

CARBONBENCH

Methodology

About the PDF. The PDF is a print edition of this Methodology page as it stood on September 29, 2026; this page remains the current reference. It replaces the July 14, 2026 v1.1 PDF, which listed five Transparency Score components weighted 30/15/20/15/20. The live system uses **six**, weighted 27/13/18/14/18/10 — a value-chain component was added in v0.2 and the others were rescaled — so anyone holding the earlier PDF should cite the live weights, not the PDF weights. The live parameter table is the file at /config/table_v1.json — its `_version` field is authoritative and the version in force is declared in every record's API response. **If this page and the PDF ever differ, cite this page (with the date you accessed it) and the version shown in the record's API response.**

Core principle

Blockchain does not make data true; it freezes verified data. The accuracy claim is carried by source citation and the verification pipeline; the integrity claim by cryptography (Principle I3). **The protocol does not prove that a record is true — it proves that the record went through a specific verification process and has not changed since.**

Data pipeline

- **Acquisition:** two record classes. (1) **Company reports** — sustainability and annual reports from company websites and exchange filing systems (e.g. KAP in Türkiye, CNINFO in mainland China, HKEXnews in Hong Kong, company investor-relations pages for US and other listed companies — aggregator sites such as responsibilityreports.com are used for discovery only, never as the source URL); every PDF gets a SHA-256 hash and a timestamp. (2) **Official-registry records** — regulator datasets (EPA GHGRP, Canada ECCC GHGRP, EU MRV, UK ETS, Australia NGER, Japan EEGS, Korea GIR); the retrieved dataset row is hashed and sealed as the record.
- **Extraction:** company reports in two stages — (A) PDF → text cache (once per report), (B) text → structured record (deterministic parsers first, a language model only for what they cannot read). Registry records are parsed directly from the regulator's dataset. Every record carries its extraction-source tag.
- **Verification:** Deterministic rules: order of magnitude, scope disclosure, grid EF unit check (0.2–1.0 kgCO₂e/kWh — calibrated on Türkiye's grid, see Limitations), internal consistency, ±40% temporal deviation, assurance class. Flagged records go to 100% human review (risk-weighted

sampling). Registry records skip the magnitude and coverage rules: their value comes from a regulatory filing.

- **Sealing:** Canonical JSON (sorted keys, fixed separators, normalized floats) → SHA-256 → (1) a hash-chained seal ledger + (2) the CarbonbenchRegistry contract on a local single-writer node — cryptographic tamper-evidence, not external notarization; public-chain anchoring is on the roadmap. Append-only: a correction = a new version; no version is ever deleted.

Transparency Score (0–100) — current version: v0.2

COMPONENT	WEIGHT	CRITERION
EF transparency	27	Whether EF values are numerically disclosed in the report (full/partial/none)
GWP recency	13	AR6 full, AR5 partial, AR4/unclear low
Methodology	18	Whether EF sources (DEFRA/IPCC/national) are referenced
Internal consistency	14	Automated check results
Scope 3 disclosure	18	Category count + total disclosure (cross-sourced in v2)
Value chain disclosure	10	Supplier/value chain program and breakdown disclosure level (added in v0.2)

The formula lives in [config/scoring.json](#), not hardcoded in the code; every change is versioned (Methodology Board principle, spec 7.1), the live version is declared in every record's API response, and the breakdown appears component by component (`scores.transparency_breakdown`). Next to every score there is a "why" panel — there is no black box. The full formula of the Accuracy Estimate (σ_T spread, evidence classes, the $\hat{q}/\hat{b}/\lambda_t$ substitution table, cap rules) is in [config/taflo v1.json](#) — and you can check it yourself against a live record in [Reproduce a score yourself](#) below.

Reproduce a score yourself — pen and paper

Every scored record's API response (`/api/record/<record_id>.json`) publishes its intermediate quantities in the accuracy block `records.dogruluk`: `.sigma_T` (σ_T , combined spread), `.delta2` (Δ^2 , reducible uncertainty), `.qhat` ($\hat{q}$, flag rate), `.bhat` ($\hat{b}$, corrected systematic bias), `.skor` (the score), `.bant_alt` / `.bant_ust` (band low / high), `.beyan_toplam` (disclosed total) and `.taflo` (the frozen parameter-table version it was scored with). The constants are in [config/taflo v1.json](#) under the key `sabitler` (constants): $k_1=33.3$, $k_2=2.0 \cdot q_kaba$, γ (gross-error weight γ)= $2.0 \cdot \lambda_t=1.0$. Two formulas close the loop:

Accuracy Estimate $S = \max(0, 100 \cdot e^{(-k_1 \cdot \Delta^2 - k_2 \cdot |\hat{b}|)} - 100 \cdot \gamma \cdot \hat{q})$ (CB-P 9.4.1)
 Plausibility band $E \cdot e^{(\pm 1.96 \cdot \lambda_t \cdot \sigma_T)}$ (CB-P 9.5.1)

Worked example — Ford Otosan 2025 (record `ford_otomotiv_sanayi_a_s_2025_v1`). The record publishes $\Delta^2 = 0.00369$, $\hat{q} = 0.0407$, $\hat{b} = 0$, $\sigma_T = 0.0646$, $E = 364,378$ tCO₂e:

$100 \cdot e^{(-33.3 \times 0.00369)} = 88.44$

$88.44 - 100 \times 2.0 \times 0.0407 = 88.44 - 8.14 = \mathbf{80.30 \rightarrow \text{the published score.}}$

Band: $364,378 \cdot e^{(\mp 1.96 \times 0.0646)} \approx 321,043 - 413,562$ tCO₂e $\rightarrow$ matches the published 321,050 – 413,553 to within $\pm 0.01\%$ (the record publishes σ_T rounded to four digits).

```
curl -s https://carbonbench.org/api/record/ford_otomotiv_sanayi_a_s_2025_v1.json |
jq '.records.dogruluk'
```

What this does and does not prove: given the published intermediates, the final score reproduces to ± 0.1 with a calculator — no access to our code required. Deriving Δ^2 , $\hat{q}$ and σ_T themselves from the raw report takes the evidence-ladder and σ dictionaries in the same config file — keys `kanit_merdiveni_U` (evidence ladder), `ef_sigma_sozlugu` (emission-factor uncertainty dictionary) and `korelasyon` (correlation) — plus the Fenton–Wilkinson combination in CB-P 9.2.1: fully specified, but a spreadsheet job rather than a two-line one. If your recomputation of a record does not match, that is a finding, not a nuisance: report it via [Support](#) and it enters the public error matrix.

Comparison — three layers

- 1. Global intensity percentile:** tCO₂e / million \$ revenue, across the whole universe. The question: "how much emission per \$1 I spend?" It is sector-independent — heavy industry naturally scores low here; this is a deliberate design choice.
- 2. Within-sector percentile:** The same intensity metric, but only within the sector group (a fair comparison). A minimum universe of 5 companies is required — below that, no score is produced.
- 3. External benchmark context — two sources:** (a) [EXIOBASE 3.8.2 \(ExioML\)](#) — **Türkiye 2022** sector intensities, kg CO₂e / € value added, VA-weighted (country-specific: a like-for-like context only for companies in Türkiye); (b) [EPA USEEIO v1.3](#) — cradle-to-gate supply-chain intensity, kg CO₂e / 2022 USD (secondary context). Caution: the EXIOBASE denominator is value added (not revenue) while EPA includes the supply chain — definitional differences from the company's own S1+2 intensity are disclosed explicitly; benchmarks never enter the score. The [Benchmark Explorer](#) shows two EXIOBASE denominators — per € of gross output (country comparison) and per € of value added (trend); they are not interchangeable.

Revenue is held in million USD. Financials in Turkish lira (TRY) taken from KAP are converted with annual average TCMB/ECB rates (`config/fx.json`). Records without revenue data fall back to the absolute S1+2 percentile and are marked with * — no fake score is produced.

Official-registry records

About two thirds of the universe are sealed extracts of regulator data (EPA GHGRP, Canada ECCC GHGRP, EU MRV, UK ETS, Australia NGER, Japan EEGS, Korea GIR), not company reports. Each program covers only part of a company: facilities above a threshold, in one jurisdiction, for the scopes the program defines (for example UK ETS: Scope 1 within the scheme boundary; NGER: Scope 1 + Scope 2; Korea GIR: one combined direct + indirect total). The registry value is therefore a **lower bound** of the company's total, and each profile states the program's boundary. Registry records carry no emission-factor disclosure; their evidence is the regulatory filing itself.

Threat model — what we defend against, what we do not

- **Defended against:** silently altering a sealed record (including by the operator), hiding the revision history, selective data deletion.
- **Not defended against:** the source report itself being wrong (physical reality is the assurance auditor's job), extraction errors outside the sample (never zeroed out; measured and bounded instead).
- **Principle I3 — physical reality:** the input is the company's disclosure. We assess its internal consistency, its methodological correctness and the quality that audit sampling can test; we do not measure or guarantee what was actually emitted. A consistently wrong disclosure produces a consistently wrong result, and we say so openly (CB-P §2.3-a).

Deep documentation

- [Anatomy of a Carbon Number \(CB-M Manual\)](#) — 9 chapters, 133 numbered failure modes, 53 standard objections; the literature roots of every formula.
- **CB-P Protocol v1.0-draft** — the shall/should-worded standard text: [docs/protocol-v1.md](#).
- **Reading error matrix (2026-07-14)** — the public measurement of our own margin of error: the disagreement rate on records resolved in the independent second-reading round, and the classification from human review (our reading error / intra-company inconsistency / false alarm). Machine-to-machine reconciliation is the LOWER bound of the margin of error; the upper bound will be provided by the human-labeled gold set. Detail: [error matrix \(docs/error-matrix.md\)](#).
- **Physical intensity (2026-07-14)** — per-tonne emissions in heavy industry (steel/cement/glass/aluminum): absolute and revenue-based percentiles are scale-biased. Production tonnage is captured deterministically from the report and positioned against

IEA/GCCA/IAI benchmark bands. It stands out in the benchmark band but never enters the score (physical-intensity bands: [config/fiziksel_yogunluk.json](#)).

- **v1.1 signal layers (2026-07-12)** — new evidence sources linked to $\hat{q}/\hat{b}$ without changing the score formula: reading-confirmation round + human-approved error classification (operator error $\neq$ organizational inconsistency; in cases of severe inconsistency, an Accuracy cap of 20) · Scope 2 dual reporting audit · EPA GHGRP official anchor (one-directional, ownership-share-weighted, full-name-matched) · electricity activity cross-check · SBTi official list + sectoral pathway disclosure verification · Transparency v0.2 value chain component. Detail: CB-P Appendix C.
- **[Badge Protocol \(CB-B v1.0\)](#)** — the criteria formulas, periodicity and independence rules for periodic performance badges (full text: [docs/badge-protocol-cb-b.md](#)).
- **[Panel Deck](#)** — a simplified record of how the methodology was negotiated.
- Every rule in the verification log is now tagged with a CB-M x.y.z taxonomy code — an exact match to the auditor's Excel sheet.

Limitations (v0.2 honesty statement)

- Part of the universe of companies in Türkiye was extracted with the first-generation pipeline (tag `legacy_v0`); it enters sample verification with priority.
- Sector groups are a rough mapping (an official GICS mapping is on the roadmap); most official-registry records are not yet classified and sit in "Other".
- The grid-factor unit check (0.2–1.0 kgCO₂e/kWh) is calibrated on Türkiye's grid; very low-carbon grids (below 0.2) and market-based zero values can be flagged for human review.
- The current chain is a local EVM (same contract, real transactions); the move to Polygon Amoy/mainnet is a single deploy.
- Write access rests with a single operator; multi-sig is on the roadmap.