95 % interval | IPCC AR5 |
|---|---|---|---|
| Gas CCGT | 473 | 435–516 | 490 (410–650) |
| Utility PV | 25.2 | 18.1–38.4 | 48 (18–180) |
| Onshore wind | 14.8 | 9.7–21.1 | 11 (7–56) |

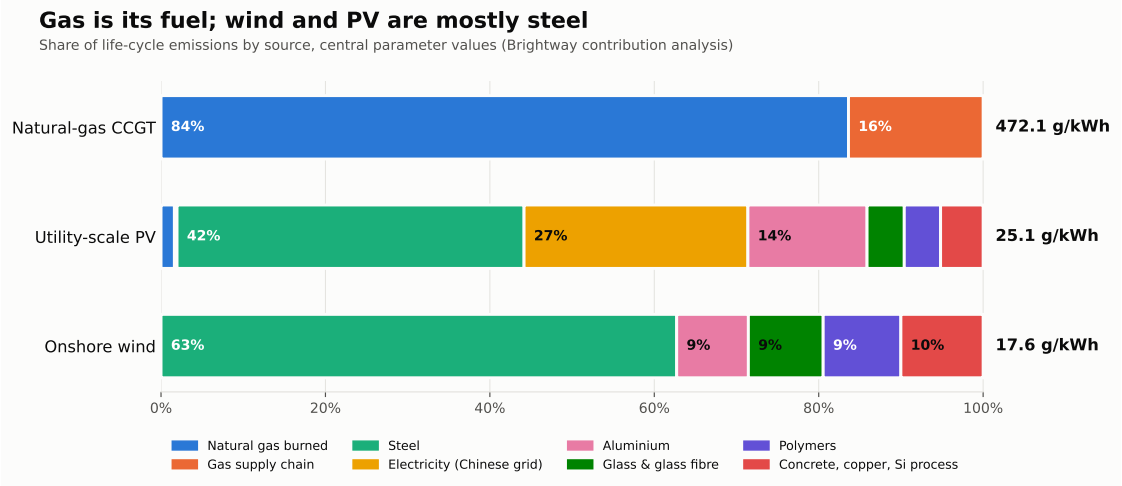



Figure 2: Contributions by source at central parameter values (Brightway contribution analysis).

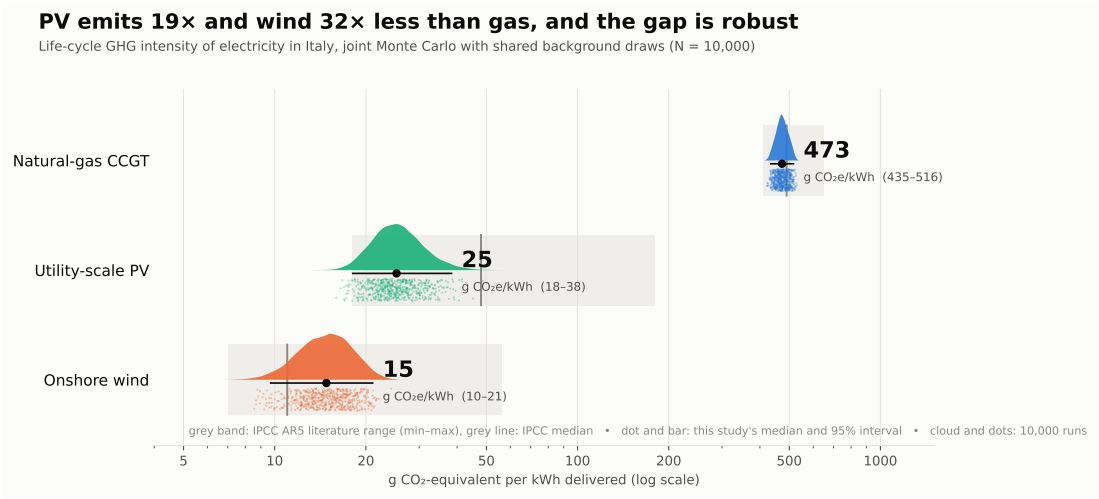



Figure 3: Joint Monte Carlo distributions on a common log scale, with IPCC AR5 ranges (grey bands) and medians (grey lines).

### *4.6. Robustness to the distributional assumptions*

Narrowing, widening or changing the shape of the spread assumed for single-source factors (glass, glass fibre, copper) leaves every result in Table 3 unchanged to the digits shown, because these factors contribute little. Halving all spreads raises P(wind<PV) to 100.0 %; doubling them lowers it to 97.0 % and raises the steel index of the difference to 0.08, so cancellation weakens but persists. The same three parameters remain the main drivers of the difference in every scenario (PV mounting steel, site irradiation and wind lifetime); only their order changes.

### *4.7. Surrogate stress-test*

With 1,000 training runs on the 32-parameter PV model, no surrogate exceeded $R^2 = 0.81$ inside the training range (Fig. 7); the relative RMSE was 8.6 % for gradient boosting, 8.8 % for the Gaussian process, 9.8 % for linear regression and 10.8 % for random forest. Outside the training range the flexible models lost most: $R^2$ fell from 0.81 to 0.62 (gradient boosting), from 0.70 to 0.60 (random forest) and from 0.80 to 0.68 (Gaussian process), whereas the linear model changed little (0.75 to 0.73). The sparse polynomial-chaos surrogate behaved like the linear one ($R^2$ 0.76 to 0.72; relative RMSE 9.5 % to 10.7 %), with the advantage that its Sobol indices follow analytically from the coefficients (they agreed with sampled indices within 0.013). The

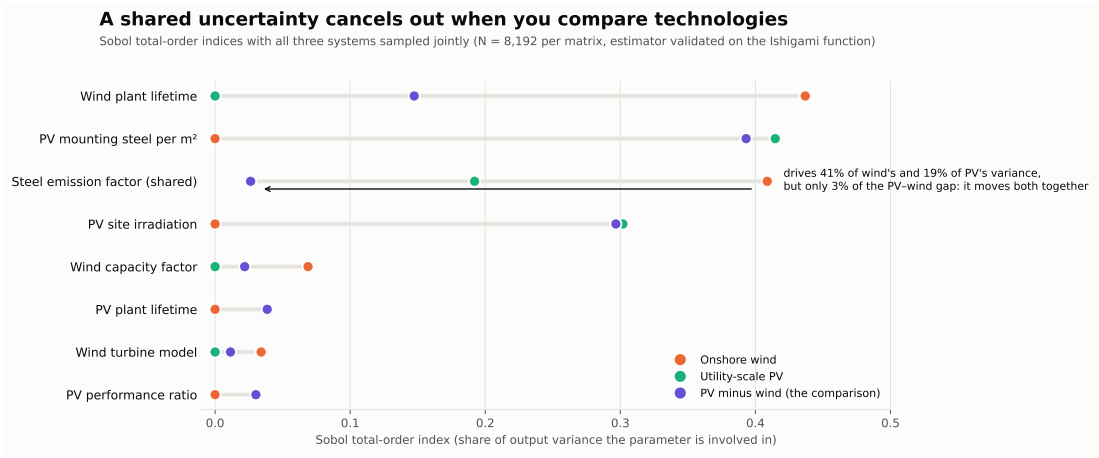



Figure 4: Sobol total-order indices for wind, PV and their difference.

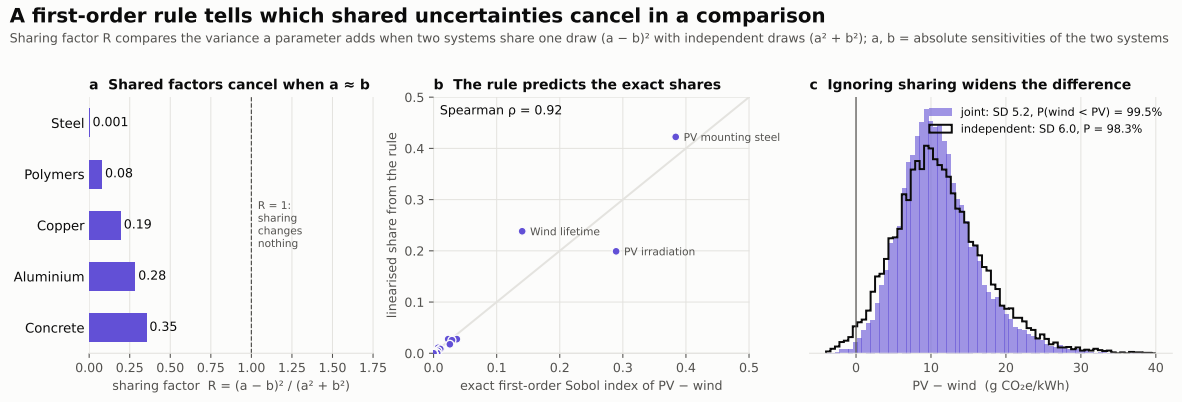



Figure 5: (a) Sharing factor of the parameters used by both PV and wind. (b) Linearised share of the variance of PV−wind predicted by the rule against the exact first-order Sobol index. (c) Distribution of PV−wind with shared and with independent draws.

largest error in a total-order index implied by a surrogate ranged from 0.13 (gradient boosting) to 0.23 (random forest). A surrogate-derived sensitivity analysis is therefore only as reliable as the surrogate itself, and the near-linear structure of the model explains why the linear baseline is hard to beat.

### *4.8. Explainable ML against exact sensitivity*

With gradient boosting trained on 1,000 runs, all three explanation methods recover the top-three drivers (100 % agreement) and rank parameters similarly to Sobol (Spearman $\rho = 0.86$–$1.00$; permutation importance $\geq 0.94$). Their importance *shares*, however, deviate by up to 0.16 (0.25 across all surrogates), and they systematically overstate PV mounting steel and understate irradiation (Fig. 8). They do detect the cancellation of the steel factor in the PV−wind comparison, ranking it 10th–14th against an exact rank of 6. In practice, ML explanations are suitable for deciding *which* parameters to investigate, but not for stating *how much* of the uncertainty each one explains.

### *4.9. Explanations under correlated inputs*

Correlation changes who receives the credit but not how uncertain the comparison is: the standard deviation of PV−wind is 5.0 at $\rho = 0$ and 5.0 g $CO_2$-eq/kWh at $\rho = 0.9$ (Fig. 9a). The exact Shapley effects, however, move. The effect of the performance ratio rises from 0.03 to 0.14 because it carries information about irradiation, and that of wind lifetime falls from 0.16 to 0.05 because the effects of the two lifetimes on the difference have opposite signs and offset each other when they are positively correlated.

Ignoring the dependence is accurate at $\rho = 0$ (distance 0.03, within noise) but degrades steadily to 0.15 at $\rho = 0.9$, and it credits the proxy (performance ratio)

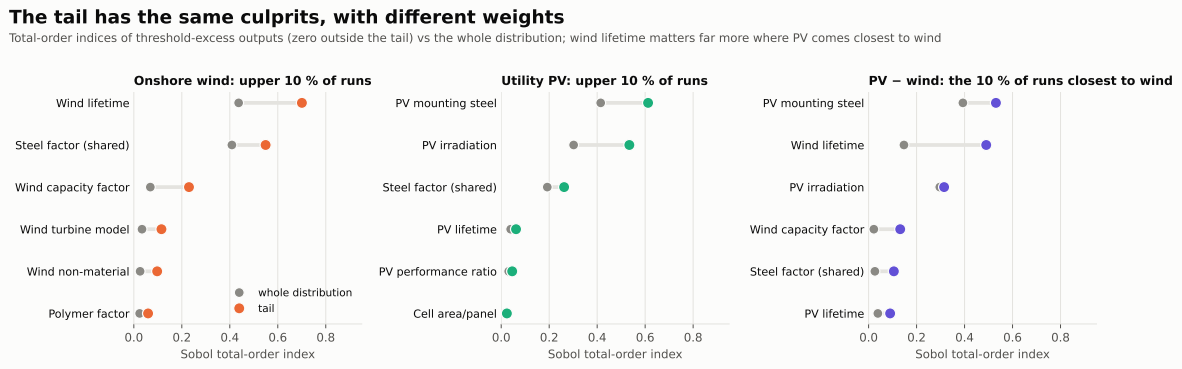



Figure 6: Total-order Sobol indices of threshold-excess outputs (tail) compared with the whole-distribution indices.

Table 3: Robustness of the main conclusions to the assumed spreads. Last column: Sobol total-order index of the steel factor for wind / PV / PV−wind.

| Scenario | P(wind<PV) | PV median (95 % interval), g $CO_2$-eq/kWh | Steel index |
|---|---|---|---|
| Baseline | 99.5 % | 25.2 (18.1–38.4) | 0.41 / 0.19 / 0.03 |
| Single-source spread ±10 % | 99.5 % | 25.3 (18.1–38.3) | 0.41 / 0.19 / 0.03 |
| Single-source spread ±40 % | 99.5 % | 25.2 (18.1–38.3) | 0.41 / 0.19 / 0.03 |
| Single-source uniform ±20 % | 99.5 % | 25.2 (18.1–38.3) | 0.41 / 0.19 / 0.03 |
| All spreads ×0.5 | 100.0 % | 25.1 (20.2–34.4) | 0.27 / 0.09 / 0.00 |
| All spreads ×2 | 97.0 % | 27.6 (16.2–58.4) | 0.35 / 0.18 / 0.08 |

with only 0.03 against an exact 0.14 (Fig. 9b,c). Permutation importance degrades in the same way (0.11 to 0.20), because shuffling one feature breaks the correlation and cannot credit its proxy (0.04). TreeSHAP summarised as covariance shares stays near 0.12–0.12 at every $\rho$, keeps the top three parameters right at every $\rho$ and credits the proxy (0.10); this supports its use for correlated LCA inputs [12]. Mean $|\phi|$ shares, the usual summary plot, are much further from variance shares (0.25–0.25), and their top-three agreement falls to 0.67 at $\rho = 0.9$. The two TreeSHAP modes are indistinguishable here (differences within the noise), so the default path-dependent mode is not worse than the interventional one in this model. The best explanation still deviates by up to 0.08 for a single parameter. Finally, the linear rule extended to dependent inputs fails (distance 0.37–0.33 from the exact effects), so we do not claim it.

### *4.10. Interactive explorer*

The explorer (Fig. 10) lets readers test the conclusions against their own assumptions. For example, moving the site-irradiation slider (or clicking a site on the map) shows the PV footprint ranging by about a quarter across Italy's regions, and lowering the mounting steel shows how the PV result approaches the wind result.

### *4.11. Limitations*

The study trades completeness for transparency: one equation per system and open data only. Its main limitations are: (i) PV inventories describe 2018–2020 technology; current modules are more efficient and made with less energy, so the PV result is likely conservative; (ii) the ground-mount dataset is heavy and dated, and it is the largest single PV driver; (iii) several background factors are single values from industry associations or secondary listings (concrete, glass, PET), given an assumed ±20 % triangular spread when no range was available; (iv) the wind uplift is calibrated on one manufacturer's LCAs; (v) the Italian wind capacity factor is a fleet average, conservative for new plants; (vi) process chemicals, metallisation pastes, transport in the silicon chain and gas-plant construction are cut off; (vii) Italy's LNG share is not modelled separately; (viii) the sharing rule is first-order and assumes independent inputs, so it is least accurate for strongly non-linear parameters such as lifetimes, and the degree-2 polynomial surrogate cannot capture the ratio structure of the model; (ix) the correlated pairs are assumed, the exact Shap-

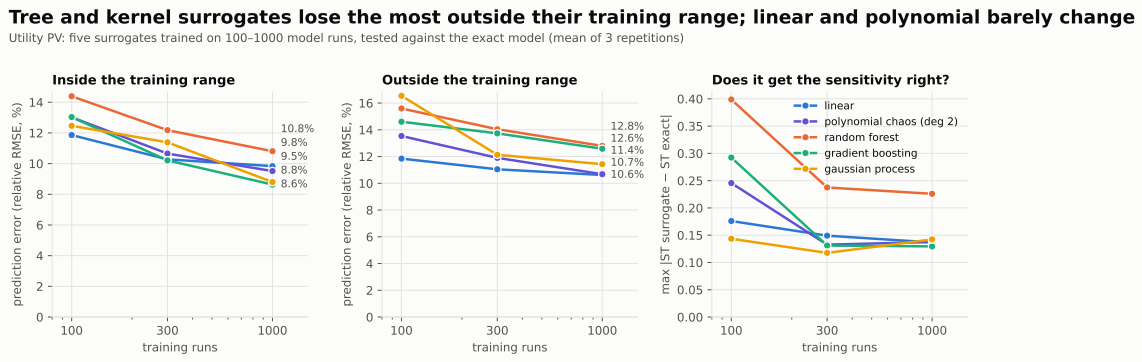



Figure 7: Surrogate prediction error inside and outside the training range, and fidelity of the implied sensitivity indices (PV).

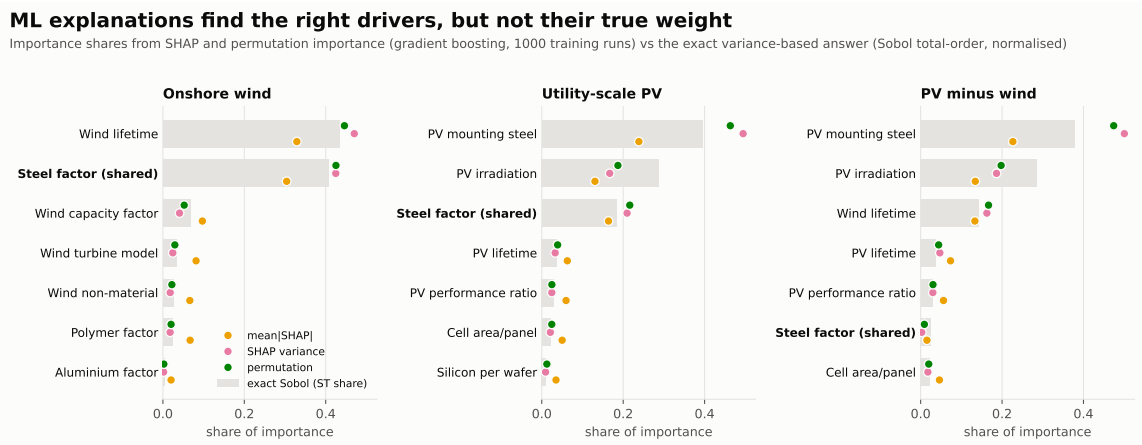



Figure 8: Importance shares from SHAP and permutation importance compared with exact Sobol total-order shares.

ley effects carry Monte Carlo noise of about 0.02, and the correlated-input test uses a reduced eight-parameter model and one surrogate type. None of these changes the gas-versus-renewables conclusion; (ii) affects the certainty of the PV-versus-wind ranking.

## 5. Conclusions

With open data alone, a compact parametric LCA reproduces the IPCC ranges for gas, wind and PV in Italy and shows that renewables emit one to two orders of magnitude less than gas under all sampled uncertainty. Sampling the technologies jointly reveals that uncertainties shared between them can dominate each result yet barely affect the comparison, which is what decisions depend on; a first-order sharing factor predicts which ones, and the conclusions hold when the assumed spreads are halved or doubled. ML surrogates and their explanations are useful for ranking drivers but should not be trusted for the size of their contributions, or outside their training range, without an exact check. When inputs are correlated, TreeSHAP summarised as variance shares was the most reliable of the explanations tested.

## Declaration of competing interest

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

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Data availability

All code, processed data, figures and this manuscript are generated by scripts in the project repository (`src/`);

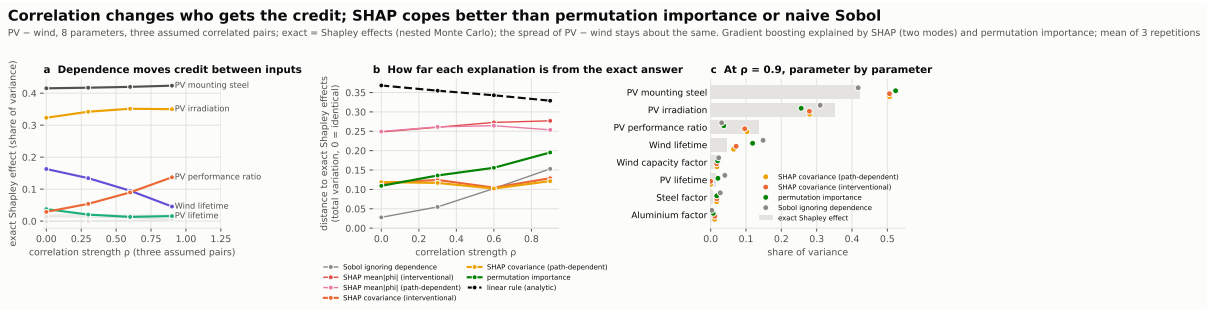



Figure 9: Explanations under correlated inputs for PV−wind. (a) Exact Shapley effects against correlation strength. (b) Total-variation distance between each explanation and the exact effects. (c) Shares per parameter at $\rho = 0.9$.

raw files are listed with URL, date, checksum and licence in `data/raw-manifest.csv`. Numbers in this manuscript are inserted automatically from the result files. No licensed data were used.

## Data attributions and disclaimer

This work contains data from Eurostat (datasets nrg_inf_epcrw, nrg_bal_peh and nrg_bal_c; CC BY 4.0), Ember (yearly electricity data and methodology; CC BY 4.0), the IEA (*Assessing emissions from LNG supply and abatement options*, 2026; CC BY 4.0) and PVGIS of the European Commission's Joint Research Centre. Capacity factors, efficiencies, yields and all other quantities derived from these sources were calculated or modified by the author; Eurostat, Ember, the IEA and the European Commission are not responsible for the derived values or for their use. Values taken from reports of the IEA PVPS, Vestas, worldsteel, the International Aluminium Institute (which owns its datasets and may revise them), the International Copper Association, the IPCC and the GHG Protocol are cited where used (Table 1 and Appendix Appendix A).

## Supplementary material

S1: interactive explorer (`figures/explorer.html`, with `explorer_data.js` and the bundled Plotly library; opens in any browser, works offline). S2: parameter table with distributions and sources (`data/processed/parameters.csv`).

## References


[1] E. A. Groen, R. Heijungs, Ignoring correlation in uncertainty and sensitivity analysis in life cycle assessment: what is the risk?, Environmental Impact Assessment Review 62 (2017) 98–109.

[2] Y. Qin, S. Suh, Does the use of pre-calculated uncertainty values change the conclusions of comparative life cycle assessments? – an empirical analysis, PLOS ONE 13 (12) (2018) e0209474. doi:10.1371/journal.pone.0209474.

[3] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems 30, 2017.

[4] P. J. G. Henriksson, R. Heijungs, et al., Product carbon footprints and their uncertainties in comparative decision contexts, PLoS ONE (2015). doi:10.1371/journal.pone.0121221.

[5] J. R. Gregory, A. Noshadravan, E. A. Olivetti, R. E. Kirchain, A methodology for robust comparative life cycle assessments incorporating uncertainty, Environmental Science & Technology (2016). doi:10.1021/acs.est.5b04969.

[6] A. Mendoza Beltrán, V. Prado, D. Font Vivanco, P. J. G. Henriksson, et al., Quantified uncertainties in comparative life cycle assessment: What can be concluded?, Environmental Science & Technology (2018). doi:10.1021/acs.est.7b06365.

[7] D. Ravikumar, T. P. Seager, S. Cucurachi, V. Prado, et al., Novel method of sensitivity analysis improves the prioritization of research in anticipatory life cycle assessment of emerging technologies, Environmental Science & Technology (2018). doi:10.1021/acs.est.7b04517.

[8] D. Marchese, M. Bates, J. M. Keisler, et al., Value of information analysis for life cycle assessment: Uncertain emissions in the green manufacturing of electronic tablets, Journal of Cleaner Production (2018). doi:10.1016/j.jclepro.2018.06.113.

[9] R. Jolivet, J. Clavreul, R. Brière, et al., lca_algebraic: a library bringing symbolic calculus to LCA for comprehensive sensitivity analysis, The International Journal of Life Cycle Assessment (2021). doi:10.1007/s11367-021-01993-z.

[10] S. Cucurachi, C. F. Blanco, et al., Implementation of uncertainty analysis and moment-independent global sensitivity analysis for full-scale life cycle assessment models, Journal of Industrial Ecology (2021). doi:10.1111/jiec.13194.

[11] M. Douziech, G. Ravier, R. Jolivet, et al., How far can life cycle assessment be simplified? a protocol to generate simple and accurate models for the assessment of energy systems and its application to heat production from enhanced geothermal systems, Environmental Science & Technology (2021). doi:10.1021/acs.est.0c06751.

[12] A. Kim, C. Mutel, S. Hellweg, Global sensitivity analysis of correlated uncertainties in life cycle assessment, Journal of Industrial Ecology (2025). doi:10.1111/jiec.70036.

[13] ISO, ISO 14040:2006 environmental management – life cycle assessment – principles and framework, Tech. rep., International

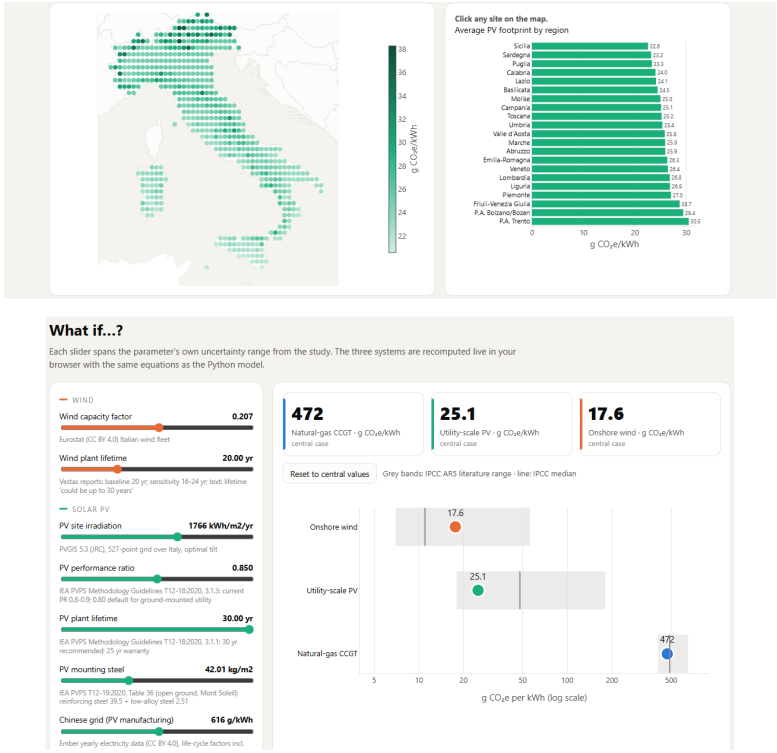



Figure 10: Interactive explorer (Supplementary material S1). Top: PV footprint at the 527 PVGIS sites, coloured by value, with regional averages. Bottom: what-if sliders and live results for the three systems against the IPCC ranges.


Organization for Standardization, Geneva (2006).

[14] ISO, ISO 14044:2006 environmental management – life cycle assessment – requirements and guidelines, Tech. rep., International Organization for Standardization, Geneva (2006).

[15] GHG Protocol, Global warming potential values (august 2024), aR6 values from IPCC AR6 WG1 Chapter 7 (Smith et al., 2021, Supplementary Material) (2024).

[16] R. Frischknecht, P. Stolz, L. Krebs, M. de Wild-Scholten, P. Sinha, Life cycle inventories and life cycle assessments of photovoltaic systems, Tech. Rep. T12-19:2020, International Energy Agency (IEA) PVPS Task 12, iSBN 978-3-907281-14-7 (2020).

[17] IEA PVPS Task 12, Methodology guidelines on life cycle assessment of photovoltaic, 4th edition, Tech. Rep. T12-18:2020, International Energy Agency Photovoltaic Power Systems Programme, iSBN 978-3-906042-99-2 (2020).

[18] European Commission, Joint Research Centre, Photovoltaic geographical information system (PVGIS) 5.3, PVcalc non-interactive service, SARAH3 radiation database, accessed 24 November 2026 (2026).

[19] S. Mali, P. Garrett, Life cycle assessment of electricity production from an onshore EnVentus V162-6.2 MW wind plant, Tech. rep., Vestas Wind Systems A/S (2023).

[20] S. Mali, P. Garrett, Life cycle assessment of electricity production from an onshore EnVentus V150-6.0 MW wind plant, Tech. rep., Vestas Wind Systems A/S (2023).

[21] O. Martinez Aguado, P. Garrett, Life cycle assessment of electricity production from an onshore V163-4.5 MW wind plant, Tech. rep., Vestas Wind Systems A/S (2025).

[22] Eurostat, Datasets nrg_inf_epcrw, nrg_bal_peh and nrg_bal_c (italy), dissemination API, accessed 24 November 2026, CC BY 4.0 (2026).

[23] ISPRA, Le emissioni di $CO_2$ nel settore elettrico nazionale e regionale, Tech. Rep. Rapporti 413/2025, Istituto Superiore per la Protezione e la Ricerca Ambientale, iSBN 978-88-448-1256-0 (2025).

[24] IEA, Assessing emissions from LNG supply and abatement options, Tech. rep., International Energy Agency, Paris, cC BY 4.0 (2026).

[25] Ember, Yearly electricity data (global) and electricity data methodology, accessed 24 November 2026, CC BY 4.0 (2026).

[26] worldsteel, Sustainability indicators 2023 report, Tech. rep., World Steel Association (2023).

[27] International Aluminium Institute, Global aluminium industry greenhouse gas emissions intensity reduction continues, with total emissions below 2020 peak, press release and statistics pages, accessed 24 November 2026 (2024).

[28] International Copper Association, Copper environmental profile, global 2023, Tech. rep., International Copper Association (2023).

[29] M. Taha, S. Vouros, K. Kyprianidis, Evaluation of environmental and economic impact of wind turbine blade manufacture at life-cycle level, in: Proceedings of the Second SIMS EUROSIM Conference on Modelling and Simulation, SIMS EUROSIM 2024, Linköping Electronic Conference Proceedings, 2025. doi:10.3384/ecp212.002.

[30] S. Schlömer, T. Bruckner, L. Fulton, E. Hertwich, A. McKinnon, D. Perczyk, J. Roy, R. Schaeffer, R. Sims, P. Smith, R. Wiser, Annex III: Technology-specific cost and performance parameters, in: Climate Change 2014: Mitigation of Climate Change.

Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2014.

[31] C. Mutel, Brightway: An open source framework for life cycle assessment, Journal of Open Source Software 2 (12) (2017) 236. doi:10.21105/joss.00236.

[32] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index, Computer Physics Communications 181 (2) (2010) 259–270.

[33] M. J. W. Jansen, Analysis of variance designs for model output, Computer Physics Communications 117 (1–2) (1999) 35–43.

[34] T. Ishigami, T. Homma, An importance quantification technique in uncertainty analysis for computer models, in: Proceedings of the First International Symposium on Uncertainty Modeling and Analysis (ISUMA), 1990, pp. 398–403.

[35] L. Breiman, Random forests, Machine Learning 45 (2001) 5–32.

[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, et al., Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[37] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, From local explanations to global understanding with explainable AI for trees, Nature Machine Intelligence 2 (2020) 56–67.

[38] A. B. Owen, Sobol' indices and shapley value, SIAM/ASA Journal on Uncertainty Quantification 2 (1) (2014) 245–251.

## Appendix A. Model parameters

| Parameter | System | Unit | Distribution | Source |
|---|---|---|---|---|
| ef_steel | shared | kg CO2e/kg | T(0.68, 1.91, 2.33) | worldsteel Sustainability Indicators 2023 (2022 data): scrap-EAF 0.68, global average 1.91, ... |
| ef_aluminium | shared | kg CO2e/kg | U(10.04, 14.8) | International Aluminium Institute press release (2023 data): sector average incl. recycling ... |
| ef_copper | shared | kg CO2e/kg | T(3.172, 3.965, 4.758) | International Copper Association, Copper Environmental Profile Global 2023: 3,965 kg CO2e pe... |
| ef_concrete | shared | kg CO2e/m3 | U(187, 255) | EPD listings (OneClickLCA): Heidelberg Materials C30/37 Norway 187 kg CO2e/m3 (A1-A3); Swedi... |
| concrete_density | shared | t/m3 | fixed 2.4 | Heidelberg Materials C30/37 EPD listing: 2400 kg/m3 |
| ef_glass | shared | kg CO2e/kg | T(0.824, 1.03, 1.236) | AGC Glass Europe: European average architectural float glass 10.3 kg CO2e/m2 at 4 mm (A1-A3,... |
| ef_glass_fibre | shared | kg CO2e/kg | T(3.832, 4.79, 5.748) | Taha, Vouros, Kyprianidis (2025), SIMS EUROSIM 2024, Linkoping Electronic Conference Proceed... |
| ef_polymer | shared | kg CO2e/kg | U(2.19, 6.59) | low: virgin PET 2.19 kg CO2e/kg (PlasticsEurope/PETCORE 2017 eco-profile, as cited by NIH); ... |
| grid_cn | shared | kg CO2e/kWh | empirical | Ember yearly electricity data (CC BY 4.0), life-cycle factors incl. upstream methane |
| ef_gas_comb | shared | kg CO2/MJ | empirical | ISPRA Rapporti 413/2025, Table 1.2 (LHV basis) |
| ef_gas_upstream | shared | kg CO2e/MJ | T(0.005, 0.0115, 0.02) | IEA (2026) Assessing emissions from LNG supply and abatement options (CC BY 4.0): all gas su... |
| eta_gas | gas | - | empirical | Eurostat nrg_bal_c (TI_EHG_MAPE_E) and nrg_bal_peh (GEP_MAPE), natural gas, Italy |
| wind_model | wind | - | choice (3 plants) | Vestas public ISO 14040/44 LCAs: V162-6.2, V150-6.0 (2023), V163-4.5 (2025) |
| cf_wind | wind | - | empirical | Eurostat (CC BY 4.0) Italian wind fleet |
| lifetime_wind | wind | yr | T(16, 20, 30) | Vestas reports: baseline 20 yr; sensitivity 16-24 yr; text: lifetime 'could be up to 30 years' |
| uplift_wind | wind | - | empirical | Calibrated on the three Vestas ISO-reviewed LCAs (src/calibrate_wind.py): Vestas result with... |
| elec_mgsi | pv | kWh/kg | LN(med 11, SD95 1.22) | IEA PVPS T12-19:2020, Table 6 (MG-Si, CN) |
| co2_mgsi | pv | kg CO2/kg | LN(med 3.58, SD95 1.22) | IEA PVPS T12-19:2020, Table 6 (fossil CO2 from reductants) |
| mgsi_per_sogsi | pv | kg/kg | LN(med 1.13, SD95 1.23) | IEA PVPS T12-19:2020, Table 7 |
| elec_sogsi | pv | kWh/kg | LN(med 49, SD95 1.1) | IEA PVPS T12-19:2020, Table 7 (solar-grade Si, Siemens, CN) |
| heat_sogsi | pv | MJ/kg | LN(med 28.8, SD95 1.1) | IEA PVPS T12-19:2020, Table 7 (cogeneration heat) |
| elec_ingot | pv | kWh/kg | LN(med 32, SD95 1.22) | IEA PVPS T12-19:2020, Table 9 (CZ mono-Si, CN) |
| gas_ingot | pv | MJ/kg | LN(med 68.2, SD95 1.32) | IEA PVPS T12-19:2020, Table 9 |
| si_per_wafer | pv | kg/m2 | LN(med 0.595, SD95 1.22) | IEA PVPS T12-19:2020, Table 12 (mono-Si wafer, CN) |
| elec_wafer | pv | kWh/m2 | LN(med 4.76, SD95 2.05) | IEA PVPS T12-19:2020, Table 12 |
| wafer_per_cell | pv | m2/m2 | LN(med 1.03, SD95 1.1) | IEA PVPS T12-19:2020, Table 16 |
| elec_cell | pv | kWh/m2 | LN(med 17.7, SD95 1.09) | IEA PVPS T12-19:2020, Table 16 |
| cell_per_panel | pv | m2/m2 | LN(med 0.935, SD95 1.24) | IEA PVPS T12-19:2020, Table 19 (panel, CN) |
| elec_panel | pv | kWh/m2 | LN(med 14, SD95 1.09) | IEA PVPS T12-19:2020, Table 19 |
| glass_panel | pv | kg/m2 | LN(med 8.81, SD95 1.33) | IEA PVPS T12-19:2020, Table 19 (solar glass) |
| al_panel | pv | kg/m2 | LN(med 2.13, SD95 1.24) | IEA PVPS T12-19:2020, Table 19 (aluminium frame) |
| cu_panel | pv | kg/m2 | LN(med 0.103, SD95 1.24) | IEA PVPS T12-19:2020, Table 19 |
| polymer_panel | pv | kg/m2 | LN(med 1.774, SD95 1.24) | IEA PVPS T12-19:2020, Table 19: EVA 0.875 + PVF 0.112 + PET 0.346 + HDPE 0.0238 + GFRP-polya... |

| Parameter | System | Unit | Distribution | Source |
|---|---|---|---|---|
| steel_mount | pv | kg/m2 | LN(med 42.01, SD95 1.89) | IEA PVPS T12-19:2020, Table 36 (open ground, Mont Soleil): reinforcing steel 39.5 + low-allo... |
| concrete_mount | pv | m3/m2 | LN(med 0.0205, SD95 1.89) | IEA PVPS T12-19:2020, Table 36 |
| steel_inverter | pv | kg/kWp | fixed 9.792 | IEA PVPS T12-19:2020, Table 40 (1 MW inverters + transformers): 9792 kg per MW |
| al_inverter | pv | kg/kWp | fixed 0.894 | IEA PVPS T12-19:2020, Table 40: 894 kg per MW |
| cu_inverter | pv | kg/kWp | fixed 2.277 | IEA PVPS T12-19:2020, Table 40: 2277 kg per MW |
| inverter_replacement | pv | - | fixed 1.2 | IEA PVPS Methodology Guidelines T12-18:2020, 3.1.1: utility inverters 30 yr with 10% part re... |
| cu_cable | pv | kg/kWp | fixed 1.3 | IEA PVPS T12-19:2020, Table 50 (570 kWp plant): copper 741 kg |
| polymer_cable | pv | kg/kWp | fixed 1.244 | IEA PVPS T12-19:2020, Table 50: polyethylene 709 kg |
| module_power | pv | kWp/m2 | fixed 0.195 | IEA PVPS T12-19:2020, Table 2: mono-Si 195 Wp/m2 (19.5%) |
| irradiation | pv | kWh/m2/yr | empirical | PVGIS 5.3 (JRC), 527-point grid over Italy, optimal tilt |
| performance_ratio | pv | - | U(0.8, 0.9) | IEA PVPS Methodology Guidelines T12-18:2020, 3.1.3: current PR 0.8-0.9; 0.80 default for gro... |
| degradation | pv | 1/yr | U(0.005, 0.007) | IEA PVPS Methodology Guidelines T12-18:2020, 3.1.4: 0.7 %-points/yr default, 0.5 in sensitiv... |
| lifetime_pv | pv | yr | T(25, 30, 30) | IEA PVPS Methodology Guidelines T12-18:2020, 3.1.1: 30 yr recommended; 25 yr warranty |