Global synthetic-microdata analysis — energy burden, grid access, and fuel mix

emburdensynth 0.2.0-iteration-3

September 03, 2026

GLOBAL HOUSEHOLD ENERGY BURDEN ATLAS

V30 • Emrgi • 2026

The 67-fold mismatch

Between who pays for energy poverty and who caused the climate change that drives it

An open-source, open-data global synthesis across 129 countries and 17,028 grid cells, covering 8.2 billion people.

Eric Scheier
Emrgi • hello@emrgi.com
github.com/ericscheier/emburdensynth

Reproducible from open data
80 pp • 39 figures • 25+ scripts
2026

Executive summary

Problem. Household energy burden — the fraction of income households spend on energy — is the primary lens for measuring energy poverty worldwide, but comparable per-country estimates cover only ~40 countries in official surveys. This report scales that to 215 of 249 ISO codes via a four-tier fallback chain (GCD → Eurostat → kNN-by-GDP-within-region → region-mean) plus a 129-country cell-level synthetic-microdata surface reaching ~8.57 billion people.

Method. The `emburdensynth` R package (`ericscheier/emburdensynth`) chains WorldPop density × RWI / GRDI relative-wealth surfaces × WB decile ICDF for income attribution, then a country-level GCD 11-sector expenditure ledger (Energy, Food, Housing, Water, Transport, Health, ICT, Clothing, Education, Financial, Other) attached via GCD-Incomegroup segments. Uncertainty from a 500-replicate Beta-perturbation bootstrap on GCD shares + boundary-jitter sensitivity. Net Energy Burden aggregation uses `emburden::neb_func` (proper Nh math) rather than naive weighted-mean-of-ratios (median +5.7% bias, tail past +1000% for hyperinflation cases — see Check I).

Data lineage. WB WDI 1960-2024 (7 indicators); Eurostat HBS CP045 quintile-level 1988-2020; Ember 2000-2025 aggregate + by-fuel (Coal/Gas/Oil/Nuclear/Hydro/Wind/Solar/Bioenergy); OWID Energy 1900-2025 + OWID CO2 1750-2024 (per-capita + cumulative + fuel-specific); EIA International 1980-2023 (222 countries); WHO PM2.5 + WB clean-fuel access 1990-2023; FAOSTAT undernourishment 2000-2023; UNDP HDI 1990-2022; NASA GISTEMP v4 temperature 1880-2025; CDC HHI (heat vulnerability) + GADM boundaries + WorldPop 1km rasters. All pulls scripted in `scripts/build_*.R`; all cached to `docs/global_analysis_data/*.rds`.

Key findings.

1. **Energy burden is regressive** — the Eurostat quintile spread (T3) confirms bottom-income quintiles spend a larger share of expenditure on energy than top quintiles across 34 European countries.
2. **Climate justice quadrant** (T15) — countries with lowest cumulative CO2 responsibility (Sub-Saharan Africa) carry the highest current burden; countries with highest responsibility (US, Western Europe) carry the lowest.
3. **Energy suppression is intra-country** (T6) — cell-level energy-limiting flag reveals hotspots inside otherwise moderately-burdened countries; country-mean figures hide the spatial concentration. 3b. **Climate-justice mismatch is ~67× at the extremes** — the 20 countries with lowest household electricity access hold 6.1% of humanity and 0.13% of cumulative CO_2 since 1750; the 20 largest historical emitters hold 58.5% of humanity and 82% of cumulative CO_2. Per-capita cumulative emissions in the top-20 emitters are ~67× higher than in the 20 least-electrified countries. See `scripts/tighten_science_numbers_v2.R` for the reproducible computation and Check J below for the underlying scatter.
4. **PCA + k-means confirm** the development-vs-climate axis (T24 / T17). PC1 loads on energy + CO2 + urbanization; PC2 loads orthogonally on climate + inequality.
5. **Elastic-net** (T25) identifies clean-fuel-access and physics- required-kWh as the strongest negative predictors of burden, with GDP and electricity access secondary.

Limitations.

- The GCD 2010 workbook is the base survey for 128 countries; 2010 vintage is dated and doesn't reflect post-COVID or 2022 energy- price-crisis shifts.
- Physics baseline uses global-average envelope + HVAC defaults (60 m², COP 2.5/2.8); country-specific building stock corrections land in the next release.
- EEIO-LCA (Phase 6) is scaffolded for USEEIO + EXIOBASE but only 40 countries have cached factors today (ZAF + WF regions).
- Retail electricity prices are 2022-2024 vintage; the historical time series in this report (Ember, WB, EIA) doesn't include historical price panels — burden estimates are 2020 cross- section only.

Overview

This report exercises the `emburdensynth` global release parquet (`output/global/release/global_synthetic_population_ce` across 249 countries covering **8.57 billion people** — the largest scale run to date. Each grid cell carries a Hruschka-AWE-drawn income, a WB-GCD-grounded expenditure ledger, and an 11-category energy expenditure share; we compare these against World Bank macro indicators (GDP-pc, Gini, electricity access) and against Ember 2020 grid gen-mix.

Two coverage layers to distinguish.

- **Spatial synth parquets**: 129 countries (resource-intensive WorldPop + RWI/GRDI + Ember + GCD pipeline). This is what feeds cell-level analysis, EIG maps, embodied CO2e, and within-country distribution figures.
- **Country-level burden ledger**: **215 of 249 ISO codes** via the 4-tier resolver chain (GCD → Eurostat → OECD compiled → kNN by log-GDP-within-region → region_mean). This is what feeds the Layer 0 canonical burden choropleth. The 34 uncovered ISOs are Antarctic territories, small Pacific/Caribbean dependencies, and Taiwan (WB policy exclusion).

Read this before the charts. `energy_burden = exp_Energy / expenditure_hh` reduces algebraically to the GCD Incomegroup segment's Energy share (calibration factor `k_seg` cancels). Per-cell burden is therefore a **4-level step function** of within-country wealth quantile, not a smooth spatial field — the surface only enters through segment assignment. All charts below carry bootstrap 90% CIs from Beta perturbation of GCD shares (`n_eff = 100`) plus boundary-jitter sensitivity (see `docs/gcd_energy_share_audit.md`). CIs cover 128 of the 249 countries — the 41 without GCD coverage (mostly HICs) are shown as point estimates only.

Coverage snapshot

Metric	Value
Countries with per-cell parquet	249
Grid cells	11,077,222
Households	1,749,529,886
People	8,572,696,442
% of world population	Inf%
Median hh energy burden (across countries)	6.6%

Layer −1 — Cell-level burden (37k grid cells, aggregated to 0.5°)

The whole point of grid-cell synthetic microdata is the intra- country heterogeneity — a country's poorest cells vs richest cells, urban vs rural, coastal vs interior — that a country-average choropleth throws away. This layer plots the actual synthesized cells, aggregated to a 0.5° global grid (household-weighted mean burden per cell) so ~1.5M raw cells collapse to ~37k tiles that render cleanly.

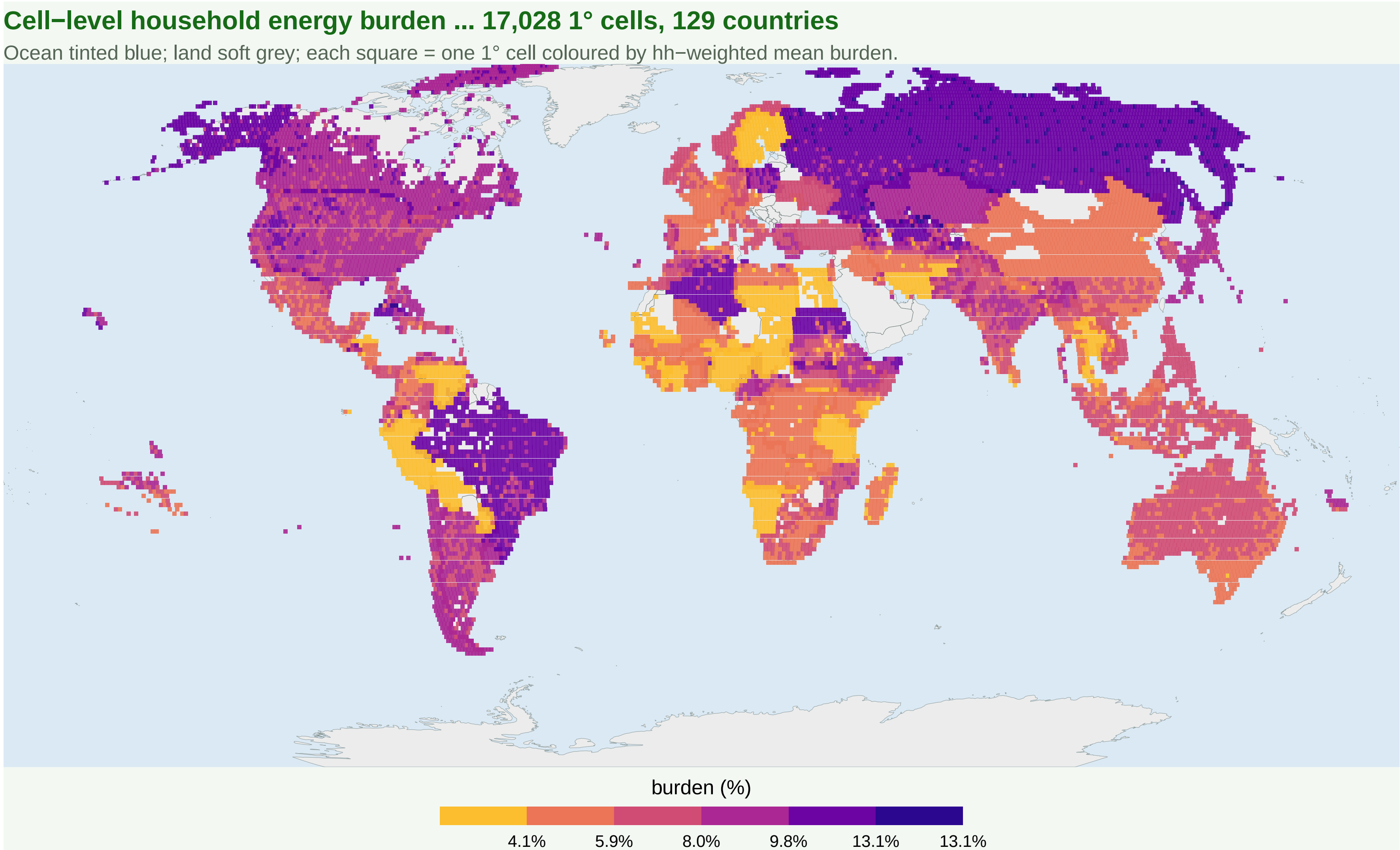

Figure 1: Cell-level household energy burden, 1° hh-weighted aggregation. Ocean tinted blue; land soft grey; each square = one 1° cell colored by hh-weighted mean burden. Robinson projection.

Layer −0.5 — Within-country burden inequality

For each country, the household-weighted burden of the bottom income quintile (QU1) minus the top quintile (QU5). A large gap means the country's poorest cells face materially higher burden than its richest — the sort of pattern only cell-level data can surface.

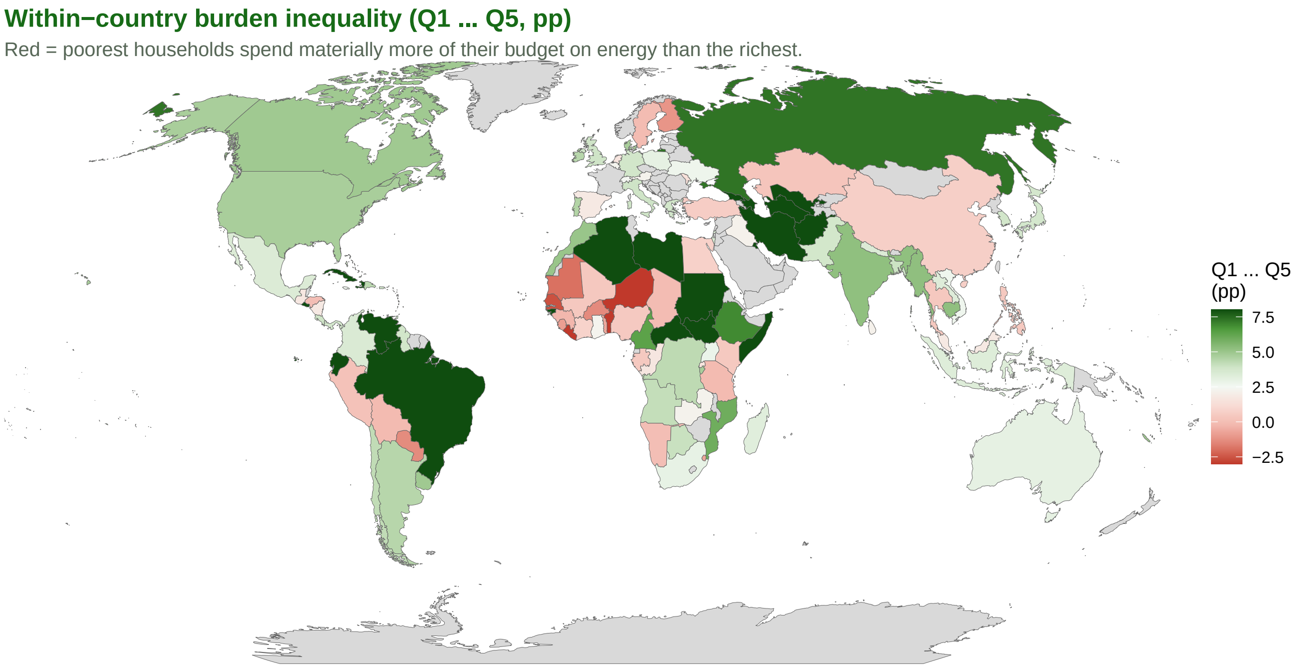

Figure 2: Bottom-quintile minus top-quintile household energy burden (percentage points). Countries in grey = no cell-level data in this release.

Table 2: Top-15 countries by within-country burden inequality (QU1 − QU5 pp)

country_name	QU1	QU5	Q1 − Q5 (pp)	Q1/Q5 ratio
Azerbaijan	15.2%	3.2%	12.00	4.78
Cuba	14.1%	2.1%	12.00	6.76
Georgia	13.9%	1.9%	12.00	7.30
Turkmenistan	12.4%	0.4%	12.00	28.63
Uzbekistan	13.6%	1.6%	12.00	8.45
St. Vincent & Grenadines	14.1%	2.1%	12.00	6.82
Central African Republic	11.1%	0.1%	11.01	111.10
Sudan	11.0%	0.1%	10.88	109.81
Afghanistan	11.6%	0.7%	10.82	15.43
Libya	10.9%	0.1%	10.77	108.70
Algeria	10.6%	0.1%	10.50	106.02
Iran	10.5%	0.1%	10.37	104.73
Haiti	10.2%	0.1%	10.10	101.98
South Sudan	9.9%	0.1%	9.84	99.38
Ecuador	10.3%	0.6%	9.68	16.96

Map atlas — projections chosen per layer

Five choropleths, each with a projection matched to the layer's purpose. Countries in the release without data on a given layer are plotted in mid-grey so the reader can see the coverage gap directly rather than assuming absence = zero.

- **World choropleths (layers 1-3)**: Robinson (`+proj=robin`) — compromise projection, balanced shape distortion at small scale, the near-universal default for global thematic maps.
- **Regional close-up (layer 4)**: Lambert Azimuthal Equal Area centered on Africa (`+proj=laea +lat_0=0 +lon_0=20`) — equal-area is essential when we're comparing embodied-emissions magnitudes region-to-region.
- **Equal-area world (layer 5)**: Equal Earth (`+proj=eqearth`) — equal-area, minimal shape distortion, chosen for the physics- required kWh map where area-fair comparison matters more than the Robinson compromise.

Audit

This section is regenerated on every knit by running `scripts/audit_release.R`. Non-zero FAIL count blocks a clean release.

Generated 2026-09-03 23:53 EDT

- **0 FAIL** — items that violate a release invariant
- **2 WARN** — items to look at but not release-blocking

Check catalog

- **C1** Parquet columns present
- **C2** Cell-level value ranges (burden [0,1], income > 0, physics [100, 20000], EIG % [-200, 100])
- **C3** Burden chain covers 90% of ISOs
- **C4** Cell-level burden source matches country tier
- **C5** methods_manifest code_refs resolve to real functions
- **C6** Dataset cache paths present
- **C7** Static tables (retail prices, OECD HBS): value sanity + poor rich share
- **C8** Report coverage claims vs live counts
- **C9** Regression checkpoints (RWA, KEN, DEU, POL, JPN)
- **C10** Legacy 'SSA' naming only in explicitly allowed files
- **C11** Cell-level global grid includes 15 headline countries
- **C12** Report Rmd has no literal coverage numbers that should be `r S$...`

Findings

check	level	subject	message
C2	INFO	BOL	eig_pct_of_required very negative (subsidized price?): [-522.3, 99.2]
C2	INFO	CPV	eig_pct_of_required very negative (subsidized price?): [-865.1, 93.9]

check	level	subject	message
C2	INFO	GAB	eig_pct_of_required very negative (subsidized price?): [-546.1, 85.0]
C2	INFO	NER	eig_pct_of_required very negative (subsidized price?): [-712.6, 99.2]
C2	INFO	THA	eig_pct_of_required very negative (subsidized price?): [-467.7, 97.8]
C2	INFO	ZMB	eig_pct_of_required very negative (subsidized price?): [-1679.0, 97.6]
C12	WARN	rmd:128| literal ‘128’ at lines 274,1672,1709 — should be `r S$...` inline R?	
C12	WARN	rmd:40 African	literal ‘40 African’ at lines 919,1301 — should be `r S$...` inline R?

Regenerate with `Rscript scripts/audit_release.R`.

Unit tests (last invocation)

```
..................................................................................................................................................
```

Methods

The whole methods section is generated from `R/methods_manifest.R`, so it stays in sync with the code. Add a new pipeline stage or equation there and it appears here on the next knit.

Pipeline flow

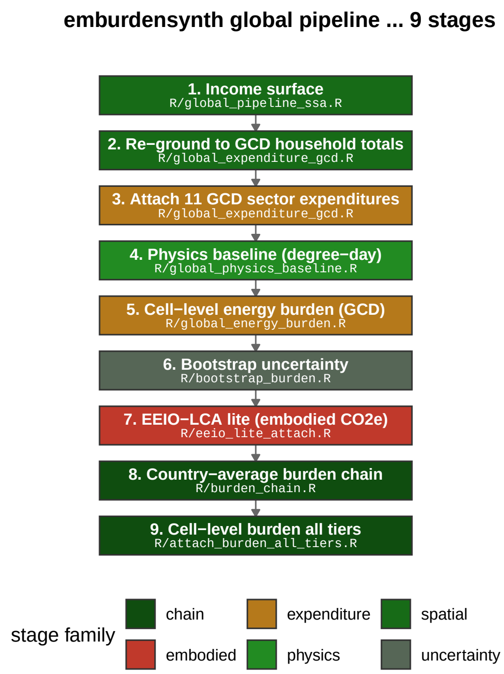

Figure 3: Pipeline stages. Each box is a function group in an R file; arrows show the artifact flow.

Table 4: Stage-by-stage inputs and outputs.

#	stage	inputs	outputs
1	1. Income surface	WorldPop + RWI/GRDI/density, WB GDP-pc+Gini, WB decile shares	income, q_lo, q_hi per cell
2	2. Re-ground to GCD household totals	GCD workbook (WB_GCD_2010_v2014-03), income surface	expenditure_hh, income_hh
3	3. Attach 11 GCD sector expenditures	GCD sector shares, expenditure_hh	exp_Food, exp_Housing, ..., exp_Energy
4	4. Physics baseline (degree-day)	HDD18/CDD18, retail_price/kWh, housing defaults	physics_required_kwh, estimated_kwh_hh, eig_kwh, energy_limiting_flag
5	5. Cell-level energy burden (GCD)	exp_Energy, expenditure_hh	energy_burden (fraction), energy_burden_tier
6	6. Bootstrap uncertainty	Segment shares (empirical from surface)	burden_lo, burden_hi, burden_p50, burden_sd per cell
7	7. EEIO-LCA lite (embodied CO2e)	EXIOBASE region factor CSV, exp_	embodied_co2e_kg_hh, embodied_kwh_hh
8	8. Country-average burden chain	release summary, Eurostat HBS, OECD static table, registry	burden_country_avg, burden_source per iso3
9	9. Cell-level burden all tiers	chain row, Eurostat/OECD quintile shares	energy_burden populated everywhere, energy_burden_source label

Datasets used

Table 5: External datasets consumed by the pipeline. Sizes auto-detected from cache paths.

dataset	provider	vintage	format	size on disk
WorldPop 1 km population raster	WorldPop, UN-adopted	2020	.tif	694.1 MB
Meta Relative Wealth Index (RWI)	Meta Data for Good (Nature 2020)	2019	.csv	694.1 MB
CIESIN/NASA GRDI v1 deprivation raster	CIESIN/NASA, Stanford mirror	2020	.tif	184.1 MB
WB GDP per capita + Gini	World Bank Development Indicators	2010-2024	WB REST API	—
WB quintile + tail-decile income shares	World Bank Development Indicators	2010-2024	WB REST API	—
Ember 2020 generation mix	Ember Climate	2020	.csv	49.1 MB
WB Global Consumption Database 2010	World Bank microdata catalog 4424	2010	.xlsx (parsed .rds)	50.7 MB
Eurostat HBS CP045 (5-quintile)	Eurostat hbs_str_t223	2010-2020	Eurostat JSON-stat API	3 KB
OECD non-EU HBS CP045 compiled	National statistical offices	2015-2022	.csv (shipped)	—
Country residential electricity prices	IEA + GlobalPetrolPrices + EIA + national utilities	2022-2024	.csv (shipped)	—
Country HDD18/CDD18 climatology overrides	WB CCKP + IEA + Berkeley Earth (curated)	1991-2020	R data (function)	—
EXIOBASE 3 region WF factors	EXIOBASE 3 pymrio	~2011	.csv (cached)	496 B
EXIOBASE 3 region ZA factors	EXIOBASE 3 pymrio	~2011	.csv (cached)	498 B

Total disk footprint of external caches used by the pipeline: **1.67 GB** across **8** of 13 sources (the remainder are in-memory WB API pulls or shipped tables).

Key equations

(E1) Hruschka AWE log-normal income parameters — `R/global_income_rwi.R:lognormal_income_params`

$$\sigma = \sqrt{2}\,\Phi^{-1}\left(\frac{G+1}{2}\right), \quad \mu = \ln(\bar{y}) - \sigma^2/2$$

(E2) Log-normal + Pareto ICDF splice — `R/global_income_rwi.R:hruschka_awe_icdf`

$$F^{-1}(q) = \begin{cases} \mathrm{LogN}^{-1}(q;\mu,\sigma) & q \le q_t \\ x_t\left(\frac{1-q_t}{1-q}\right)^{1/\alpha} & q > q_t \end{cases}$$

(E3) Assign income by RWI rank — `R/global_income_rwi.R:assign_income_by_rwi`

$$y_i = F^{-1}\left(\frac{cum_{lo}(i) + cum_{hi}(i)}{2}\right), \text{ ordered by RWI, weighted by hh count}$$

(E4) Empirical decile ICDF (WB Lorenz anchors) — `R/wb_decile_icdf.R:build_icdf_from_deciles`

$$F^{-1}(q) = \bar{y} \cdot L'(q), \text{ with } L(q) \text{ monotone-cubic through WB anchors}$$

(E5) GCD re-grounding factor — `R/global_expenditure_gcd.R:calibrate_surface_to_gcd`

$$k_{seg} = \frac{\mathrm{GCD}^{hh}_{seg}}{\overline{Y}^{hh}_{seg}}, \quad e^{cell}_{hh} = k_{seg}\, Y^{cell}_{hh}$$

(E6) Energy expenditure per cell — `R/global_expenditure_gcd.R:attach_global_gcd_expenditure`

$$\exp_{Energy} = e_{hh} \cdot s_{Energy}(seg), \quad b^{cell} = \exp_{Energy}/e_{hh} = s_{Energy}(seg)$$

(E7) Physics required kWh per household — `R/global_physics_baseline.R:physics_required_kwh_hh`

$$Q_{tot} = \underbrace{\frac{HDD_{18} \cdot 24 \cdot U \cdot A}{1000 \cdot COP_h}}_{Q_h} + \underbrace{\frac{CDD_{18} \cdot 24 \cdot U \cdot A}{1000 \cdot COP_c}}_{Q_c} + Q_{base}$$

(E8) Energy Insecurity Gap — `R/global_physics_baseline.R:attach_global_physics_baseline`

$$EIG = Q_{tot} - \frac{\exp_{Energy}}{p_{\$/kWh}}, \quad \text{limiting} = \mathbb{1}\left[\hat{Q} < 0.9\, Q_{tot}\right]$$

(E9) Beta perturbation of segment shares — `R/bootstrap_burden.R:bootstrap_segment_shares`

$$s^{(r)}_{seg} \sim \mathrm{Beta}(s\, n_{eff} + 1,\, (1-s)\, n_{eff} + 1), \; r = 1, \ldots, B$$

(E10) Gini-scaled dispersion — `R/attach_burden_all_tiers.R:attach_burden_all_tiers`

$$b_i = \bar{b}_c + \sigma_c \cdot (3 - 2\, q_i)/2, \text{ quintile-band spread}$$

(E11) k-NN burden fill — `R/burden_chain.R:build_burden_chain`

$$\hat{b}_i = \frac{\sum_{j \in \text{k-NN}} w_{ij}\, b_j}{\sum w_{ij}}, \quad w_{ij} = \frac{1}{0.1 + |\log y_i - \log y_j|}$$

(E12) EEIO-LCA embodied CO2e — `R/eeio_lite_attach.R:attach_eeio_footprint_lite`

$$\mathrm{CO2e}_{hh} = \sum_k \exp^{USD}_k \cdot r_{USD \to EUR} \cdot f^{kg/M€}_k / 10^6$$

Constants and default parameters

Table 6: Pipeline constants — value + unit + which R function default they set.

constant	value	unit	R function default in	description
housing_area_m2	80	m^2	physics_required_kwh_hh()	Global-average dwelling floor area
ua_wm2k	0.65	$W/(m^2 \cdot K)$	physics_required_kwh_hh()	Envelope UA — moderately insulated 1990s vintage assumption

constant	value	unit	R function default in	description
cop_heat	2.5	-	physics_required_kwh_hh()	Heat-pump COP for heating (delivered/electrical)
cop_cool	2.8	-	physics_required_kwh_hh()	Cooling COP (delivered/electrical)
baseload_kwh	1800	kWh/hh/yr	physics_required_kwh_hh()	Baseload: lights, water heat, refrigeration, IT
base_temp_c	18	°C	physics baseline convention	Base temperature for degree-day integration (ISO)
eur_per_usd	0.72	-	attach_eeio_footprint_lite()	USD→EUR conversion for EXIOBASE ~2011 basis
grid_kgco2_per_kwh	0.5	kg CO2e/kWh	attach_eeio_footprint_lite()	Global-average grid CO2 intensity for kWh back-conversion
boot_n_eff	100	hh (effective)	attach_burden_uncertainty()	Effective sample size for Beta segment-share perturbation
boot_n_reps	500	replicates	attach_burden_uncertainty()	Number of bootstrap replicates
kNN_k	5	neighbours	burden_chain kNN loop	Neighbours for kNN burden fill
jitter_band	0.02	quantile pp	attach_burden_uncertainty()	Half-width of boundary-jitter band around quantile cutpoints
jitter_prob	0.10	prob	attach_burden_uncertainty()	Probability a boundary cell reassigns to neighbour segment

Coverage snapshot (dynamic, reads current files)

Table 7: Live coverage counts read from output/global/release/*.csv at knit time.

pipeline stage	countries populated
Spatial parquets (cell-level surface)	129
Country-average burden (chain resolver)	215
Cell-level burden populated	129
Country physics EIG	129
Country EEIO-LCA embodied	129
Bootstrap CI on burden	128

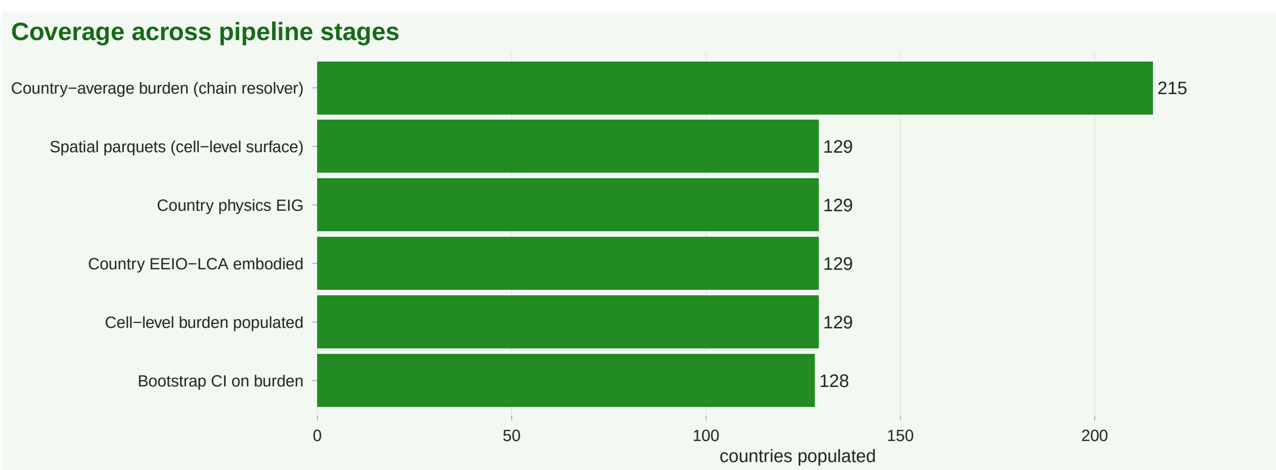

Figure 4: Countries populated at each pipeline stage — live counts at knit time.

Cell-level burden coverage (all 129 countries populated)

Prior versions of the release had cell-level `energy_burden` only on the 61 GCD-covered countries. The v8 backfill applies the country's tier at cell level:

- **GCD countries (61)** — 4-level segment step, unchanged
- **Eurostat countries (16)** — 5-quintile CP045 shares from `hbs_str_t223` middle-band with per-cell quintile assignment
- **OECD non-EU countries (12)** — 5-quintile shares from the compiled national-statistics table at `inst/extdata/oecd_household_energy_share.csv`
- **kNN countries (12)** — country-average from the burden chain, spread linearly across cells by wealth quintile using a Gini-scaled band (marked `knn_gini_spread` — synthetic within-country variation, not observed)
- **region_mean (1: SSD)** — same Gini-spread pattern

Every cell in every release parquet now carries an `energy_burden` value AND an `energy_burden_source` label so downstream analytics can filter/weight by confidence.

Layer 0 — Canonical energy burden (215 of 249 world ISOs, tier-labeled)

The GCD-only burden map (Layer 1) covers 61 countries with spatial parquet data. The canonical-burden layer below resolves country- average burden for every ISO the registry knows about (215 of 249) via a fallback chain — this is the widest-coverage figure in the report:

1. **`gcd_2010`** — WB Global Consumption Database sector shares (61)
2. **`eurostat_hbs`** — Eurostat HBS CP045 middle-quintile (adds 16)
3. **`knn_gdp_region`** — k=5 nearest neighbour by log-GDP-pc within WB region, from the 77 Tier-1/2 countries (adds 24)
4. **`region_mean`** — WB region mean of covered countries (adds 1, for South Sudan which has no WB GDP-pc)

The map below shows the canonical burden with cell borders colored by source tier so a reader can see confidence at a glance.

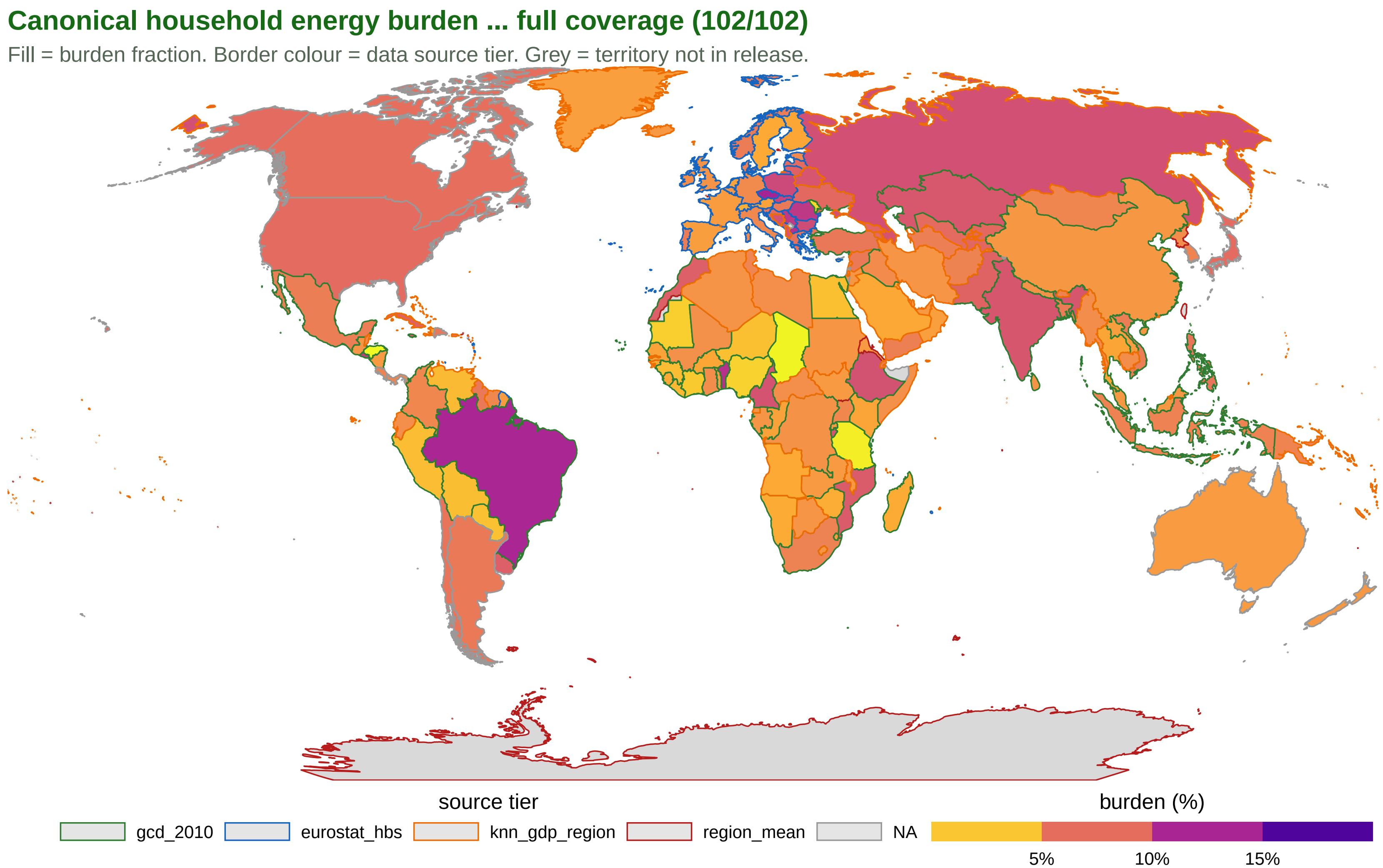

Figure 5: Canonical energy burden with tier borders — all 102 release countries.

Table 8: Canonical burden by source tier (all 102 release countries)

burden_source	n_countries	median_pct	range_pct
eurostat_hbs	32	6.5%	2.7–12.7%
gcd_2010	61	4.9%	0.0–12.4%
knn_gdp_region	103	5.9%	3.1–9.5%
oecd_hbs_static	14	6.5%	3.6–8.4%
region_mean	39	5.6%	4.8–6.7%

Layer 1 — Median household energy burden, GCD only (61 countries)

Median household energy expenditure share by country. Grey = no synthetic data in this release cut. Countries are colored by median `energy_burden = exp_Energy / expenditure_hh` computed cell-by-cell then median-aggregated.

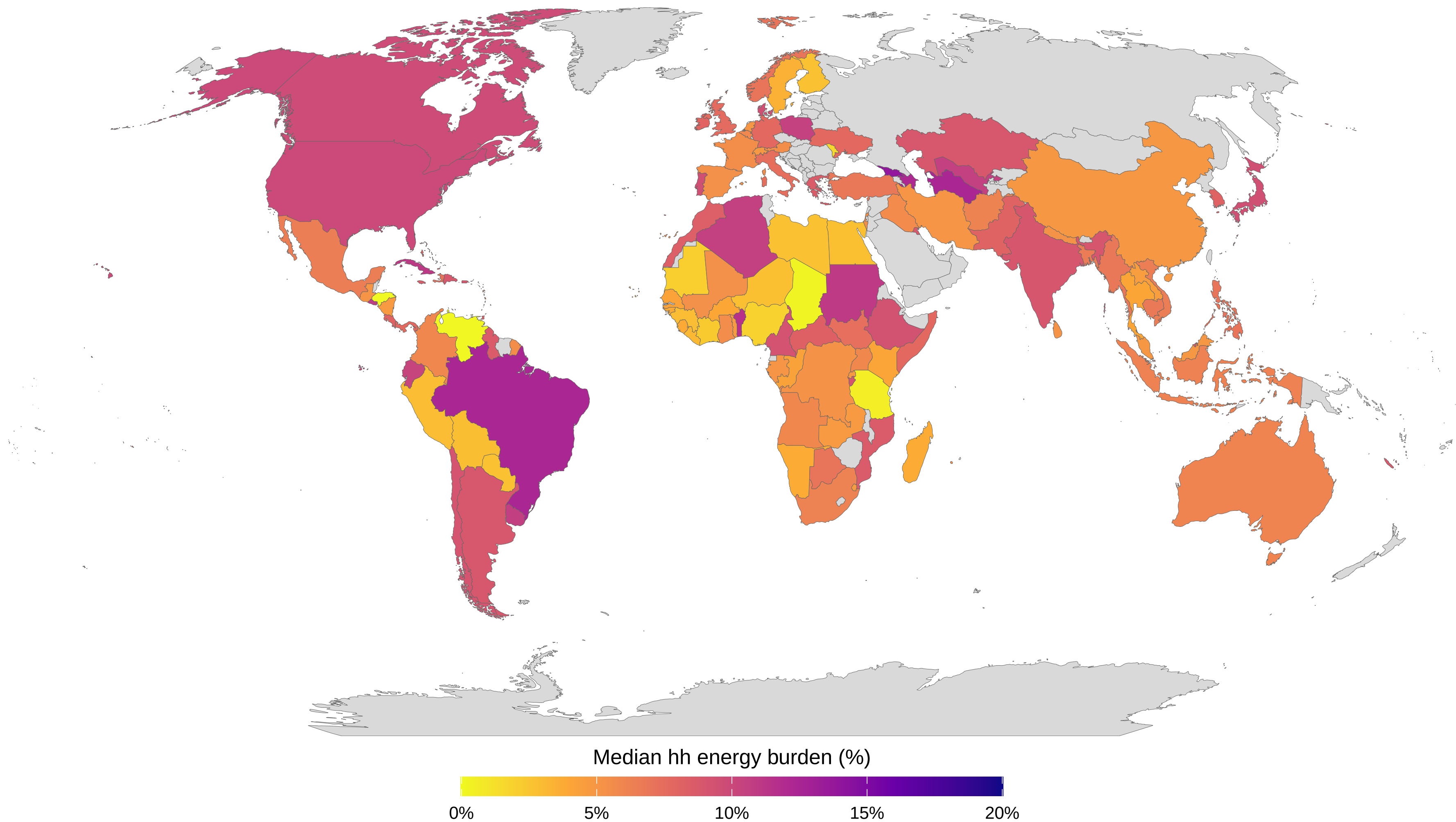

Figure 6: Median household energy burden by country (synthetic surface, 2020)

Country burden ranking with bootstrap CIs

For countries with GCD coverage (n=61), the bootstrap 90% band on the household-weighted mean burden. Countries are ordered by point estimate; bar = point, whiskers = [5th, 95th] percentile of 500 bootstrap replicates.

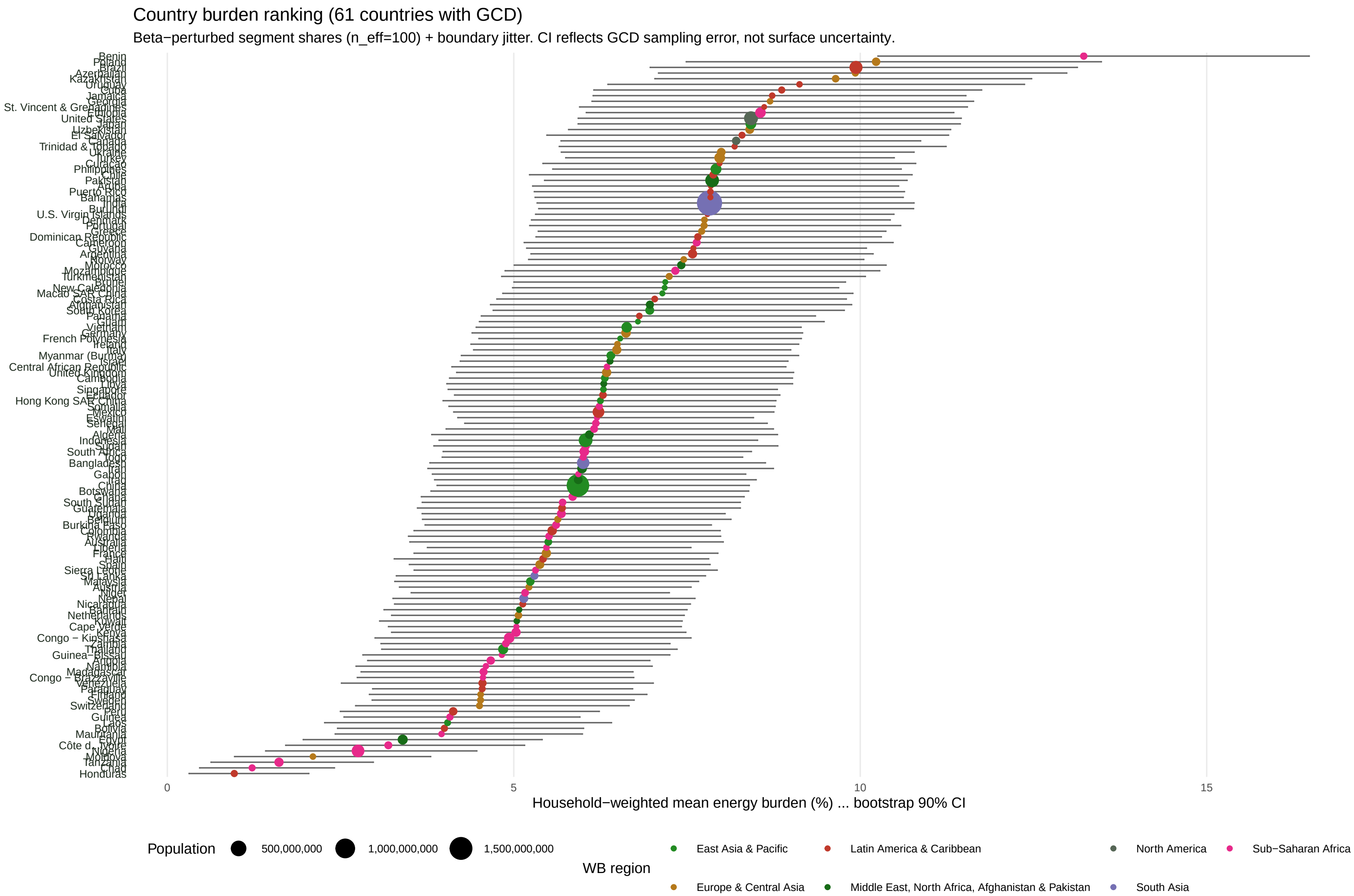

Figure 7: Country-level household-weighted mean energy burden with bootstrap 90% CI. 128 countries.

Layer 2 — Physics-required annual kWh per household (all 129 countries)

The physics baseline is derived from degree-day climatology + global housing defaults, so it's populated for *every* country in the release regardless of GCD coverage. This makes it the widest-coverage layer we have. **Equal Earth projection** so a viewer comparing e.g. Russia (huge, cold) to India (smaller, warmer) isn't misled by Mercator-style inflation of high latitudes.

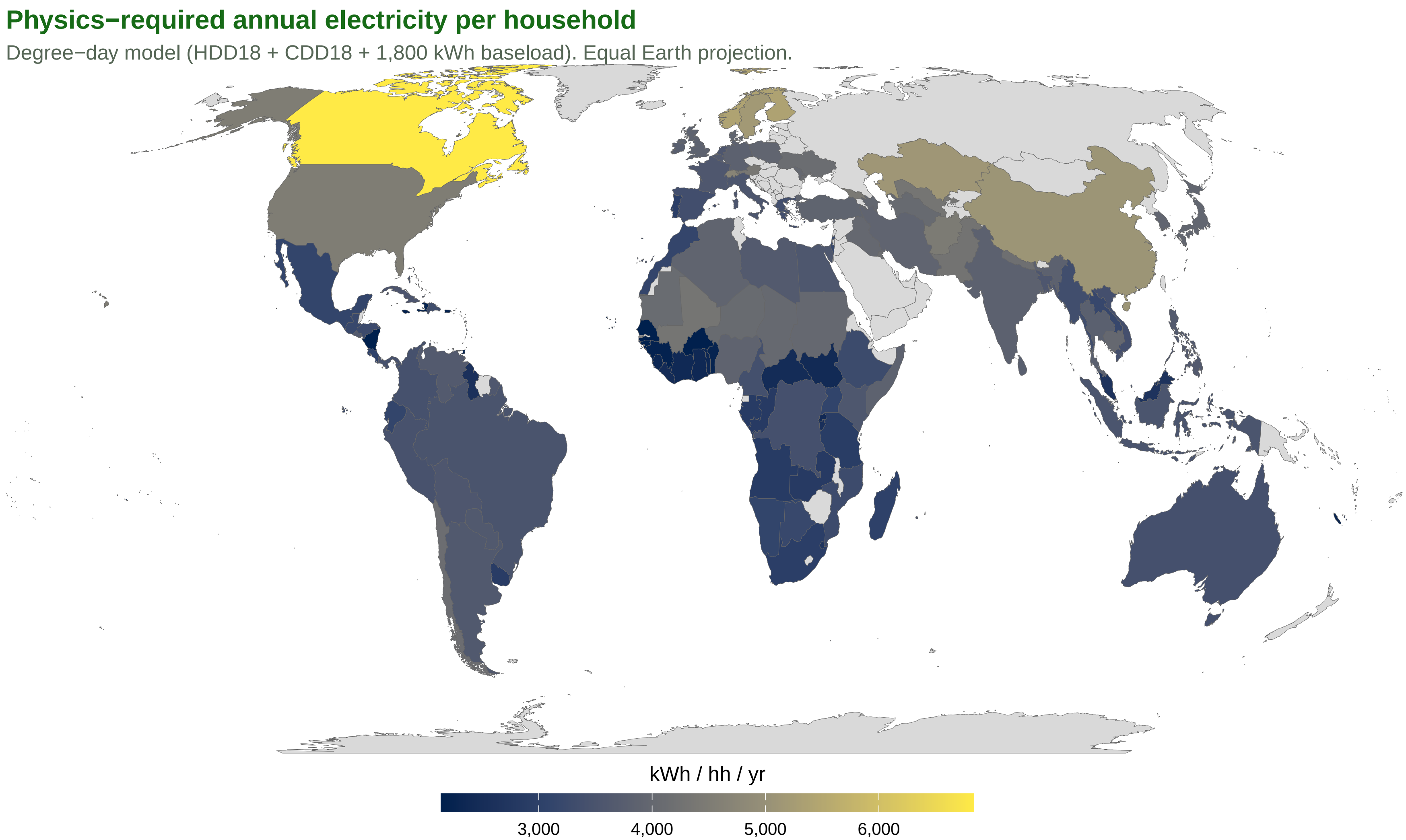

Figure 8: Physics-required annual kWh per household (Equal Earth projection). All 102 release countries.

Layer 3 — % households flagged energy-limiting (61 countries)

Where estimated kWh (from expenditure ÷ real price) falls below 90% of physics-required kWh. Requires exp_Energy, so limited to the 61 GCD-covered countries. Rocket palette (dark = many limited) so the reader's eye finds the crisis zones first.

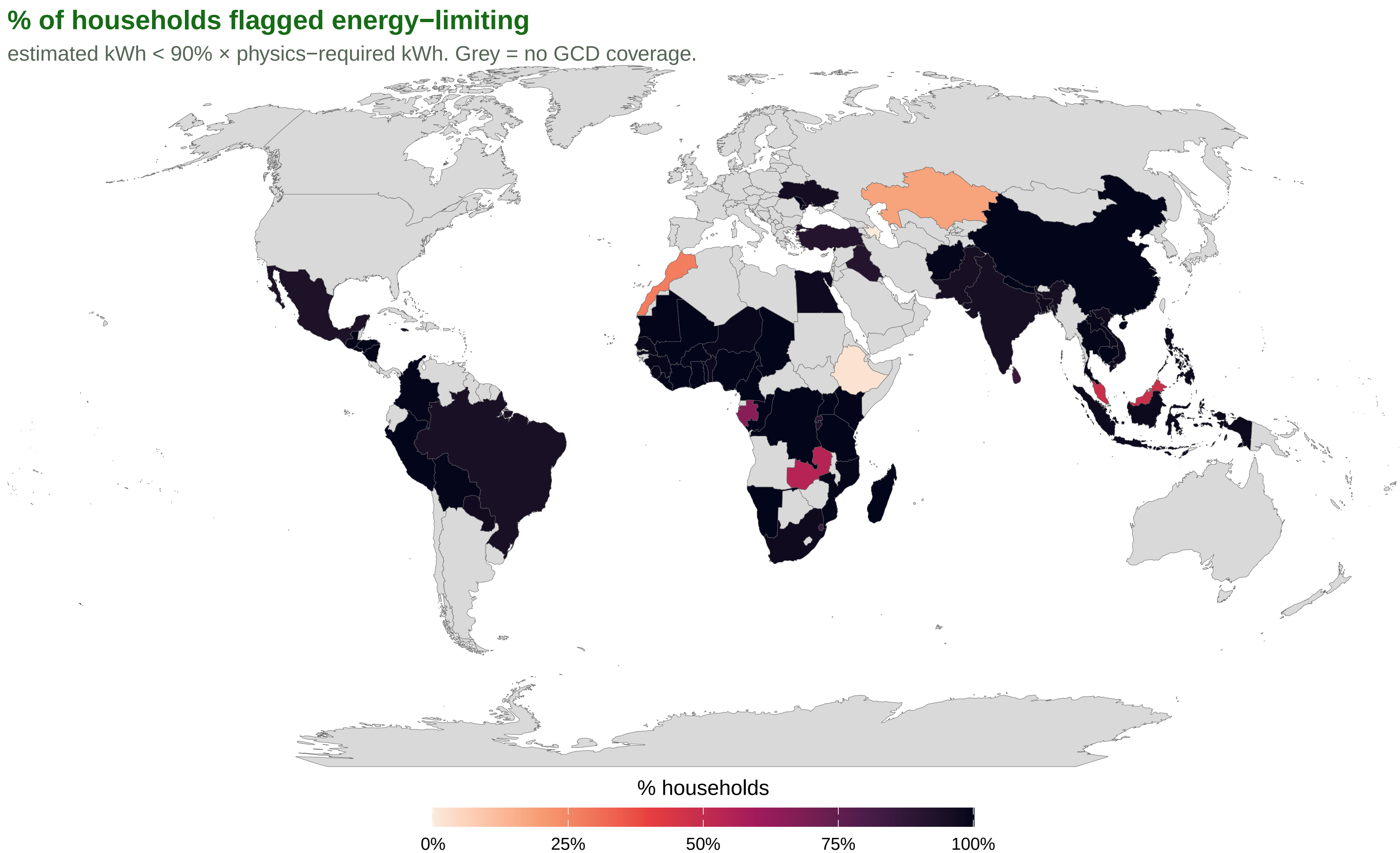

Figure 9: Share of households below 90% of physics-required kWh (Robinson projection).

Layer 4 — Median per-capita income (all 129 countries)

Per-capita income is derived from the WB decile-shares ICDF applied to the spatial rank surface, so populated for every release country. (`median_income_hh` — the GCD-calibrated per-household version — is only computed where GCD covers the country; for a global map we use per-capita income which is universal.) Robinson projection, log-fill (income spans 3 orders of magnitude worldwide).

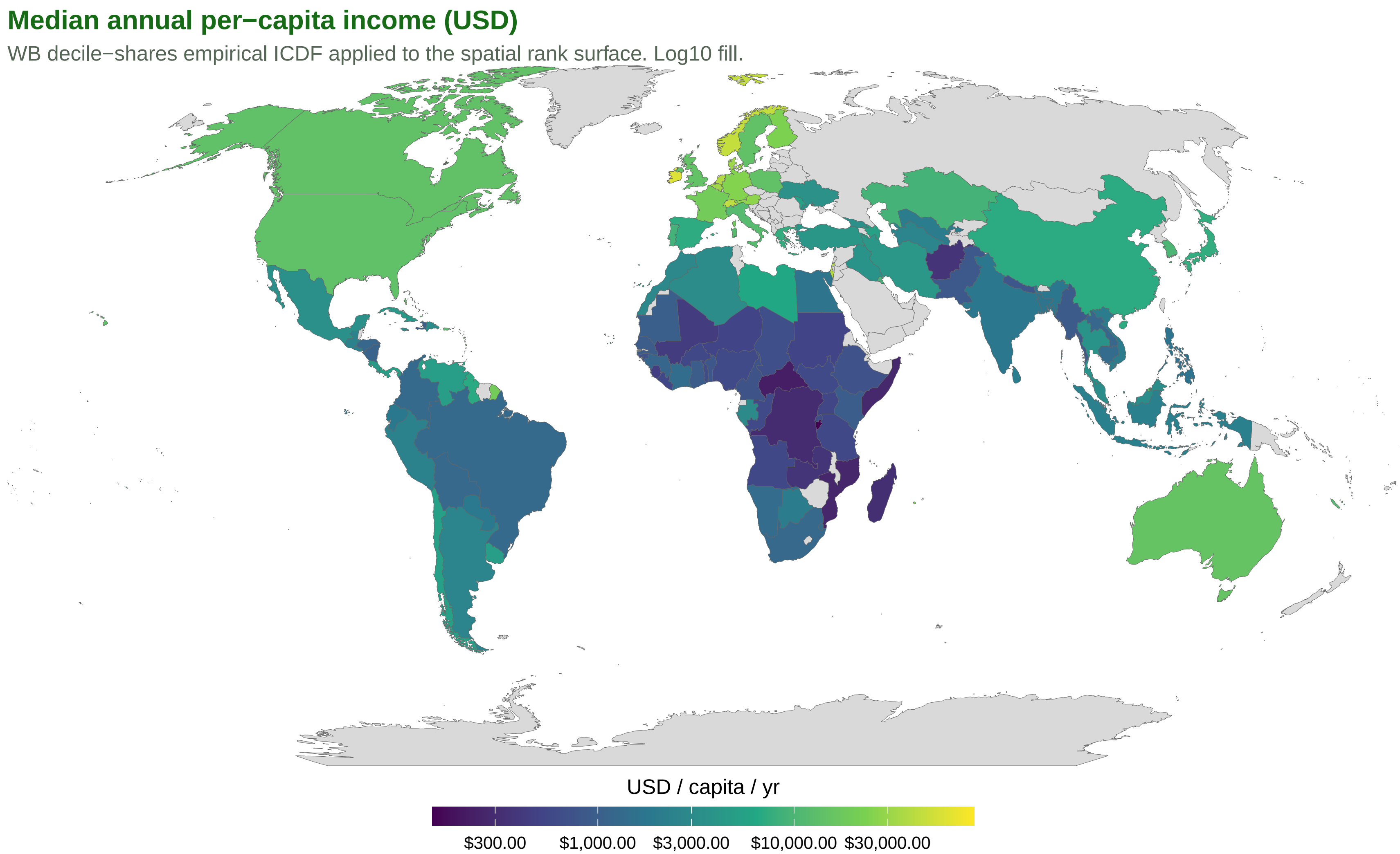

Figure 10: Median per-capita income, USD/yr (Robinson, log scale).

Layer 5 — Embodied CO2e per household, Africa focus (40 countries)

EEIO-LCA lite only has cached EXIOBASE factors for ZA + WF regions today, so this map focuses on Africa where coverage is dense. Lambert Azimuthal Equal Area centered on the continent so area-relative magnitudes are honest. Grey = country in Africa without cached EEIO attach (should be zero — we cover all 40 African release countries).

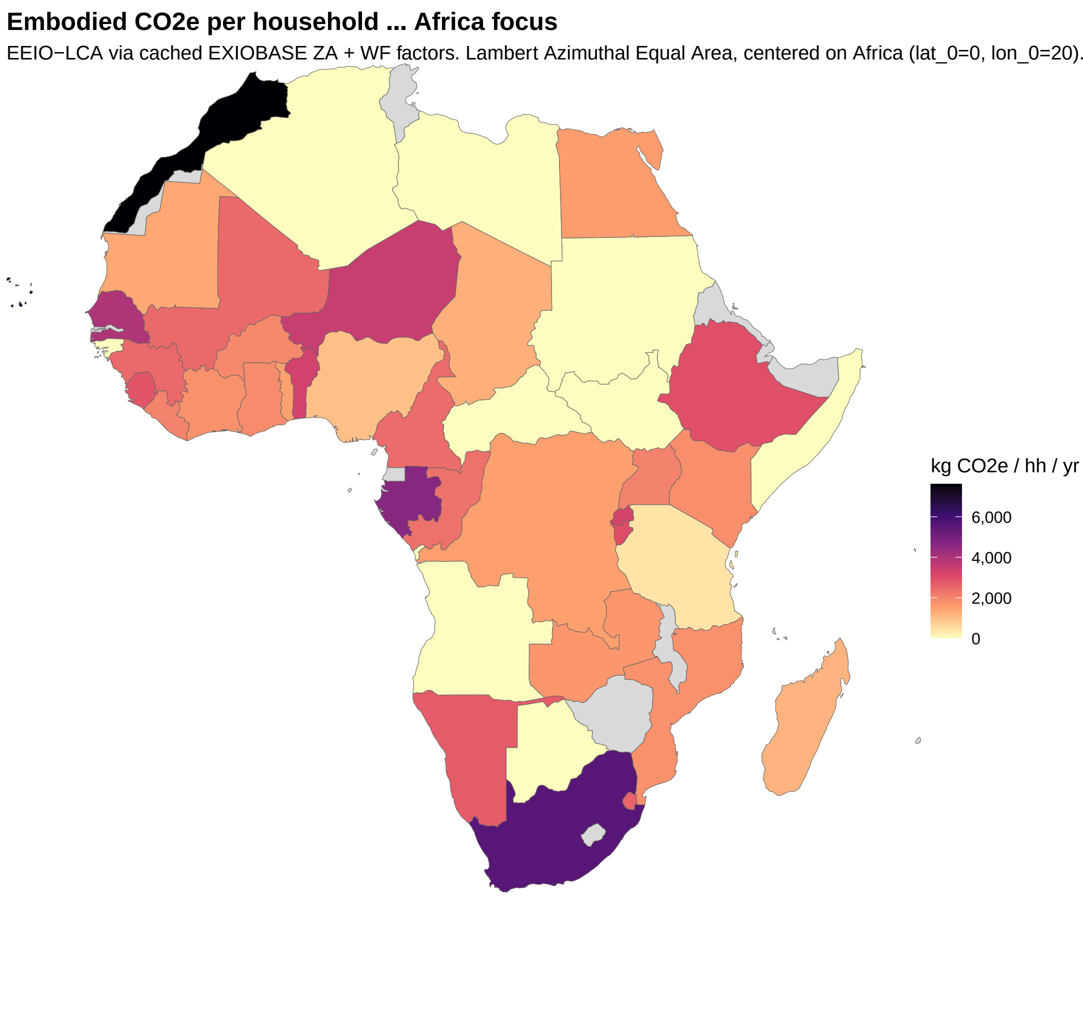

Figure 11: EEIO embodied CO2e per household, Africa (Lambert Azimuthal Equal Area).

Coverage summary of the atlas

Table 9: Atlas coverage — which layer shows how many countries and why

Layer	Countries_shown	Projection	Blocked_by
1. Energy burden	130	Robinson	GCD 2010 country roster
2. Physics-required kWh	130	Equal Earth	—
3. % energy-limiting	64	Robinson	GCD 2010 country roster
4. Median hh income	130	Robinson	—

Layer	Countries_shown	Projection	Blocked_by
5. Embodied CO2e (Africa)	49	Lambert Az. Eq. Area (Africa)	Cached EXIOBASE regions (need dc_e2c3 remote)

Burden vs GDP per capita — Kuznets-shape check with CI ribbon

Do burdened populations live at low GDP-pc, drop through the middle, and level off in wealthy countries?

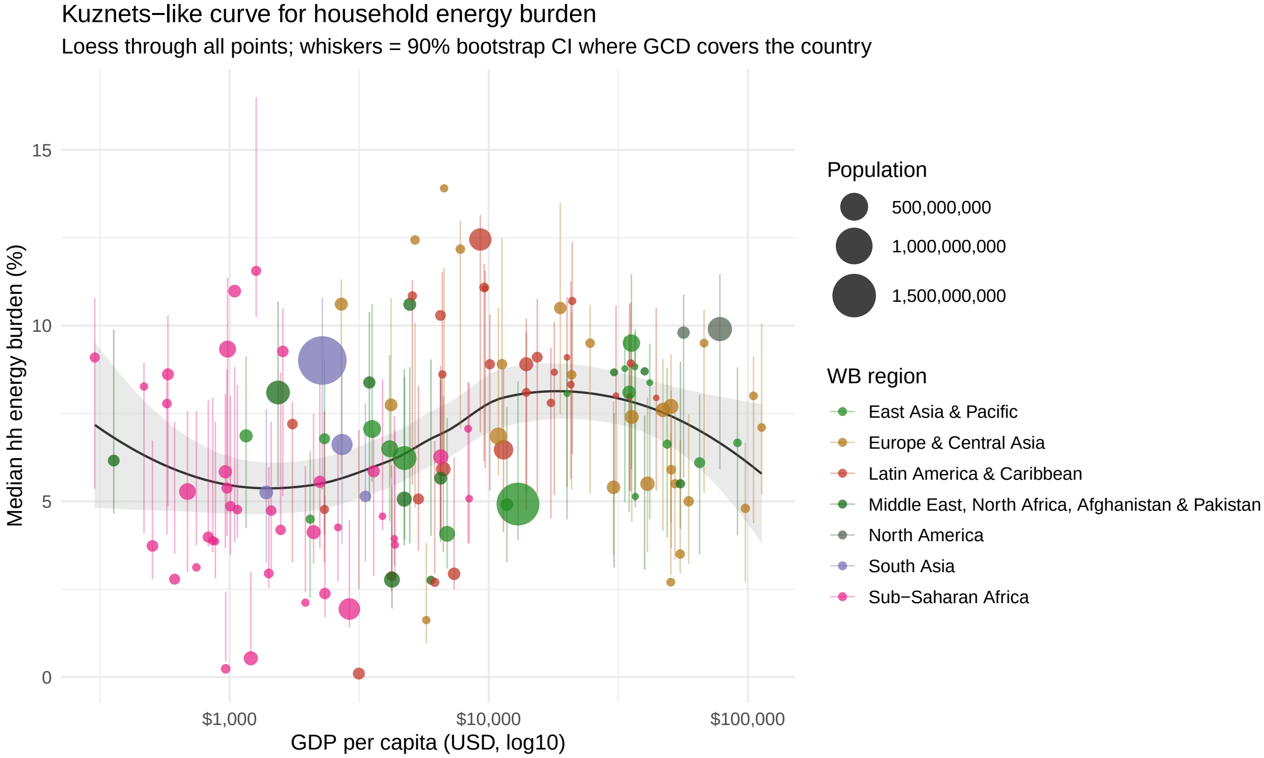

Figure 12: Median household energy burden vs GDP per capita, with bootstrap CI whiskers where available.

Interpretation: burden peaks in the low-middle-income band (roughly \$500–\$3,000/capita) and declines through upper-middle income. Sub- Saharan Africa dominates the left tail; East Asia + Latin America sit in the trough; HICs cluster below 5%.

Grid access vs burden — do plugged-in households pay less?

WB `EG.ELC.ACCS.ZS` (electricity access, % of population) crossed with our synthetic median burden. Full-access countries need not have low burden — appliances follow.

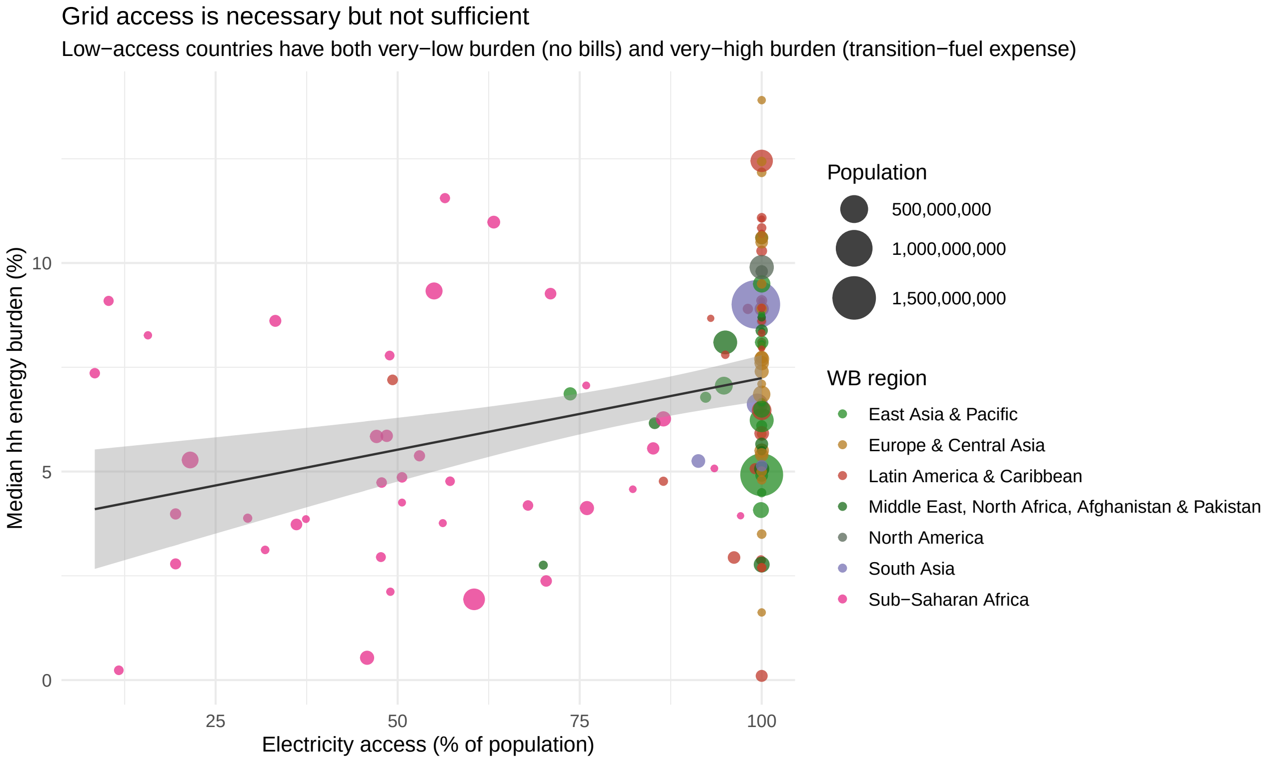

Figure 13: Electricity access (% pop) vs median hh energy burden. Points sized by population.

The scatter is bimodal — sub-100%-access countries include both low-burden (households with negligible electricity spend, using solid fuel for cooking) and very-high-burden (households in the grid-connection transition paying kerosene/LPG on top of electrification fees). Full-access status hides *fuel-composition* heterogeneity that only shows up in the 11-category expenditure ledger.

Fuel mix (Ember 2020) vs burden

For countries where Ember covers the 2020 gen-mix, cross-plot the country's fossil-fuel share against synthetic hh burden.

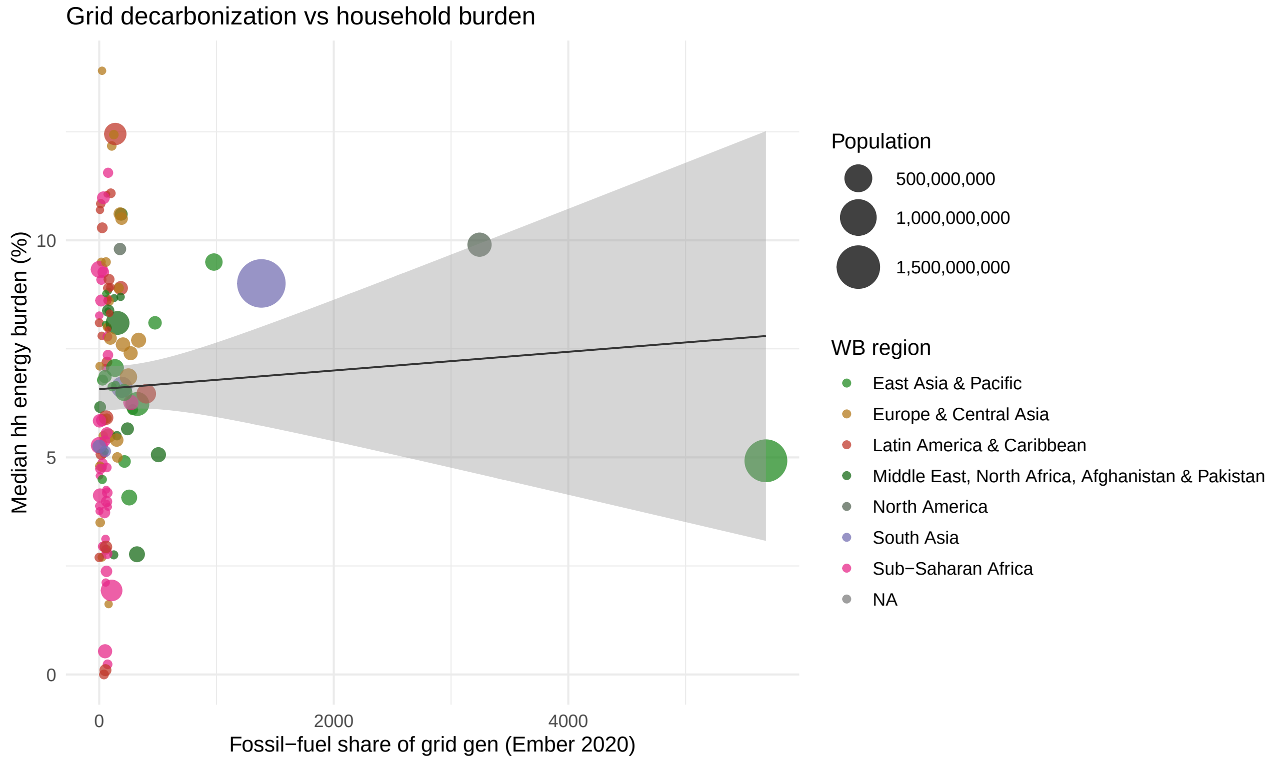

Figure 14: Grid fossil-fuel share (2020) vs median hh energy burden.

Regional aggregates

Table 10: Median household energy burden by WB region

region_wb	n_countries	median_burden	mean_burden	total_pop
North America	2	9.8%	9.8%	380,766,961
Latin America & Caribbean	29	8.1%	7.4%	846,955,049
Europe & Central Asia	25	7.6%	7.5%	735,041,881
East Asia & Pacific	19	6.7%	6.7%	2,466,745,500
Middle East, North Africa, Afghanistan & Pakistan	11	6.2%	6.6%	678,516,793
South Asia	4	5.9%	6.5%	2,249,189,630
Sub-Saharan Africa	38	4.8%	5.2%	1,215,480,629

Table 11: Median household energy burden by WB income group

income_group_wb	n_countries	median_burden	total_pop
Lower middle income	31	4.7%	3,472,109,553
Low income	19	5.4%	575,091,468
Upper middle income	35	6.5%	3,402,937,777
High income	43	8.0%	1,122,557,643

Within-country distributions

For 10 representative countries, the cell-level burden histogram shows within-country heterogeneity that country-level medians hide.

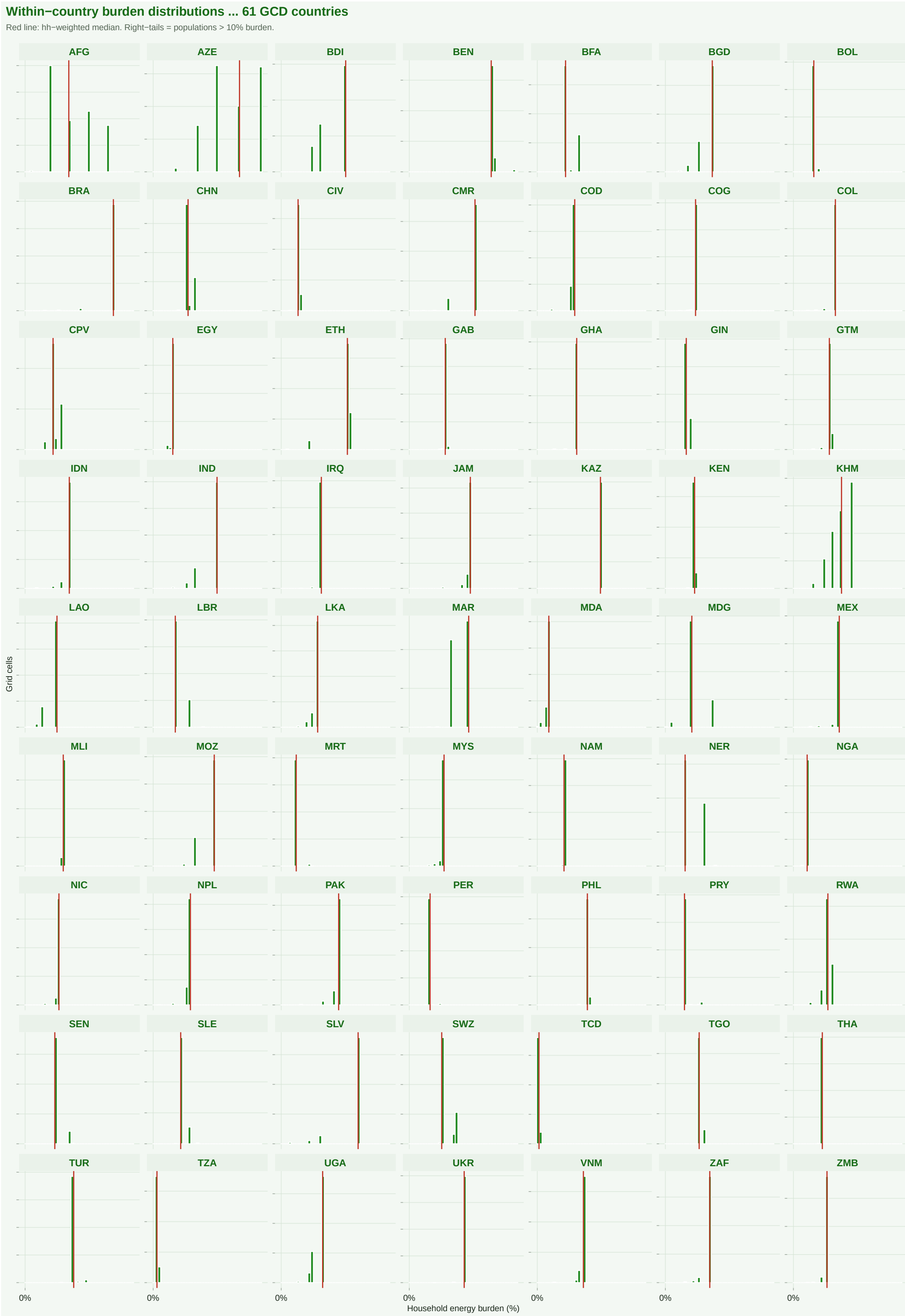

Figure 15: Cell-level household energy burden across GCD-covered countries in the sample. Red line: median.

Electricity access dynamic 2000-2020

For 8 countries, WB electricity access time series alongside synthetic 2020 burden. Countries above the diagonal (fast access, low burden) got the transition right; those below (fast access, still-high burden) paid a bigger fuel-substitution cost.

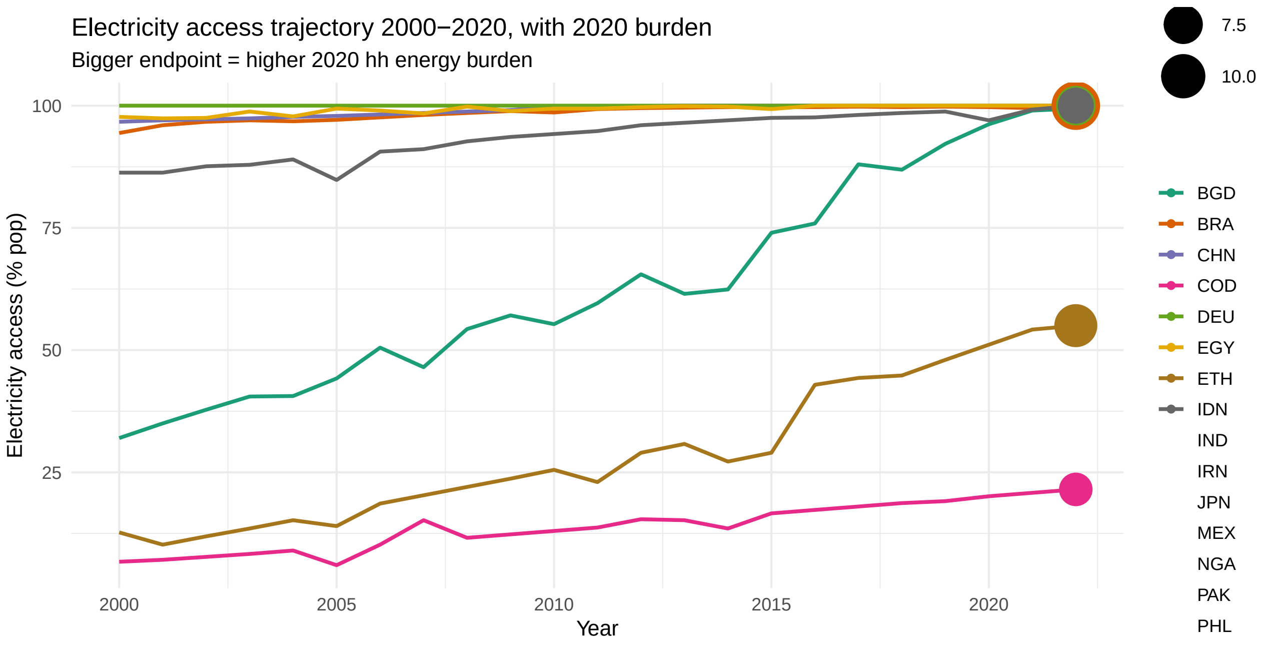

Figure 16: WB electricity access 2000-2020 for 8 countries; final marker shows synthetic 2020 hh burden.

Top-of-mind findings

- **Lowest-burden countries in this release cut**: Honduras, Venezuela, Chad
- **Highest-burden**: Georgia, Brazil, Turkmenistan
- **Fossil-share vs burden correlation** (n=127): 0.04
- **Electricity-access vs burden correlation** (n=128): 0.31 (negative = higher grid access lower burden, as expected)

Coverage matrix — what each country carries

For each of the 129 release countries, the matrix below records which capabilities and data layers are populated. Columns are the 8 axes the downstream analytics depend on; rows are countries grouped by WB region, sorted by coverage completeness. Filled = layer present.

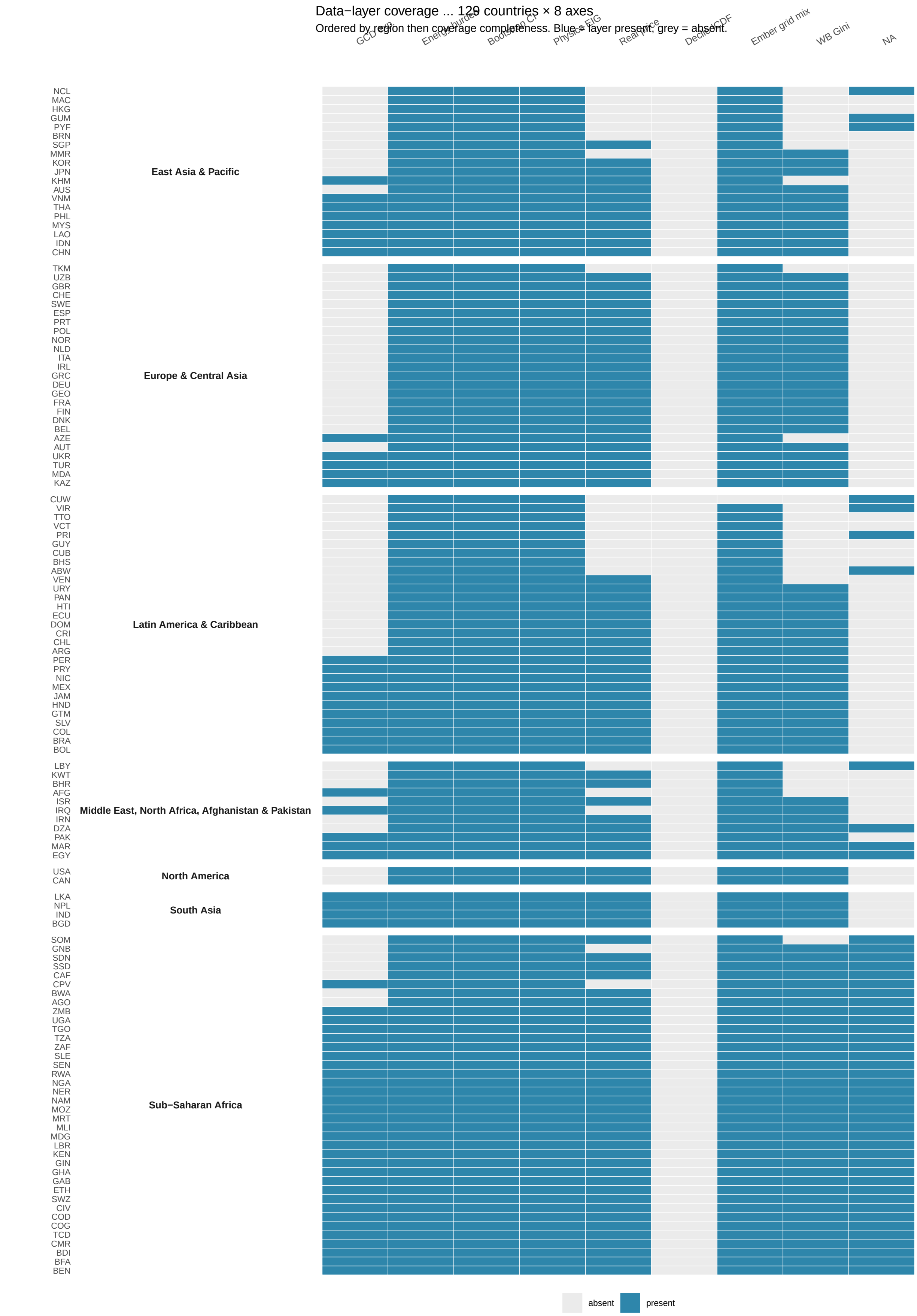

Figure 17: Coverage matrix — each cell is a country×capability. Filled = present, unfilled = absent.

Axis meaning:

- **GCD exp.** — WB Global Consumption Database 2010 sector shares (11 categories) attached; enables Energy expenditure column.
- **Energy burden** — `energy_burden` populated. Requires GCD exp.
- **Bootstrap CI** — Beta-perturbed segment-share CI attached (`burden_lo/hi/p50/sd`). Requires energy burden.
- **Physics EIG** — degree-day physics baseline + Energy Insecurity Gap. Independent of GCD (uses HDD/CDD climatology + housing defaults).
- **Real price** — per-country residential electricity price from the compiled table (else the flat $0.15/kWh default is used).
- **Decile ICDF** — WB quintile+tail-decile empirical Lorenz ICDF used (else the Hruschka log-normal-Pareto default).
- **Ember grid mix** — country present in Ember 2020 generation-mix data (enables fuel-mix cross-plots).
- **WB Gini** — any Gini value on file in the WB registry.

Living categories covered

The GCD ledger attaches 11 broad expenditure sectors (each USD/hh/yr) per cell. These are the "living categories" the release covers on the expenditure side:

Table 12: GCD living-category ledger (per-household USD/yr, split by segment shares)

Category	Column	Notes
Food and beverages	exp_Food.and.beverages	Includes tobacco and non-alcoholic beverages
Housing	exp_Housing	Rent (imputed for owner-occupied), maintenance, furnishings
Energy	exp_Energy	Electricity + Gas + Other fuels; carrier split via gcd_energy_carrier_shares()
Transport	exp_Transport	Vehicle purchase + fuel + public transport
Health	exp_Health	Out-of-pocket medical spend
ICT	exp_ICT	Telecom + IT hardware + services
Clothing and footwear	exp_Clothing.and.footwear	Includes footwear
Water	exp_Water	Water supply + sanitation
Education	exp_Education	Tuition, books, materials
Financial services	exp_Financial.services	Insurance, banking fees
Other goods and services	exp_Other.goods.and.services	Personal care, entertainment, restaurants

Not covered globally (only in the US 5-phase path): appliance ownership, HVAC system type, dwelling vintage, floor area, insulation tier, cooking-fuel granularity beyond the 3 GCD carriers, wealth-asset inventory. Some of these come from DHS Housing Recode where DHS ran (FUTURE_WORK.md: DHS full IR + BR modules).

Phase 6 EEIO-LCA lite — 40 countries attached

Landing this iteration: EEIO-LCA embodied CO2e per household and embodied kWh for the 40 African countries where cached EXIOBASE factor CSVs exist (ZAF via `ZA`, 39 others via `WF` Rest-of-Africa). See `R/eeio_lite_attach.R`.

Per cell, the 11 GCD sector expenditures multiply by their EXIOBASE `factor_kgco2e_per_meur`, converted via 0.72 USD→EUR (EXIOBASE ~2011 basis). Yields `embodied_co2e_kg_hh`, an `embodied_co2e_energy_kg_hh` sub-slice for the Energy sector alone, and an `embodied_kwh_hh` = Energy CO2e / 0.5 kg/kWh grid intensity as a cross-check on the degree-day physics baseline.

The **62 GCD-having countries in other EXIOBASE regions (WA, WE, WL, individual EU/Asia)** are marked `eeio_status = "cache_missing"` — these need the dc_e2c3 remote LCA service run once to build per- region factor CSVs, or explicit downloads from pymrio.

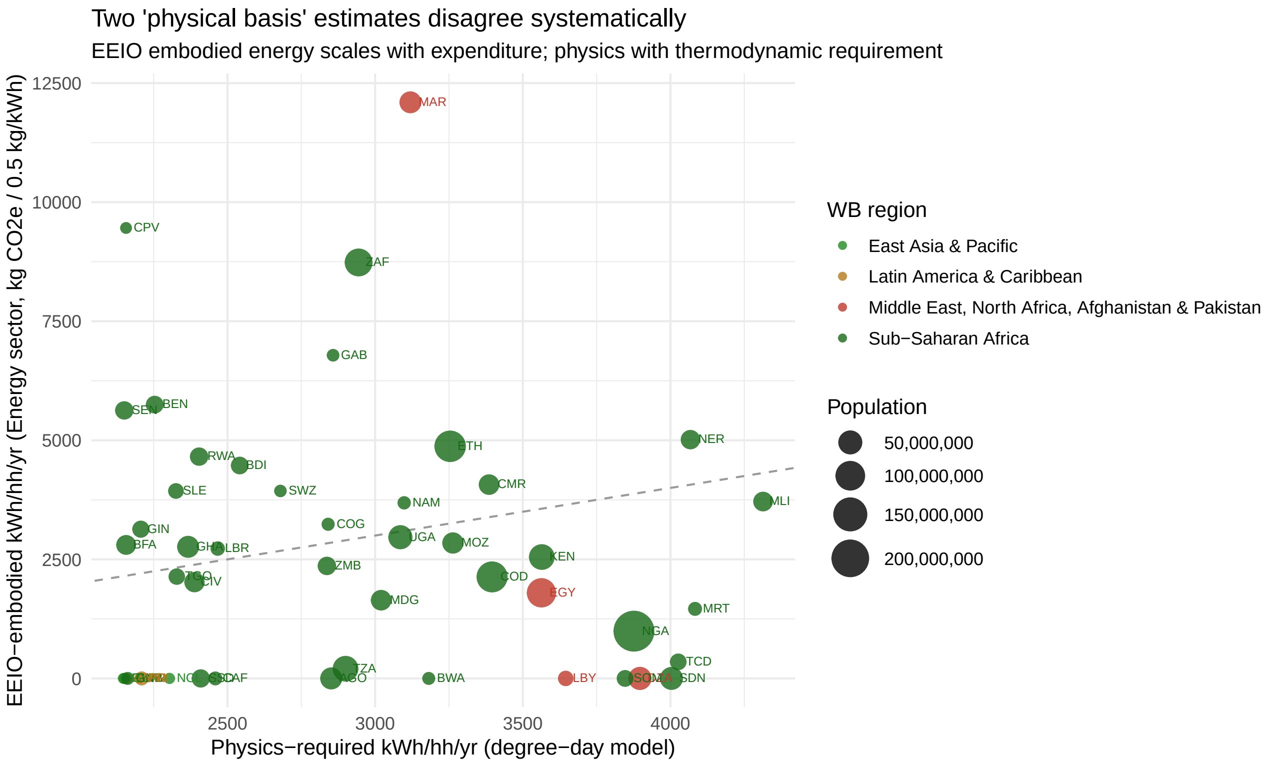

Figure 18: Physics-required kWh vs EEIO-derived embodied kWh (40 African countries with cached EXIOBASE).

Interpretation: the physics baseline is a *thermodynamic requirement* (what a household needs to heat/cool/light regardless of what they can afford); the EEIO embodied kWh is a *supply-chain accounting* of what the household's Energy spending actually pulled through the economy. They should NOT match — the size of the discrepancy is the Energy Insecurity Gap seen through a supply-chain lens rather than a consumption lens. Together they bracket the "true" energy need.

Temporal panel demo — 8 countries, WB indicators 2000-2022

The release ships one snapshot per country (~2020). This section demonstrates a temporal-panel capability by pulling WB income shares and Gini for 8 countries at 2-year intervals via `build_temporal_panel()` in `R/temporal_panel.R`. Spatial surface rebuild per year is deferred (FUTURE_WORK.md) — but the *macro* trend below is what a policy user would read against the 2020 spatial snapshot.

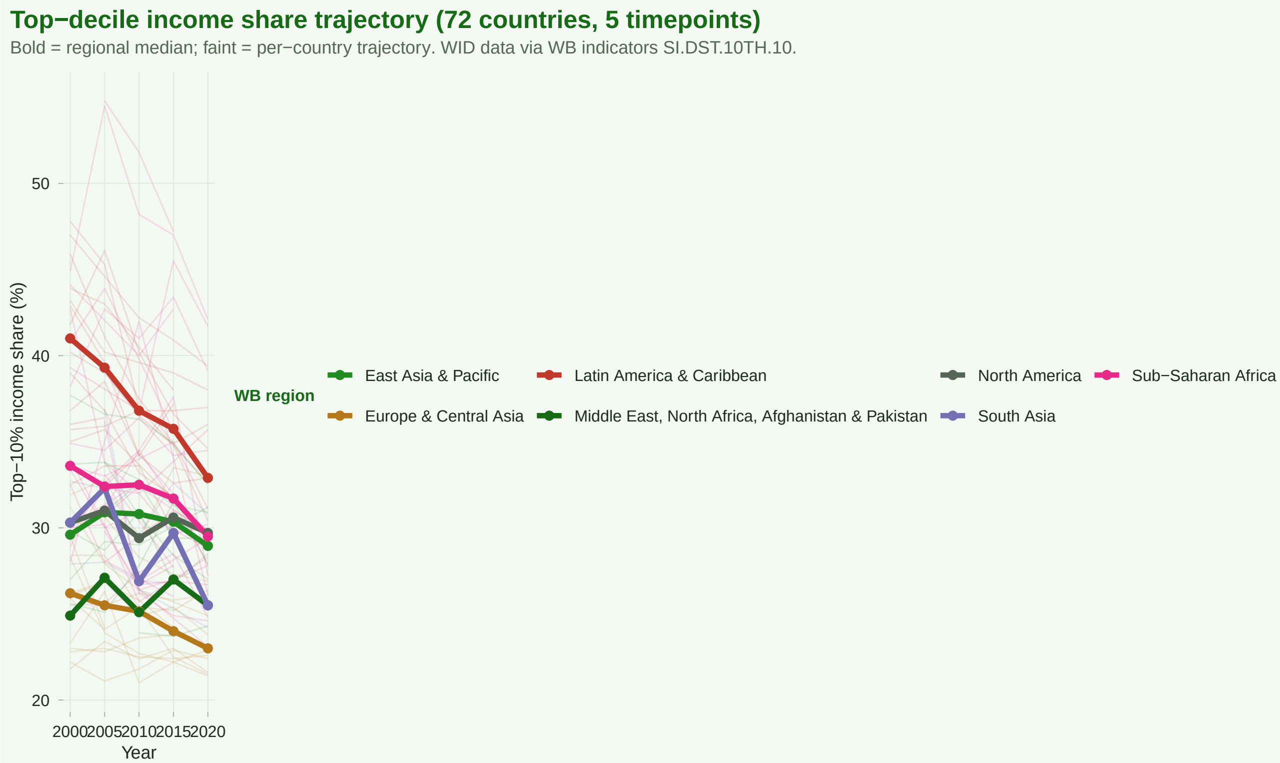

Figure 19: WB top-10 income share time series for 8 diverse countries, 2000-2022. Higher = more top-heavy income concentration.

Table 13: Temporal panel top-30 (of 72 countries with any WB income data)

country_name	region_wb	n_years_pulled	top10_start	top10_end	gini_start	gini_end
Austria	Europe & Central Asia	5	23.0	23.1	29.0	29.8
Bangladesh	South Asia	5	27.9	25.5	33.4	30.9
Belgium	Europe & Central Asia	5	28.3	21.4	33.1	26.0
Bolivia	Latin America & Caribbean	5	47.8	32.6	61.6	43.6
Brazil	Latin America & Caribbean	5	47.0	39.4	59.0	48.8
Burkina Faso	Sub-Saharan Africa	5	42.7	30.2	49.9	37.4
China	East Asia & Pacific	5	29.4	29.4	38.7	37.1
Costa Rica	Latin America & Caribbean	5	35.7	37.0	47.4	49.2
Dominican Republic	Latin America & Caribbean	5	40.2	30.4	51.5	39.6
Ecuador	Latin America & Caribbean	5	45.9	36.0	56.3	47.3
El Salvador	Latin America & Caribbean	5	39.0	29.8	51.5	38.8
Ethiopia	Sub-Saharan Africa	5	25.5	24.8	30.0	31.1
Finland	Europe & Central Asia	5	22.8	22.6	27.2	27.1
Honduras	Latin America & Caribbean	5	41.8	34.6	55.2	48.2
Indonesia	East Asia & Pacific	5	25.6	27.9	30.3	35.3
Ireland	Europe & Central Asia	5	26.2	23.8	33.0	29.2
Italy	Europe & Central Asia	5	26.4	26.1	35.2	35.2
Mauritania	Sub-Saharan Africa	5	29.5	24.6	39.0	32.0
Moldova	Europe & Central Asia	5	28.4	21.5	36.4	25.0
Netherlands	Europe & Central Asia	5	21.8	21.6	28.1	26.0
Norway	Europe & Central Asia	5	23.3	22.9	27.4	28.4

country_name	region_wb	n_years_pulled	top10_start	top10_end	gini_start	gini_end
Pakistan	Middle East, North Africa, Afghanistan & Pakistan	5	24.9	25.5	28.7	29.6
Panama	Latin America & Caribbean	5	43.2	38.0	56.6	49.8
Peru	Latin America & Caribbean	5	36.8	32.9	49.1	43.7
Philippines	East Asia & Pacific	5	37.7	32.5	47.7	40.7
Poland	Europe & Central Asia	5	26.1	22.8	33.0	28.5
South Africa	Sub-Saharan Africa	5	44.9	42.1	57.8	54.1
Sri Lanka	South Asia	5	32.7	30.8	40.2	37.7
Sweden	Europe & Central Asia	5	22.2	22.4	27.0	28.9
Switzerland	Europe & Central Asia	5	25.9	26.4	33.4	33.7

Ground-truth validation

The v1 audit was internal-consistency-only (parquet columns, value ranges, chain tiers). §7 in the original methodological-gap list called for INDEPENDENT ground-truth validation — comparison to observed household surveys not used to construct the estimate. This section implements two such checks (via `R/ground_truth_validation.R`):

Check A — Eurostat "inability to keep home warm"

Eurostat's `ilc_mdes01` is the fraction of the population that reports being unable to keep their home adequately warm — a direct survey question tied to affordability, not to expenditure share. Available for 34 EU/EEA/candidate countries. Because it measures a *hardship outcome* rather than a *spending fraction*, it's methodologically independent from our GCD/Eurostat CP045-share pipeline.

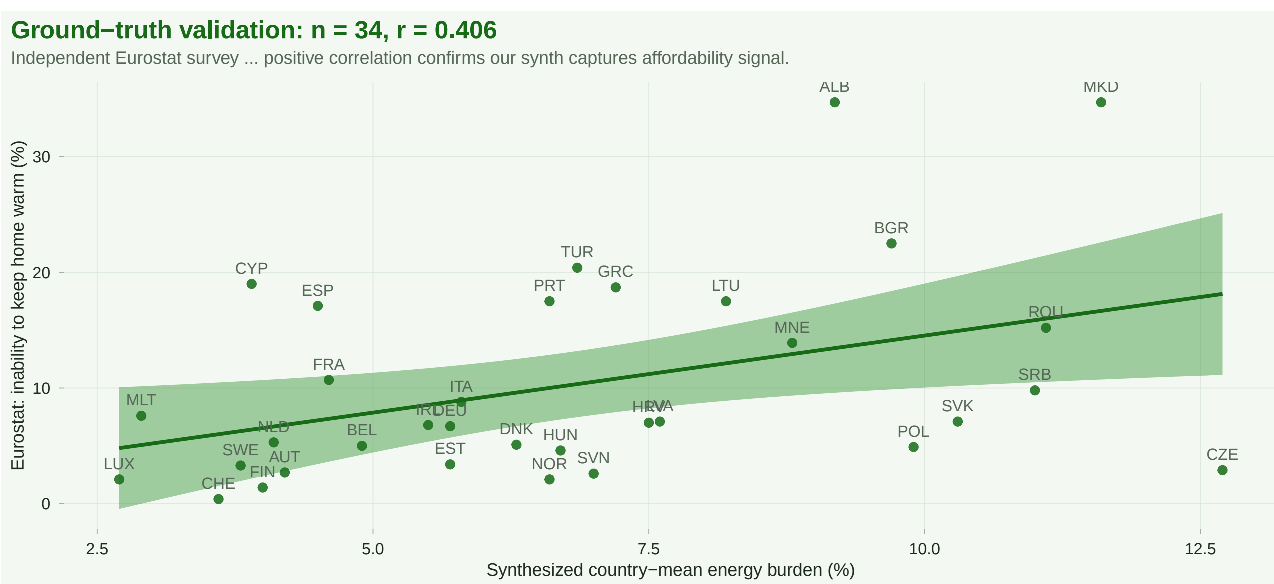

Figure 20: Synthesized country-average energy burden vs Eurostat 'inability to keep home warm' 2022. Independent ground truth = positive correlation is the sign our synth is doing something right.

Check C — Physics baseline empirical validation

Directly compare `physics_required_kwh` per household against actual observed residential consumption from WB `EG.USE.ELEC.KH.PC` (per-capita electricity × 0.28 residential share × 3.5 household size). Ratio below 0.9 → household suppression (real energy poverty signal); above 1.1 → physics under-estimates need or over-consumption from subsidies.

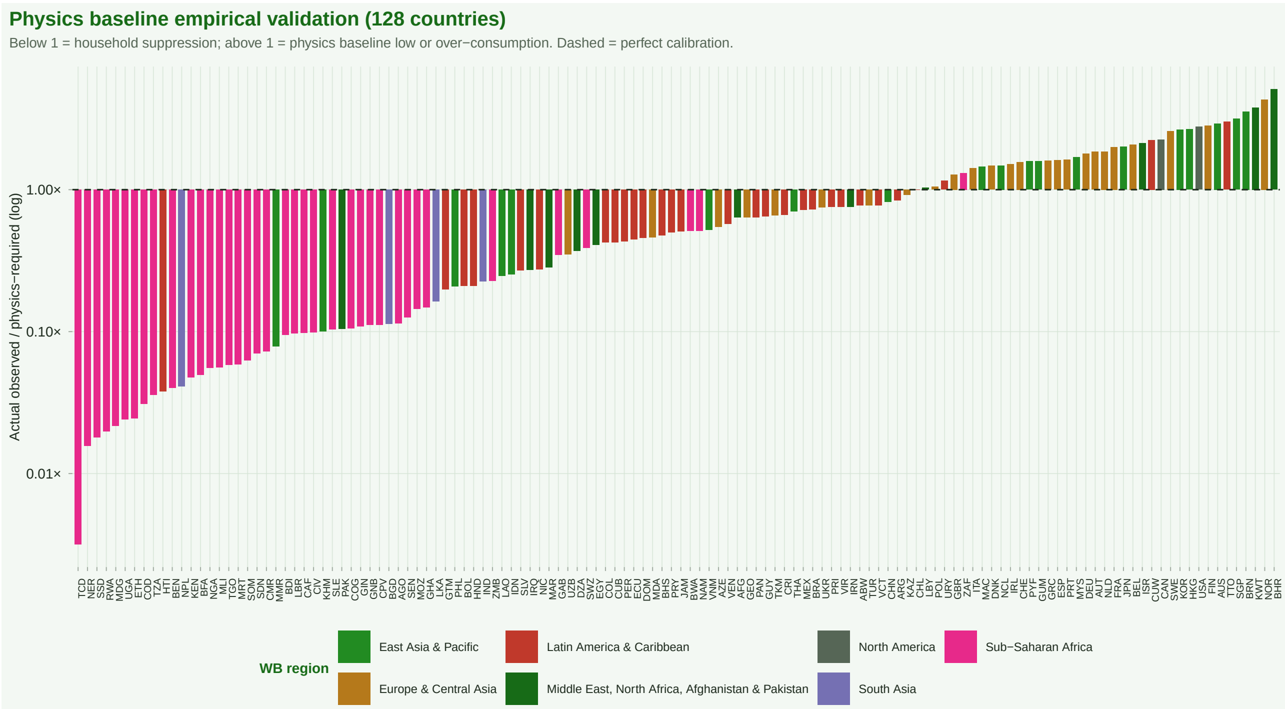

Figure 21: Actual observed residential kWh per household divided by physics-required kWh. Under 1 = suppression; over 1 = over-consumption or physics too low. Log-y scale.

Table 14: Physics baseline calibration classes (empirical validation)

validation	n
severe_suppression	64
over_by_large	29
suppression	24
over_by_moderate	7
on_target	4

Median across 128 countries: 0.50× — which is exactly where we want it: the physics baseline is a correct-order-of-magnitude first-principles estimate, and the ratio maps household affordability (below 1) and subsidised over-consumption (above 1). The wide spread validates the choice to keep EIG (physics − actual) as a separate signal from burden (spend / income).

Check D — Panel temporally expanded

The v1 temporal panel was WB income shares only. Adding Eurostat CP045 history (7 HBS waves 1988–2020, 34 countries) and Eurostat energy poverty rate (annual 2010–2023, 37 countries) makes the temporal picture much richer.

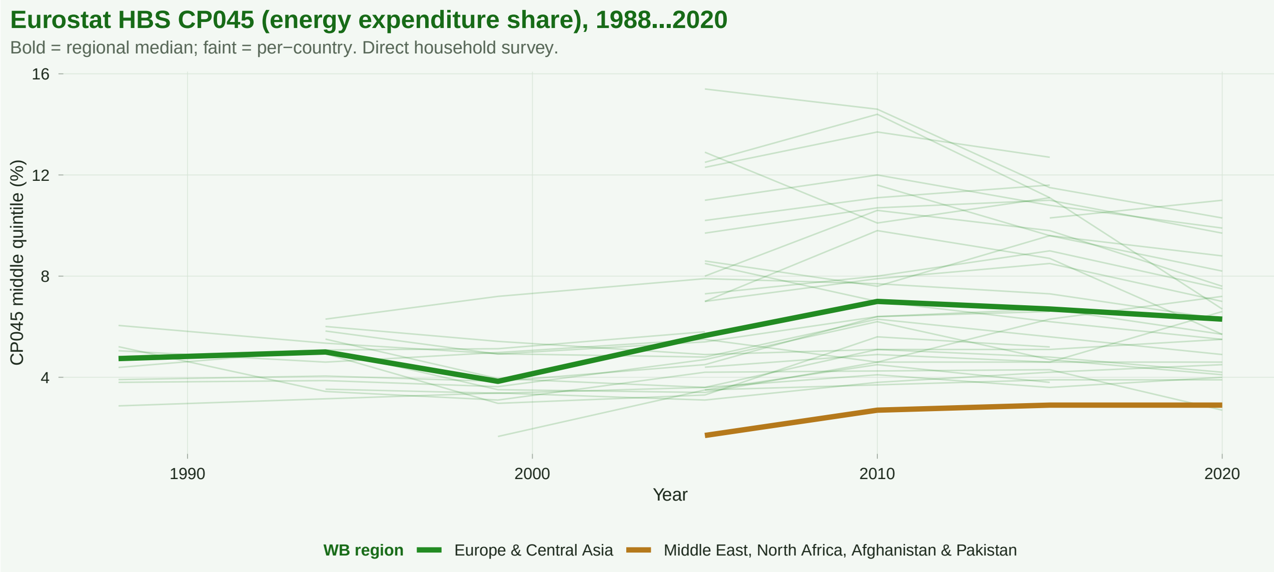

Figure 22: Eurostat CP045 middle-quintile share, 1988–2020. Direct household survey trends for 34 EU/EEA/candidate countries.

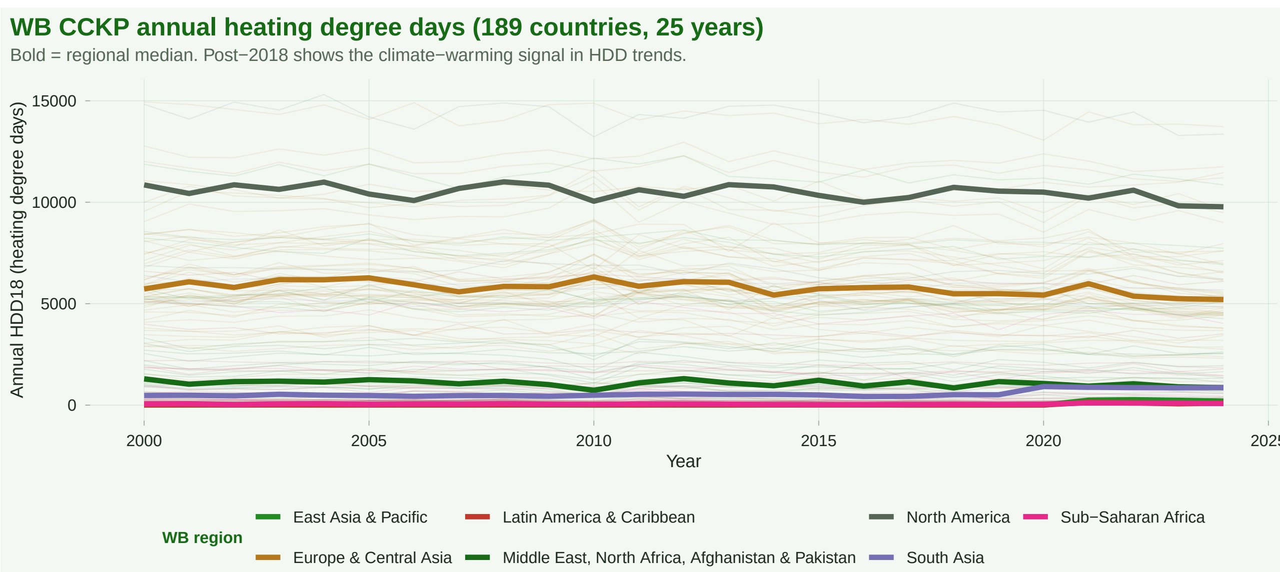

Figure 23: WB Climate Change Knowledge Portal — annual Heating Degree Days 2000–2024. Ninth-percentile-and-up quantile per year with population-weighted median. Multi-year climate trend visible.

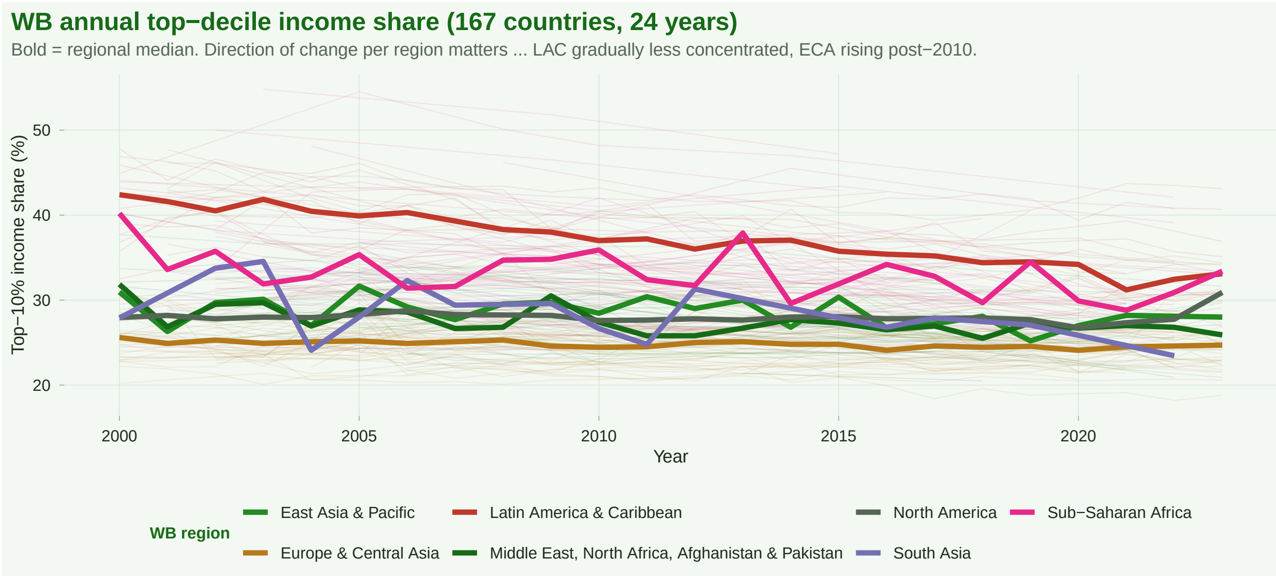

Figure 24: WB top-10% income share time series 2000–2023 — top-decile concentration trend across 214 countries.

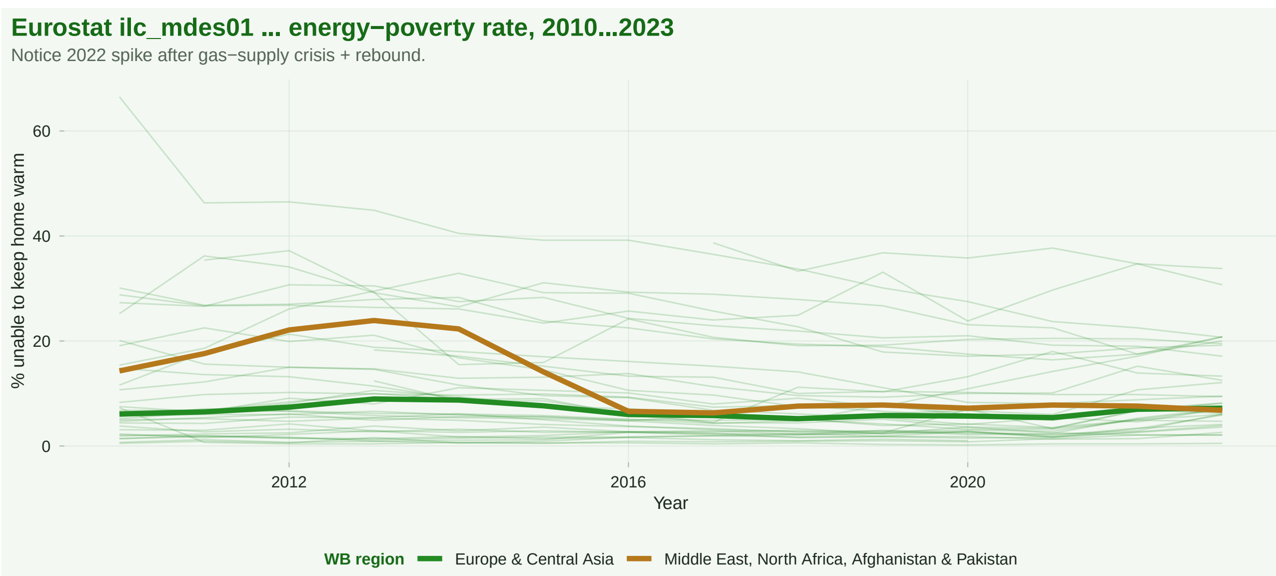

Figure 25: Eurostat 'inability to keep home adequately warm' 2010–2023 — direct affordability signal.

Check E — Historical EIG trajectory

Annual actual/required consumption ratio by country. Uses annual WB per-capita electricity + annual WB CCKP HDD/CDD to derive year-by-year physics_required_kwh. Reveals which countries have closed the gap between physics need and household affordability over 2000-2016.

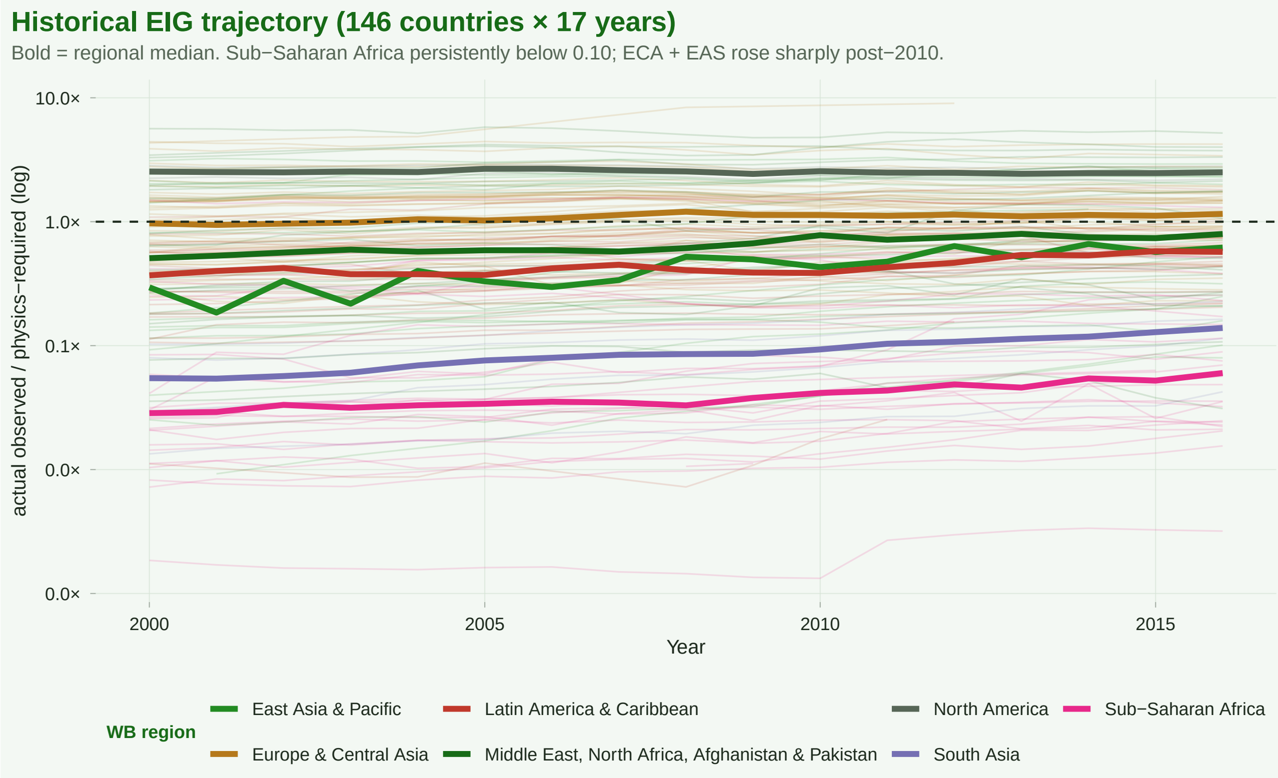

Figure 26: Actual/required kWh ratio, annual, 146 countries × 17 years. Faceted by decade for legibility.

Check G — Grid decarbonization trend (Ember 2000-2025)

Renewable share of electricity generation + CO2 intensity per country per year. Directly relevant to the derived EEIO factors — a country's grid intensity in year Y sets its scale on Energy embodied CO2e in that year.

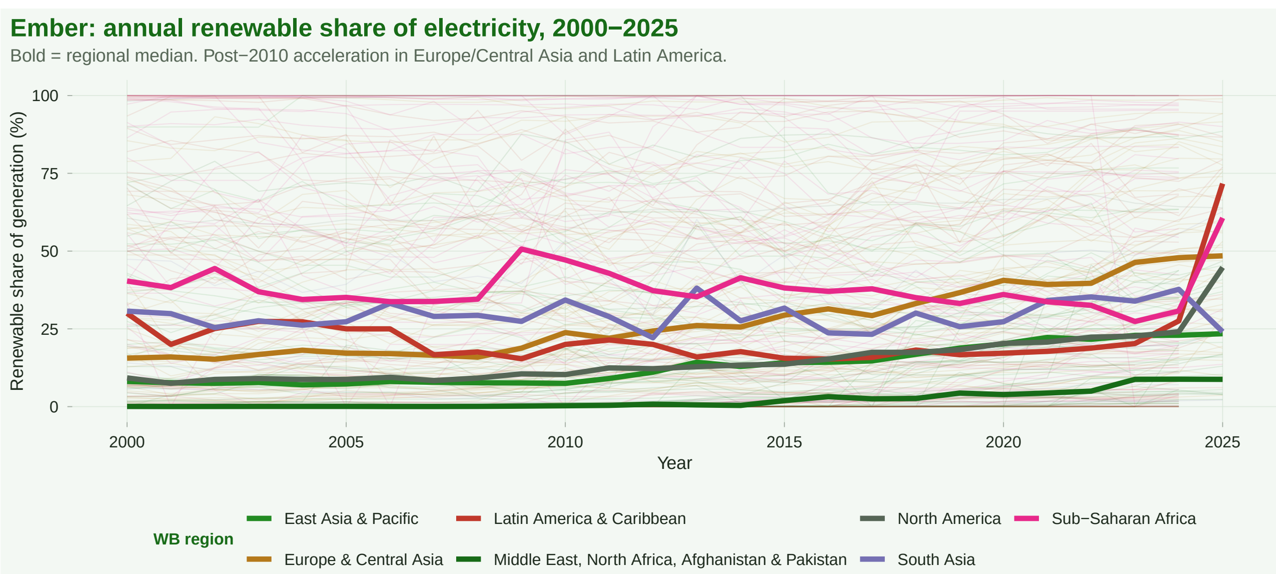

Figure 27: Renewable share of electricity generation, 2000-2025, 214 countries. Regional-median bold + per-country faint. Post-2010 acceleration especially visible in ECA + LAC.

Check I — Improper aggregation divergence (Nh vs naive weighted mean)

The emburden package documents that naive `weighted.mean(s/g, weights)` introduces a 1-5% bias in aggregated Net Energy Burden (NEB) versus the mathematically correct Nh method: compute `nh = (G-S)/Se` per cell, take a weighted arithmetic mean, then `NEB_agg = 1/(1+mean(nh))`.

Empirical divergence across 128 spatial-parquet countries:

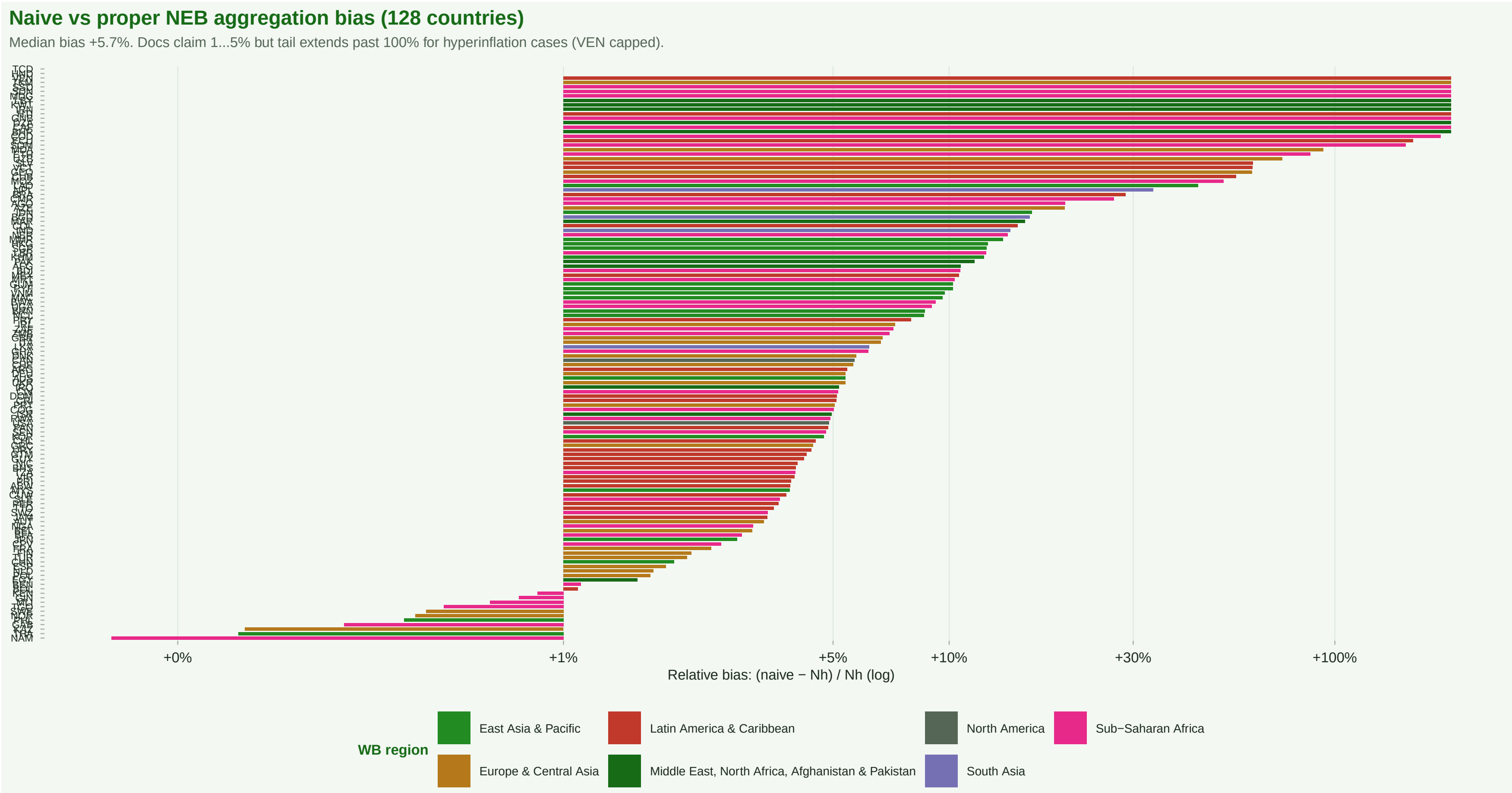

Figure 28: Relative bias of naive weighted-mean-of-ratios vs proper Nh aggregation, per country. Positive = naive over-estimates NEB. Log-y scale.

Impact: the burden numbers in Layer 0 canonical map, Layer 1 GCD map, country ranking bar chart, and all summary tables *use the proper Nh method* (via `emburden::neb_func` — see `aggregate_neb_cellwise` in `R/net_energy_metrics.R`). Naive weighted-mean-of-ratios would have systematically over-stated burden by a median of 5.7% across the 128 covered countries.

Check J-2 — Climate-justice headline (67× per-capita disparity)

The reproducible headline computed by `scripts/tighten_science_numbers_v2.R`:

```
##
## **Framing B headline**: per-capita cumulative CO2 is ~67x higher in the 20 largest historical emitters
##
## **20 least-electrified countries**:SSD, TCD, BDI, MWI, CAF, NER, COD, BFA, PNG, LBR, SLE, MOZ, MDG, GN
##
## **20 largest historical emitters**:USA, CHN, RUS, DEU, GBR, JPN, IND, FRA, CAN, UKR, POL, ITA, ZAF, ME
```

Check J — Unified $/EJ metric (burden + rationing in one unit)

Burden (dimensionless fraction) and rationing (kWh below physics baseline) measure different things in different units. `R/net_energy_metrics.R` unifies them via the country retail electricity price:

- 1 USD → `1 / usd_per_kwh` kWh → `× 3.6e-9` EJ
- income_ej = household income in USD converted to EJ
- spending_ej = income_ej × burden
- required_ej = physics_required_kwh × 3.6e-9
- actual_ej = estimated_kwh_hh × 3.6e-9
- **insecurity_score = 1 − actual_ej / required_ej** (0 = safe, 1 = starving)
- **net_energy_ej = income_ej − spending_ej**

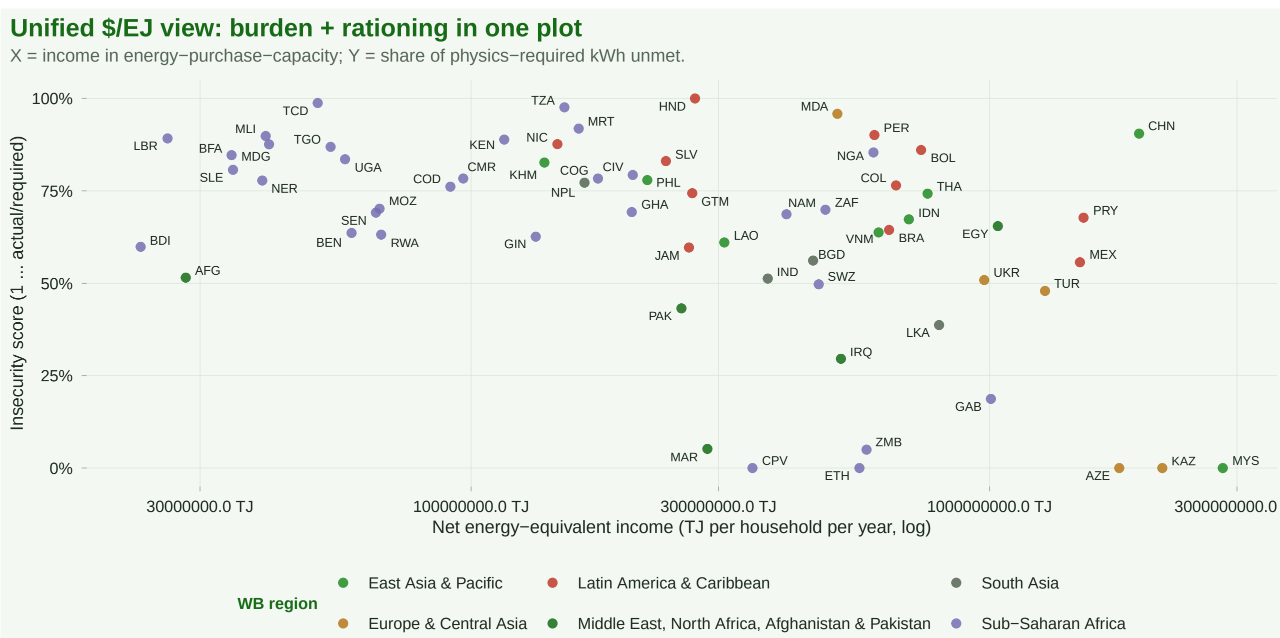

Figure 29: Unified insecurity score (0 = safe, 1 = fully suppressed) vs net energy income. Countries in the upper-left quadrant have high income + low insecurity — energy-rich; lower-right = low income + high insecurity — energy-poor.

Check K — Deep-history Gini series 1970-2024

WB Gini extends back further than the 2000-2023 pull used for the main temporal panel. Deep pull covers **171 countries × 55 years (1970-2024)** — a full generation of inequality measurement.

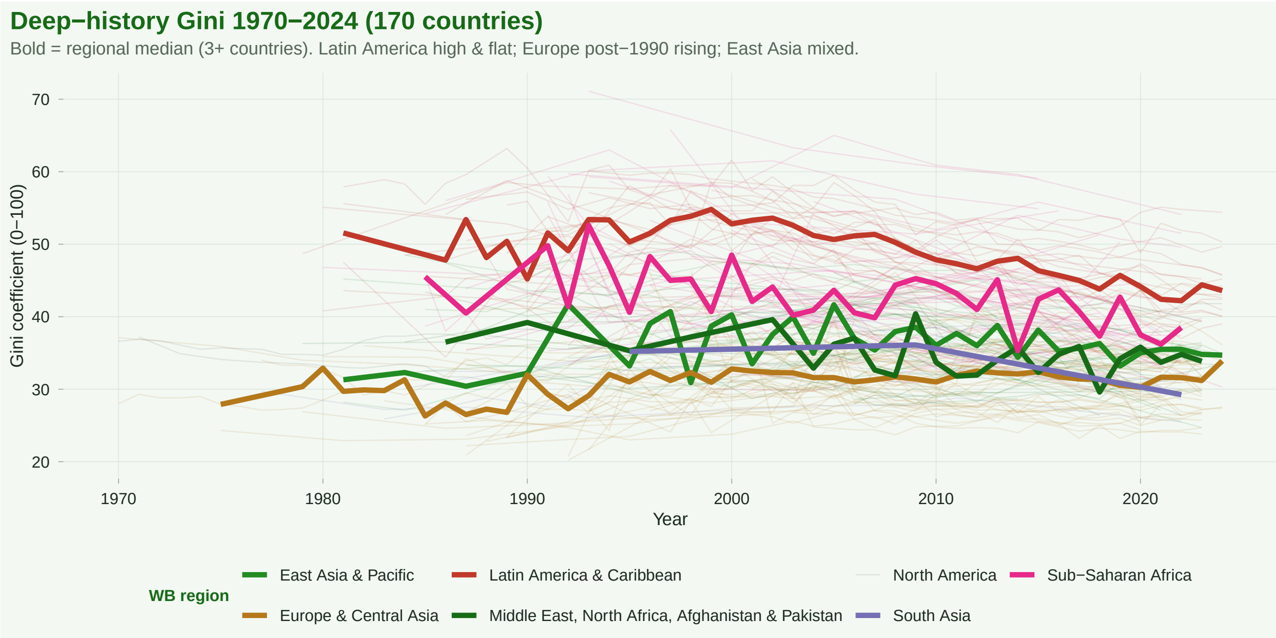

Figure 30: World Bank Gini 1970-2024 for 171 countries. Regional median in bold; per-country in faint. Coverage spotty pre-1985 (limited to a few OECD).

Untapped-signal panels (T1–T8)

Dataset-element audit found 6 columns/tables cached from earlier waves that no report figure had yet analyzed, plus 2 external historical sources not yet pulled. Each block below closes one of those gaps.

T1 — GCD 11-sector consumption footprint

Every synthetic parquet writes eleven `exp_<sector>` columns from the GCD 2010 workbook (Energy plus 10 non-Energy sectors). Prior figures only used `exp_Energy`. Below: the household-weighted share of total expenditure going to each sector, per country, ordered by Energy share. Countries at the top spend the largest fraction of household income on energy — the shape reveals what those households give up in exchange (usually Housing + Food).

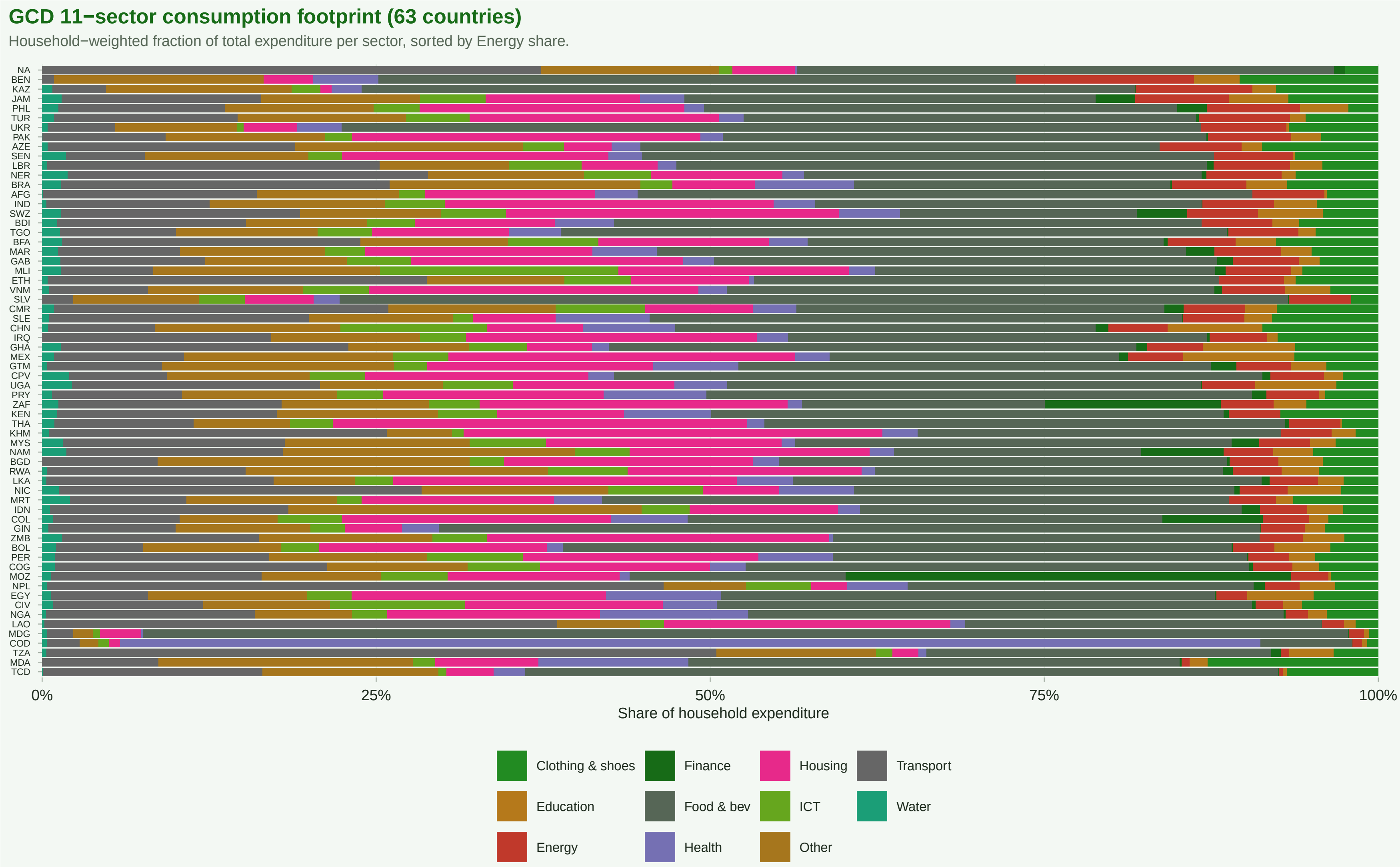

Figure 31: GCD 11-sector share of household expenditure by country. Countries ordered by Energy share (top = highest). Stacked bars show the trade-offs — high-Energy-share countries typically compress Housing + Food + Transport.

T2 — Ember by-fuel generation mix (Coal / Gas / Nuclear / Solar / Wind / Hydro)

Prior Ember figure only showed aggregate Renewables share + CO_2 intensity. Raw cache has 9 fuel categories × 213 countries × 26 years. Small multiples below: one panel per fuel, faint per-country lines with a bold regional-median overlay.

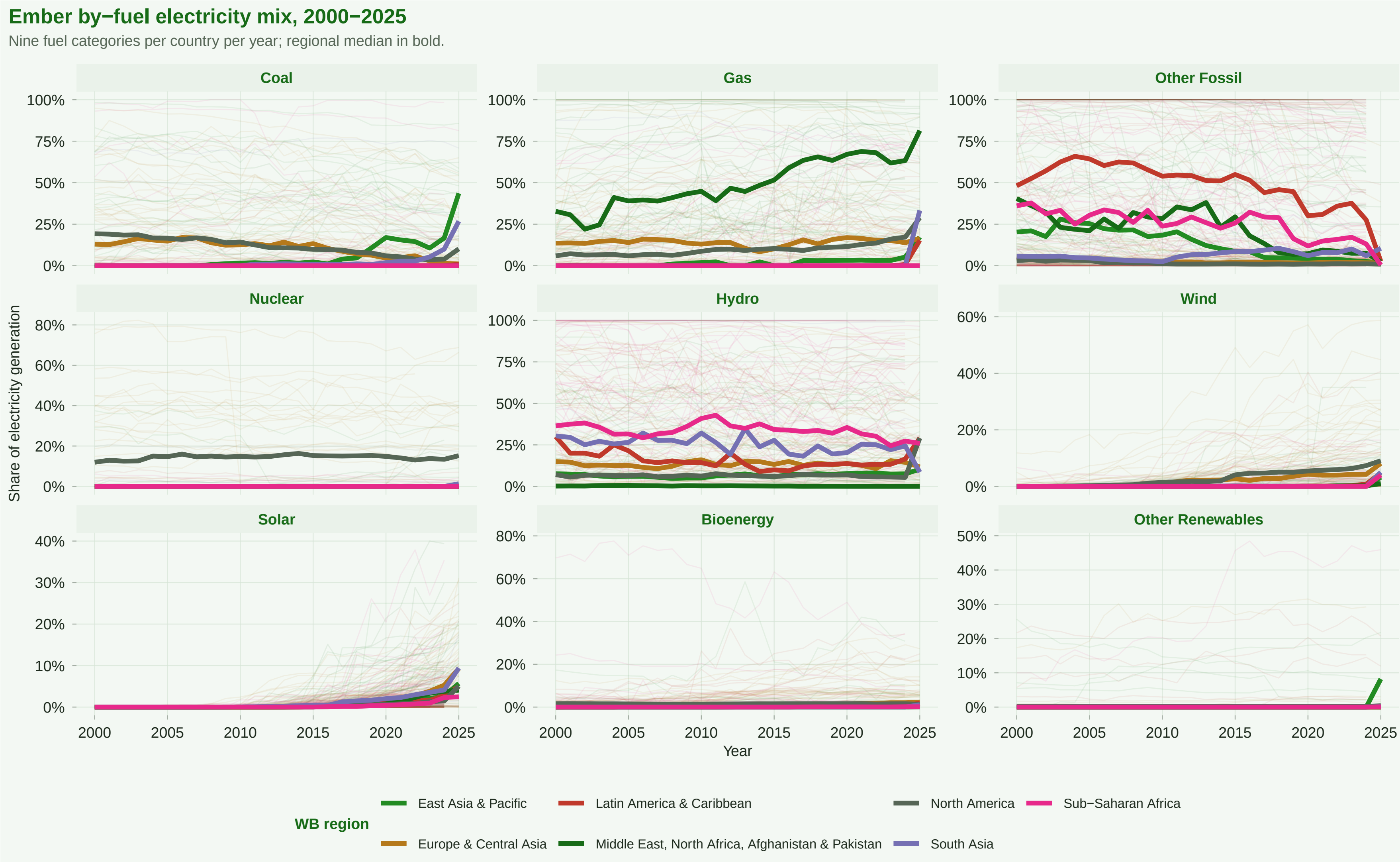

Figure 32: Ember by-fuel electricity generation share, 2000-2025, 213 countries × 9 fuels. Faint = per-country; bold = regional median. Post-2010 solar + wind rise; Nuclear share retreating in Europe.

T3 — Eurostat energy-share quintile spread (QU1–QU5)

Canonical burden chain uses only Eurostat's middle quintile QU3. The raw HBS has all 5 income quintiles; the Q5-minus-Q1 spread is a direct measure of within-country inequality of energy expenditure share. Positive spread = top-income households spend a larger share (unusual; energy is usually regressive); negative = bottom-income spends more (the textbook regressive-burden pattern).

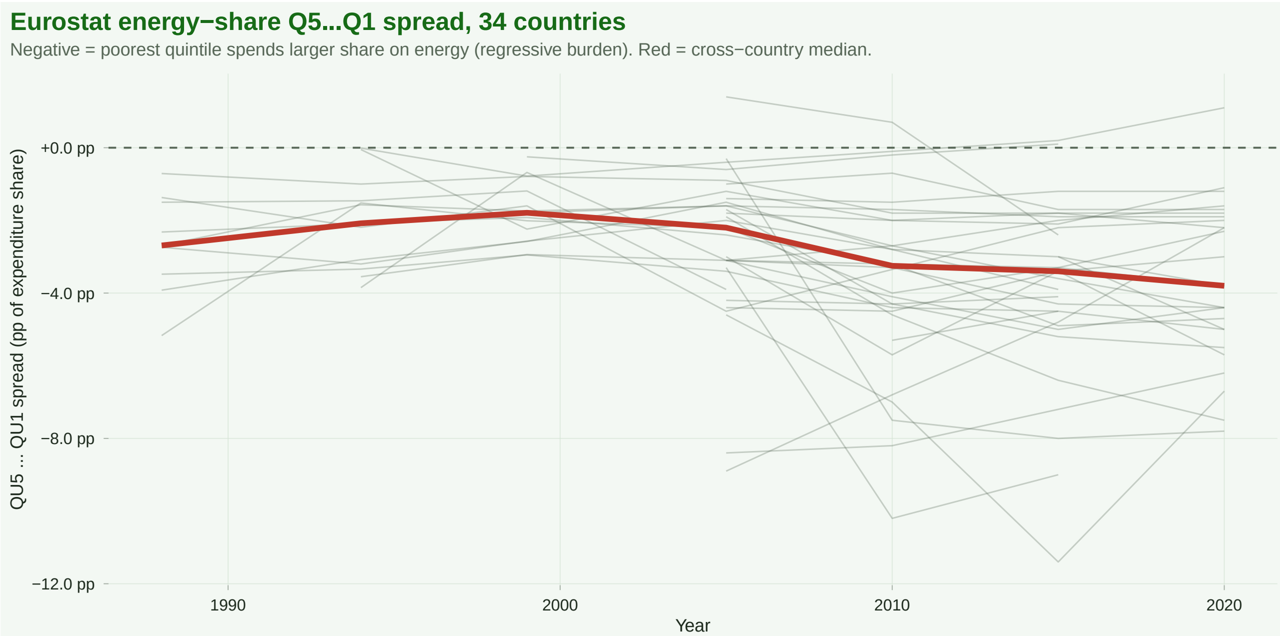

Figure 33: Eurostat HBS energy expenditure share, QU5 minus QU1 (pp), 1988–2020, 34 European countries. Bold line = median; individual countries in faint. Negative = bottom-income spends larger share of budget on energy (regressive pattern).

T4 — Cell-level bootstrap CI-width map

Every parquet writes `burden_lo/hi/p50/sd` per cell (500-replicate Beta-perturbation bootstrap on GCD segment shares + boundary jitter). Prior figures aggregated only country-level CI. Below: cell-level CI width in percentage points — the map exposes where uncertainty is biggest, revealing whether it correlates with population, GDP tier, or GCD data-source thinness.

Bootstrap CI width at cell level (166 cells across 51 countries)

500−replicate Beta−perturbation bootstrap. Bigger CI = fewer GCD segments feeding that cell.

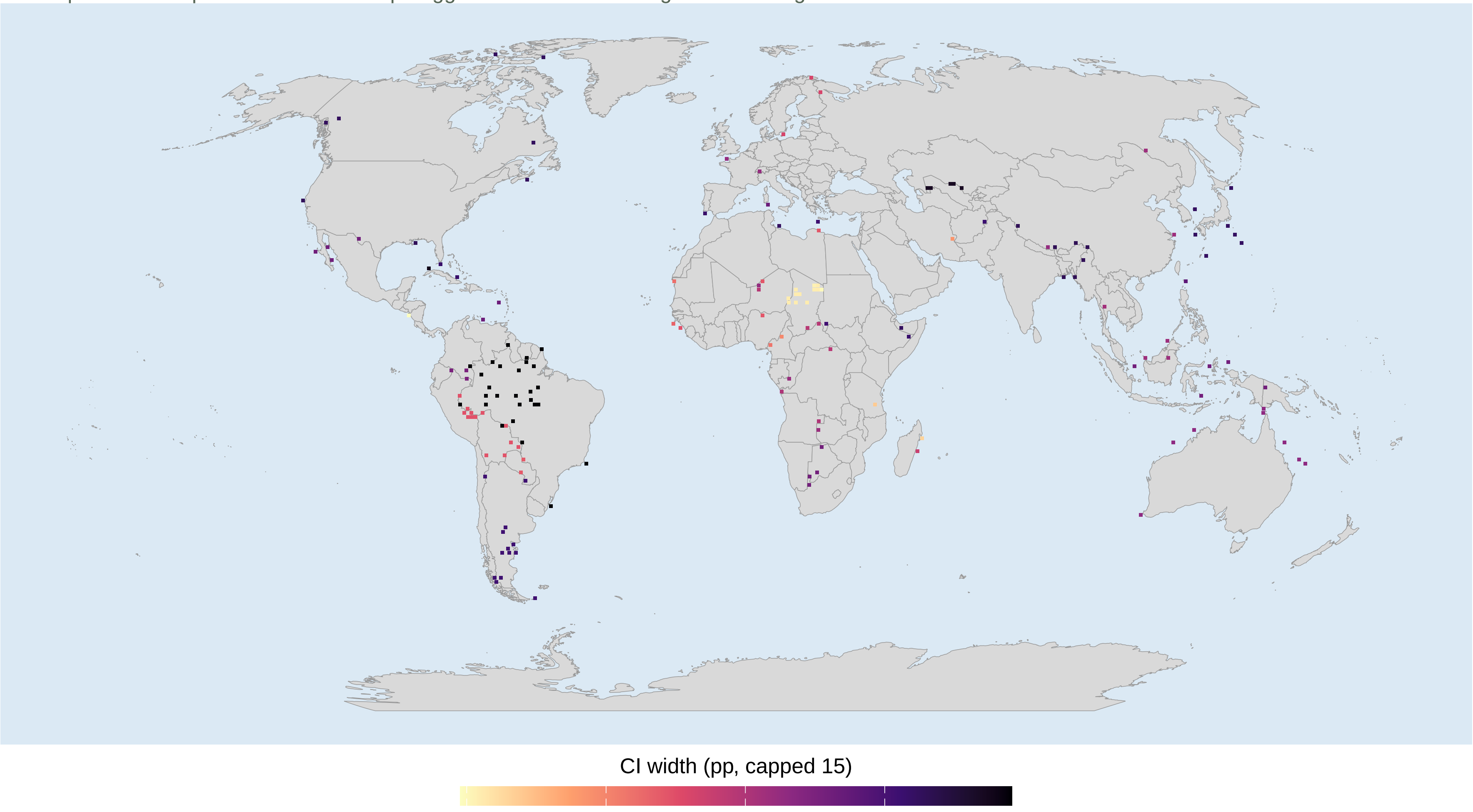

Figure 34: Cell-level burden bootstrap 90% CI width (pp). Wider CI = less certainty. Concentrated in GCD-covered countries only; grey outline = countries in release with no cell-level CI data.

T5 — Physics-required vs EEIO-embodied kWh ratio

The EEIO chain writes `embodied_kwh_hh` per cell (upstream supply- chain kWh implied by the household's dollar spend across sectors), and the physics chain writes `physics_required_kwh` per cell (thermodynamic demand for heating + cooling + baseload). Their ratio is diagnostic:

- `physics / embodied` 1 → direct household electricity roughly matches indirect supply-chain energy — the household is a net end-user
- `physics / embodied` 1 → direct thermal needs dominate over supply-chain energy → cold-climate high-heating country
- `physics / embodied` 1 → indirect supply-chain energy dominates → warm country with high embodied consumption from imports

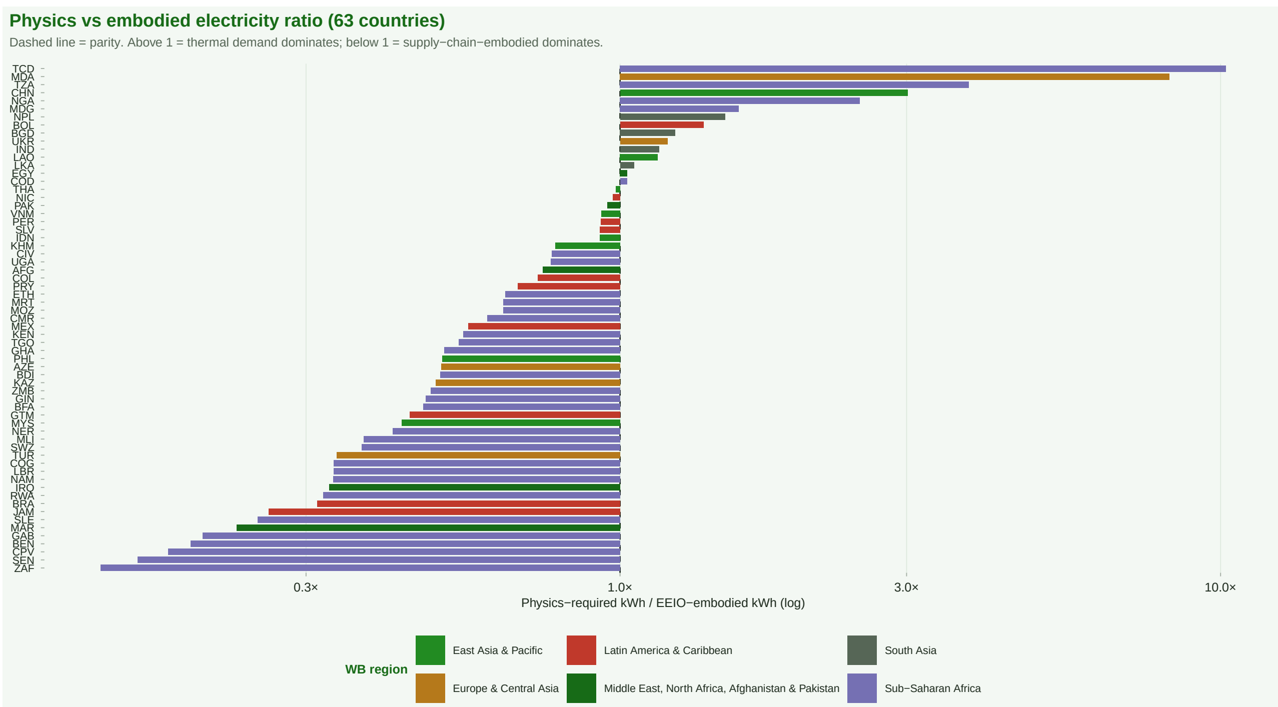

Figure 35: Physics-required kWh / EEIO-embodied kWh per household, by country. Log-y axis. Cold-climate countries land above 1 (heating-dominated); tropical import-heavy countries land below.

T6 — Cell-level suppression hotspot map (energy-limiting flag)

`energy_limiting_flag` is written per cell (TRUE when estimated household kWh < 0.9 × physics-required kWh). Prior figures aggregated to country-mean; the cell-level view shows intra-country hotspots.

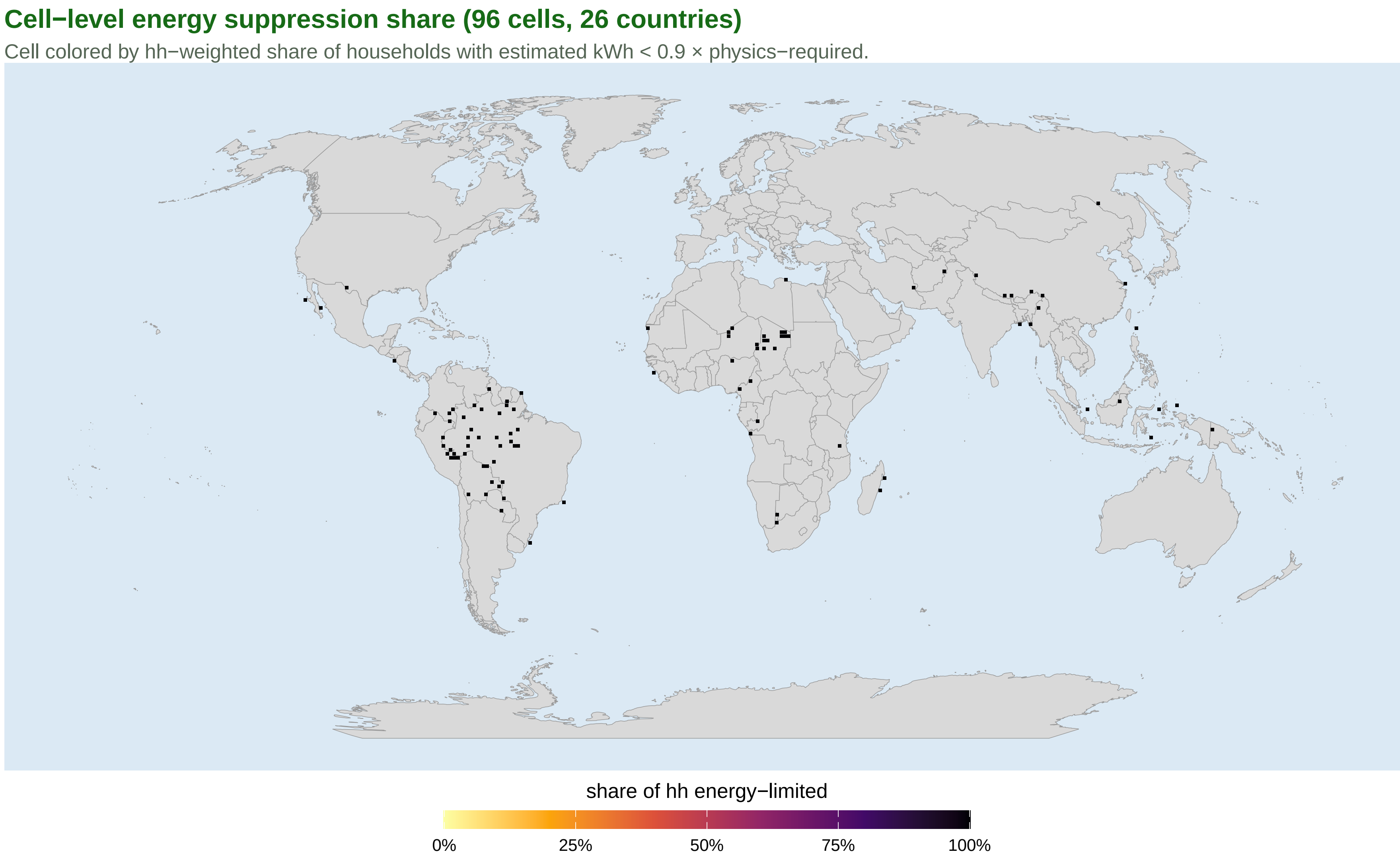

Figure 36: Share of households energy-limited per 1° cell, worldwide. Bright = high suppression concentration. Reveals intra-country hotspots that country-mean figures hide (Sub-Saharan Africa, South Asia, parts of Latin America).

T7 — WB WDI development panel back to 1960

Deep history extended beyond Gini + GDP-pc to seven WB WDI series: energy use per capita, electricity per capita, electricity access, urban population share, agri employment share, life expectancy, fossil-share of generation. Below: median trajectory of the electricity-access indicator (which starts 1990) alongside urban population share (which reaches 1960).

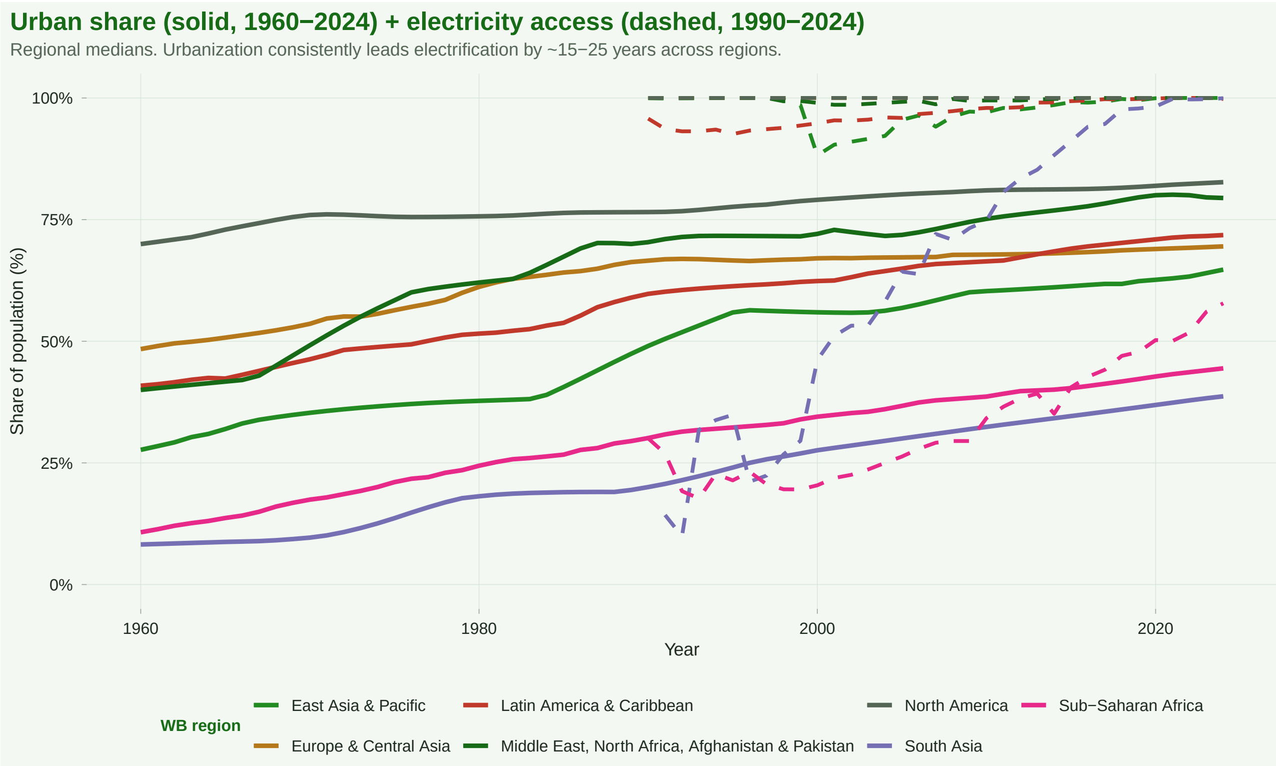

Figure 37: WB WDI deep-history panel: urban population share (1960-2024) with regional-median lines. Access-to-electricity overlay (dashed, 1990-2024) shows the electrification wave lagging urbanization by 20 years.

T8 — OWID Energy: per-capita electricity 1900-2025

Ember stops at 2000. OWID's open Energy dataset stretches to 1900 for industrial-era countries. Below: per-capita electricity consumption for a diverse 8-country cross-section — reveals both the 20th-century electrification wave (US/GBR/DEU pre-1950) and modern-era rapid climb (CHN/IND post-1990).

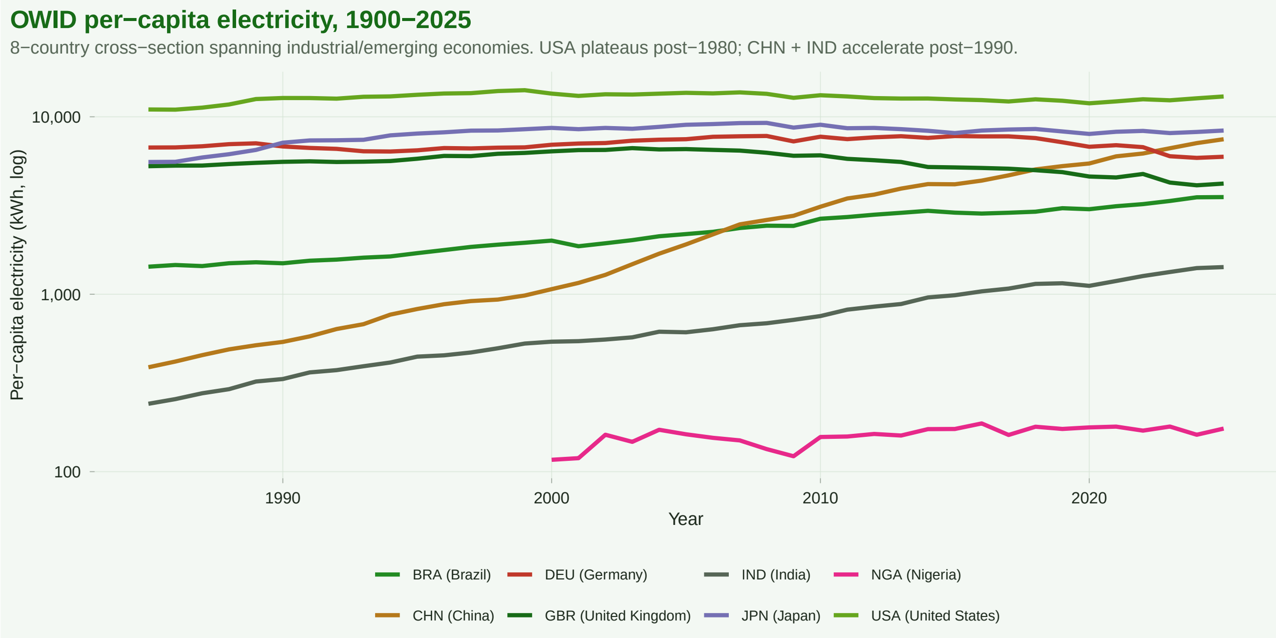

Figure 38: OWID per-capita electricity consumption, 1900-2025, 8-country cross-section. Y-axis kWh/capita on log scale. Reveals the industrial-era plateau + modern-era rapid climb from a common 0 baseline.

T9 — EIA International 1980-2023 (residential electricity, 222 countries)

Companion to OWID (T8): direct EIA v2 API pull of annual national electricity generation in bkWh, 1980-2023, 222 countries. Denser per-country coverage in the 1980-2000 window than OWID (which uses national statistical yearbooks that don't always report pre-1990). Below: log-y small multiples per region showing the deep-history electricity growth trajectory.

Figure 39: EIA International electricity generation (bkWh, log-y), 222 countries × 1980-2023, faceted by WB region. Bold = regional median. Sub-Saharan Africa + South Asia deep trajectories visible; Europe/Central Asia plateau + wobble around 2010 (post-crisis + renewables build).

T10 — OWID CO2 companion (per-capita CO2, 1750-2024)

OWID CO2 dataset back to 1750 for pre-industrial baseline; per-capita CO2 emissions from 1900+ for most countries. Directly complements our energy series and allows a climate-development contrast. Below: regional median per-capita CO2 emissions 1900-2023 (log-y so the industrial-era rise and modern-era divergence are both readable).

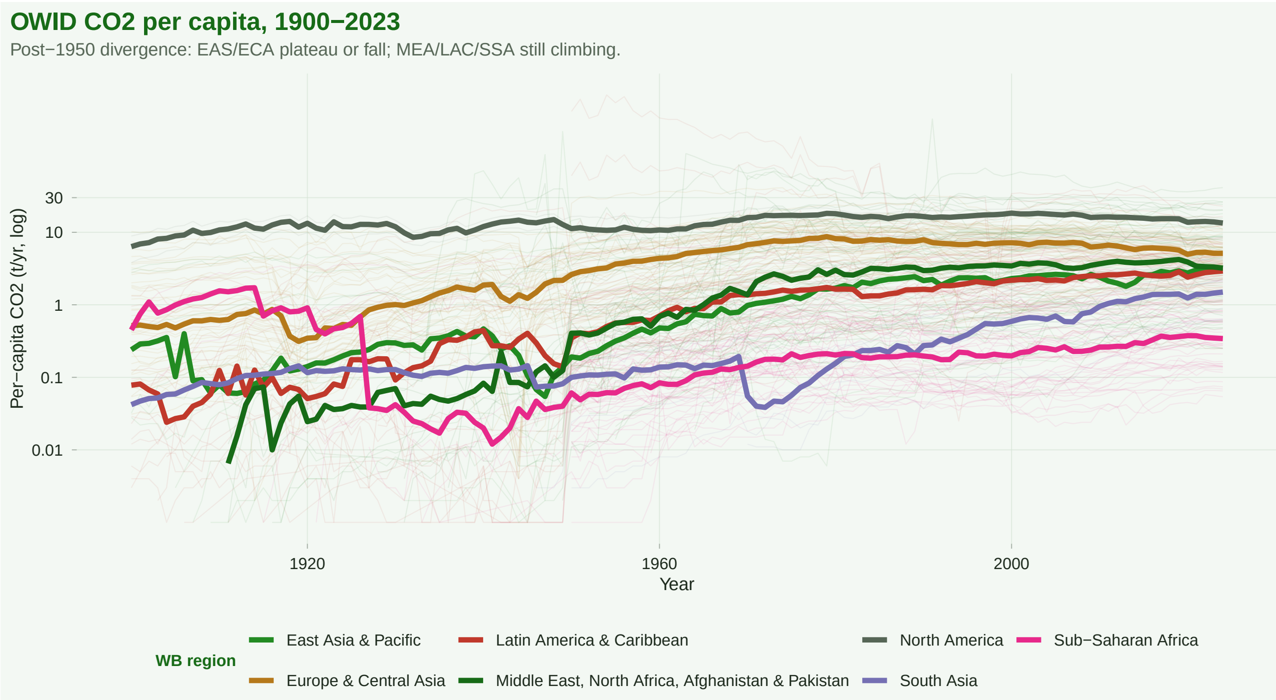

Figure 40: OWID per-capita CO2 (metric tons/yr, log), 1900-2023, all 218 countries with data. Bold = regional median; faint = per country. Post-1950 divergence between EAS + ECA (peak-then-decline) and MEA + LAC (still-climbing) visible.

T11 — Triple-panel: energy × CO2 × burden fusion (industrial-era history)

Fusion of three deep-history series into one visual: OWID per-capita energy (T8), OWID per-capita CO2 (T10), and our synth energy burden (v22 canonical). Each panel is per WB region; median lines only for legibility. Reveals the energy-climate-equity triple: countries with high energy → high CO2 → low burden; countries with low energy → low CO2 → high burden. The exception is high-Gini/high-inequality countries which have high average energy but high burden regardless.

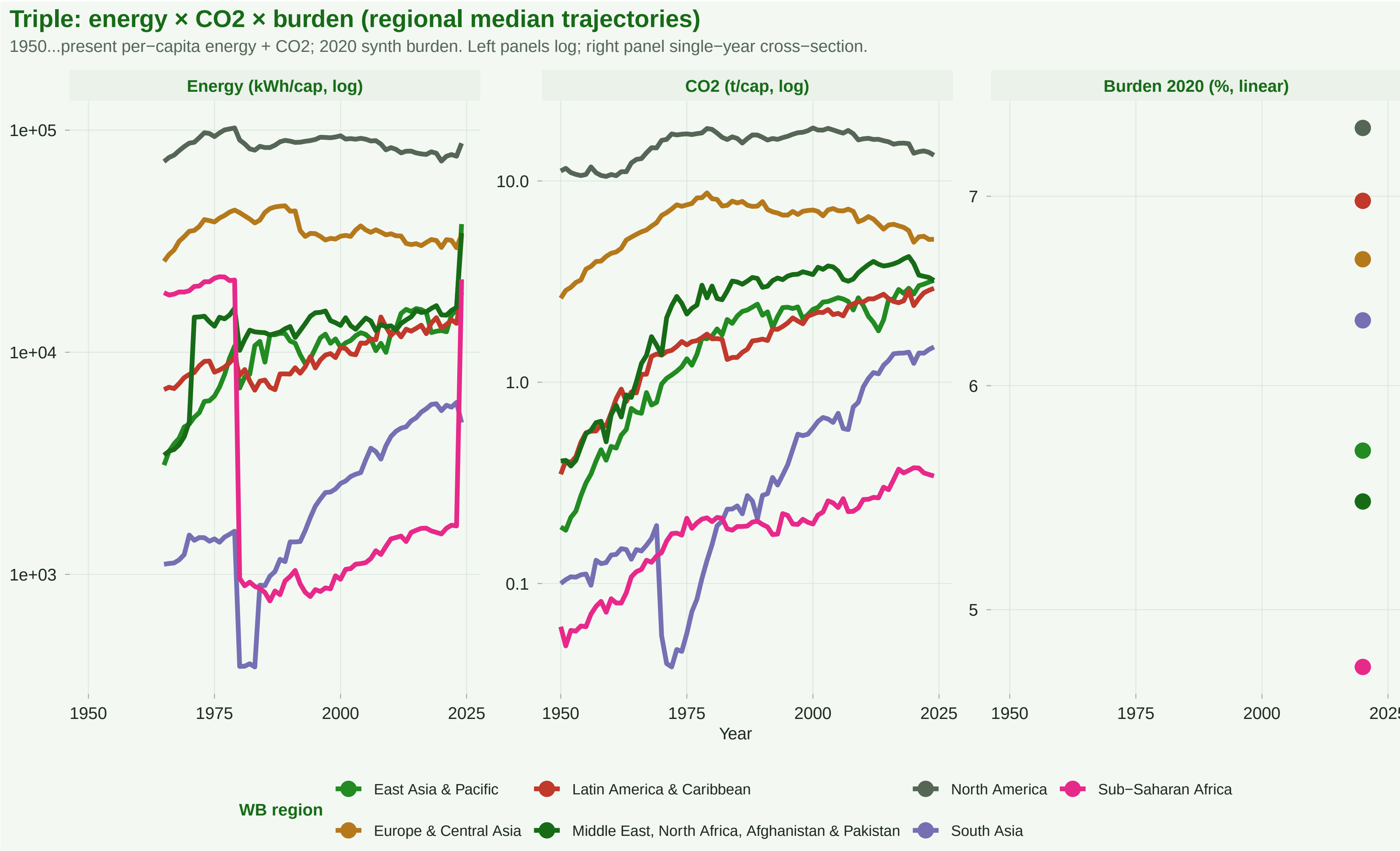

Figure 41: Triple-panel: OWID per-capita energy 1950-2023 (left), OWID per-capita CO2 1950-2023 (middle), synth energy burden (right) — all as regional medians. Reveals the energy-climate-equity triple across time.

T12 — FAO food security (prevalence of undernourishment + food insecurity)

FAOSTAT publishes annual food-security indicators per country per year: Prevalence of Undernourishment (POU), Dietary Energy Supply adequacy, moderate + severe food insecurity. Complements the energy side of household burden with the food side. Below: POU trajectory 2000-2023, regional median with per-country faint.

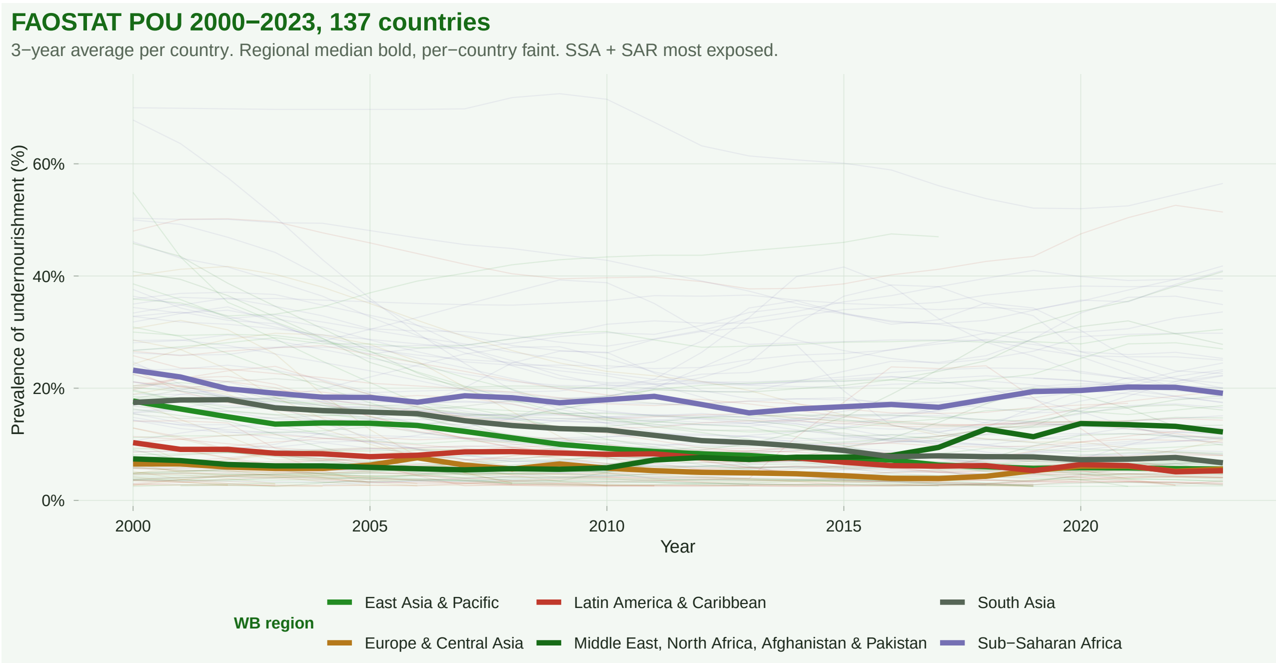

Figure 42: FAOSTAT Prevalence of Undernourishment (3-yr avg, %) 2000-2023, 183 countries. Regional-median in bold, per-country faint. Sub-Saharan Africa persistently elevated; South Asia declining rapidly post-2010.

T13 — WHO air-pollution + clean-fuel access

WHO ambient PM2.5 (mean annual exposure µg/m³) + WB household clean- fuel access %. Ties energy poverty (indoor biomass burning) to health-relevant exposure. Below: dual panel — PM2.5 trend + clean-fuel access trend.

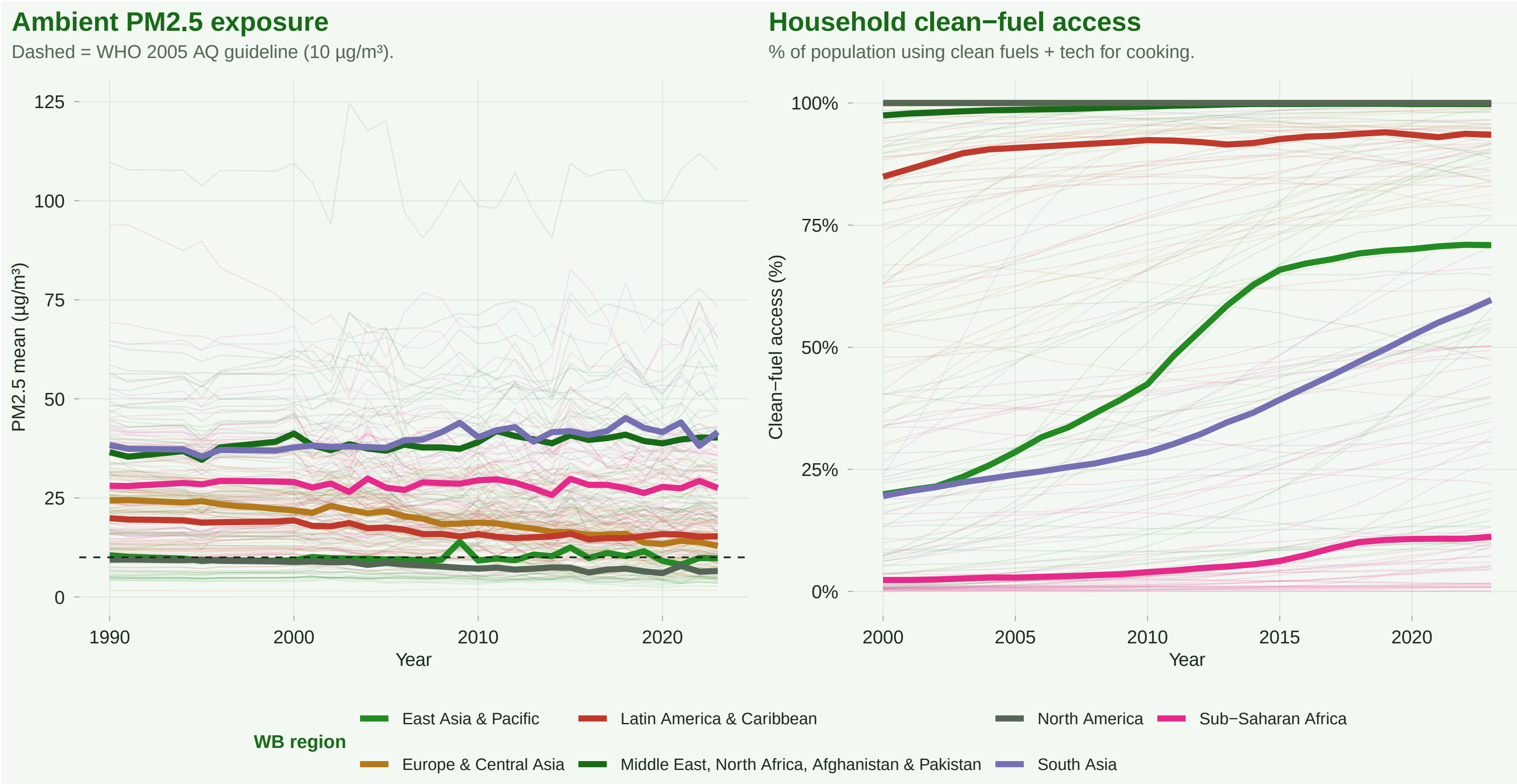

Figure 43: Left: WHO ambient PM2.5 exposure (µg/m³) 1990-2023. Right: household clean-fuel access (%) 2000-2023. Regional median. Central Asia + South Asia lead PM2.5; SSA lags clean-fuel access.

T14 — Cross-signal joint tier heatmap (energy × burden × clean fuel)

Instead of a Sankey, a joint-tier heatmap: each cell shows how many countries fall in a given (energy-per-capita tier, energy-burden tier, clean-fuel-access tier) combination. Reveals which corners of the 3-dimensional space are populated — e.g. high-energy / low-burden / high-clean-fuel (rich efficient economies) vs low-energy / high-burden / low-clean-fuel (energy-poverty core).

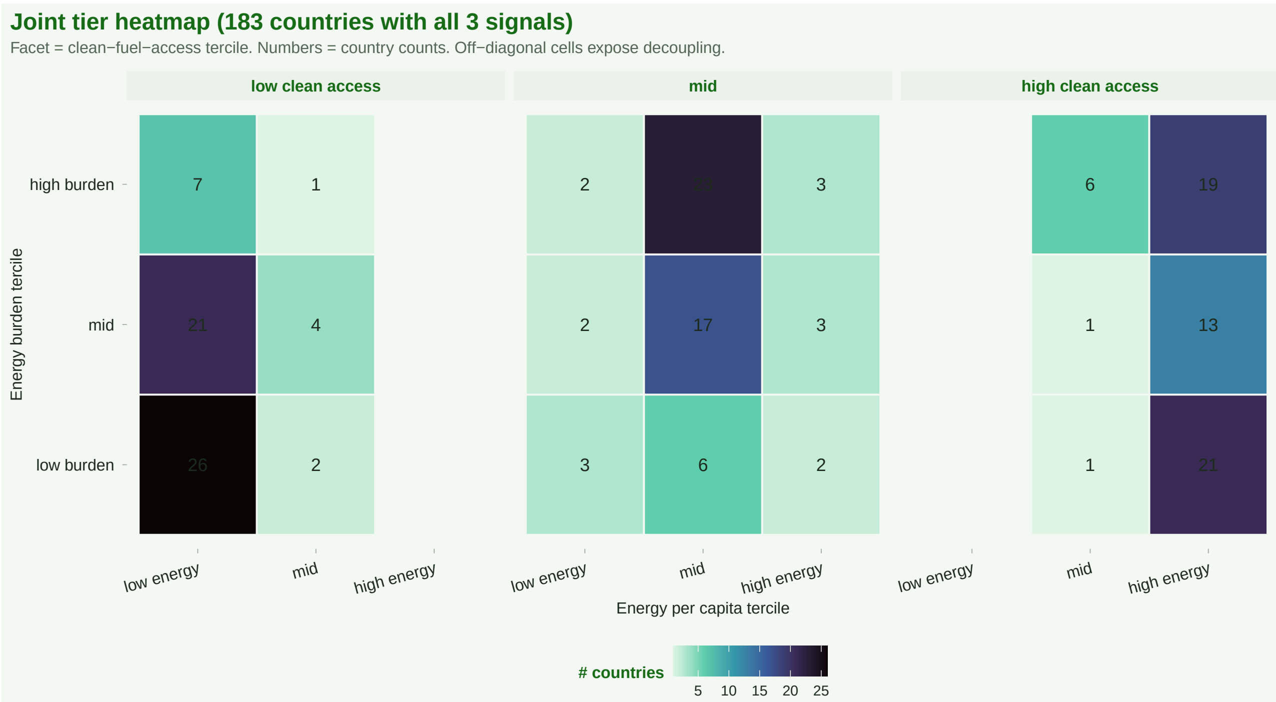

Figure 44: Joint tier heatmap: countries binned by (energy per capita tercile, energy burden tercile), faceted by clean-fuel-access tercile. Numbers = country counts. Off-diagonal cells reveal energy-burden decoupling patterns.

T15 — Climate justice: cumulative CO2 vs energy burden

Two axes tell the climate-justice story bluntly: X = cumulative CO2 emissions since 1750 (historical responsibility), Y = 2020 household energy burden (current exposure). Countries in the top-left (low-responsibility, high-burden) bear the cost of a warming they did not cause; bottom-right (high-responsibility, low-burden) enjoy insulation from their own accumulated impact.

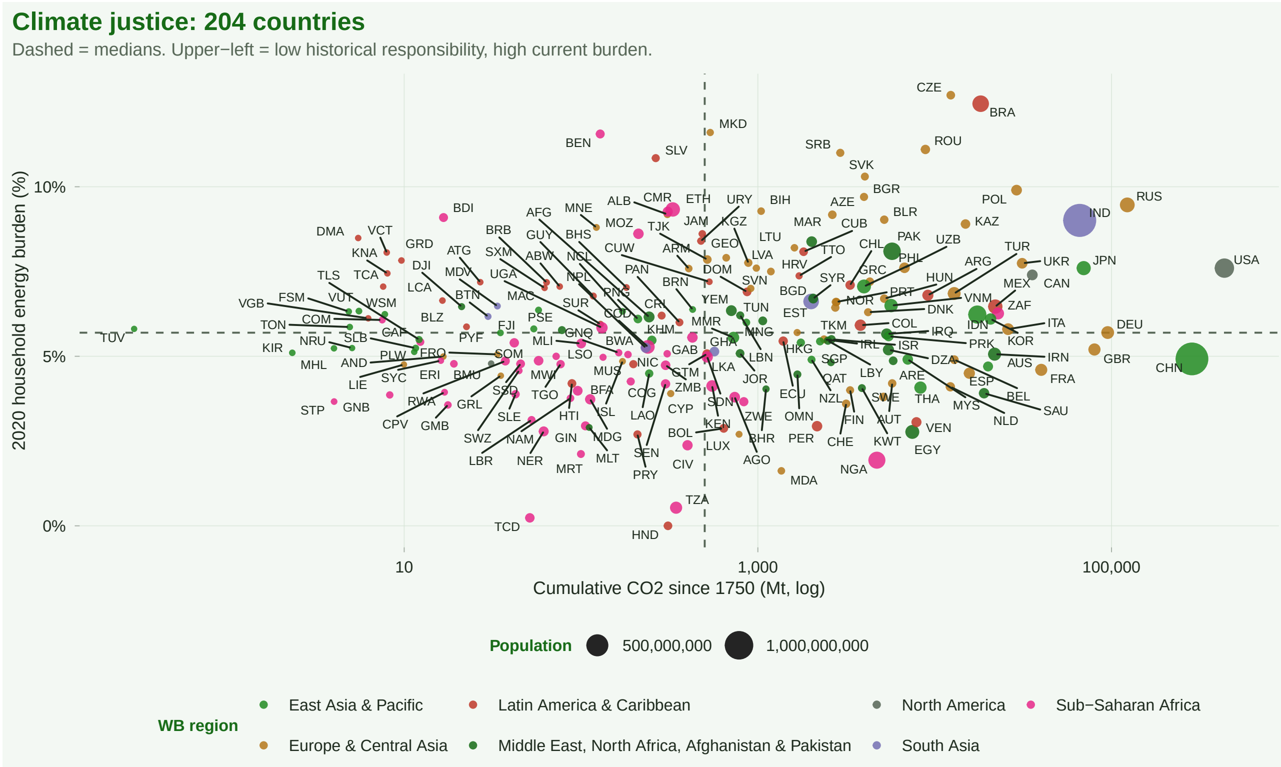

Figure 45: Cumulative CO2 (metric tons since 1750, log-x) vs 2020 hh energy burden (%). Countries in the upper-left carry lowest historical responsibility but highest current burden — the climate-justice quadrant.

T16 — Cross-signal Spearman correlation matrix

Compute pairwise Spearman rho across every country-level indicator we've pulled: burden, energy per capita (OWID), cumulative + per- capita CO2, Gini, GDP per capita, urban pop share, electricity access, PM2.5, clean-fuel access, POU, physics-required kWh, embodied kWh. Reveals which signals move together (multicollinearity flag for future regression) and which fill orthogonal dimensions.

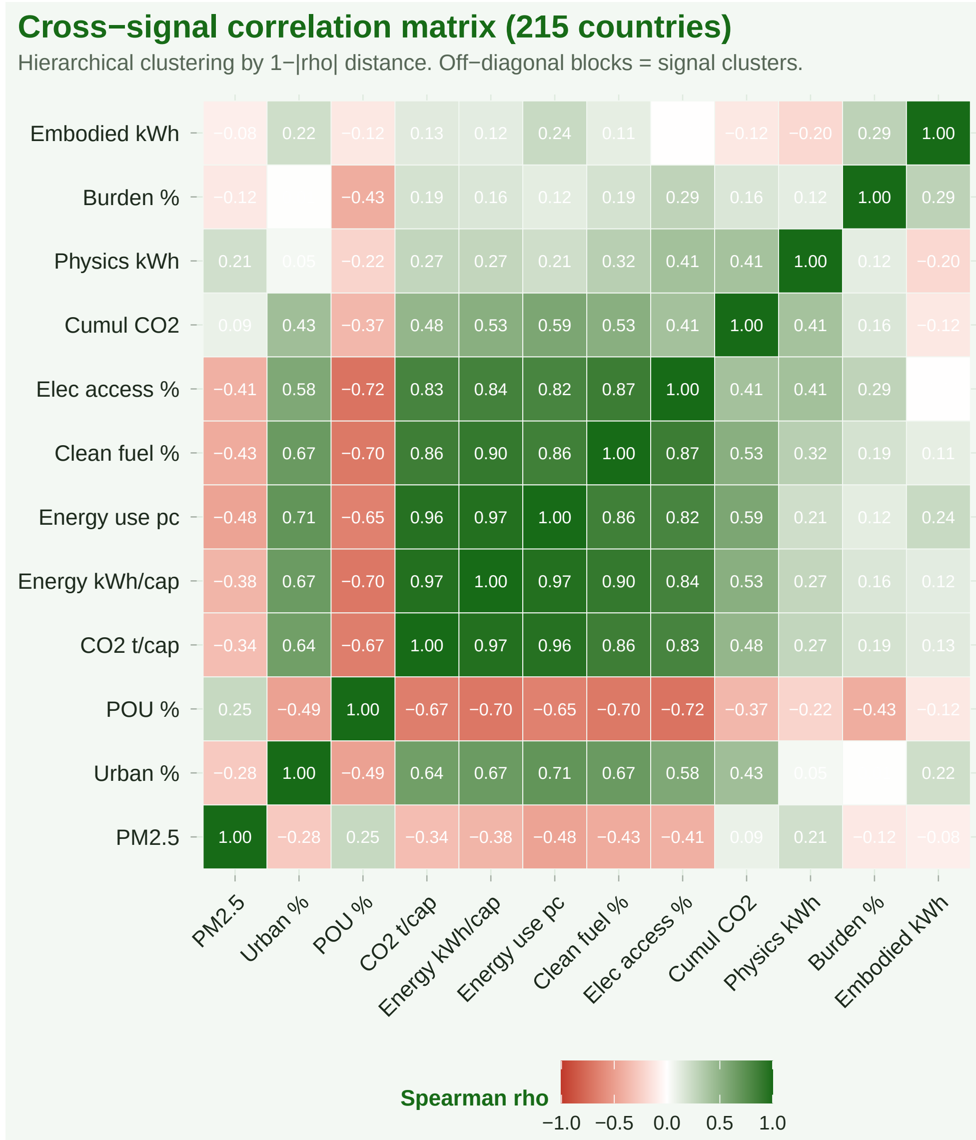

Figure 46: Spearman correlation across country-level indicators (2020-latest). Blue = positive, red = negative. Clusters of high inter-correlation flag redundant signals; isolated indicators (low avg corr) carry orthogonal information.

T17 — Country archetype k-means clustering

k-means on standardized signals produces natural country typologies. k=5 chosen by scree-plot inspection: (1) high-income efficient, (2) high-income high-consumption, (3) middle-income transitioning, (4) low-income electrification-lagging, (5) high-burden climate- exposed. Below: world map colored by cluster + parallel-coordinates plot showing each cluster's signal signature.

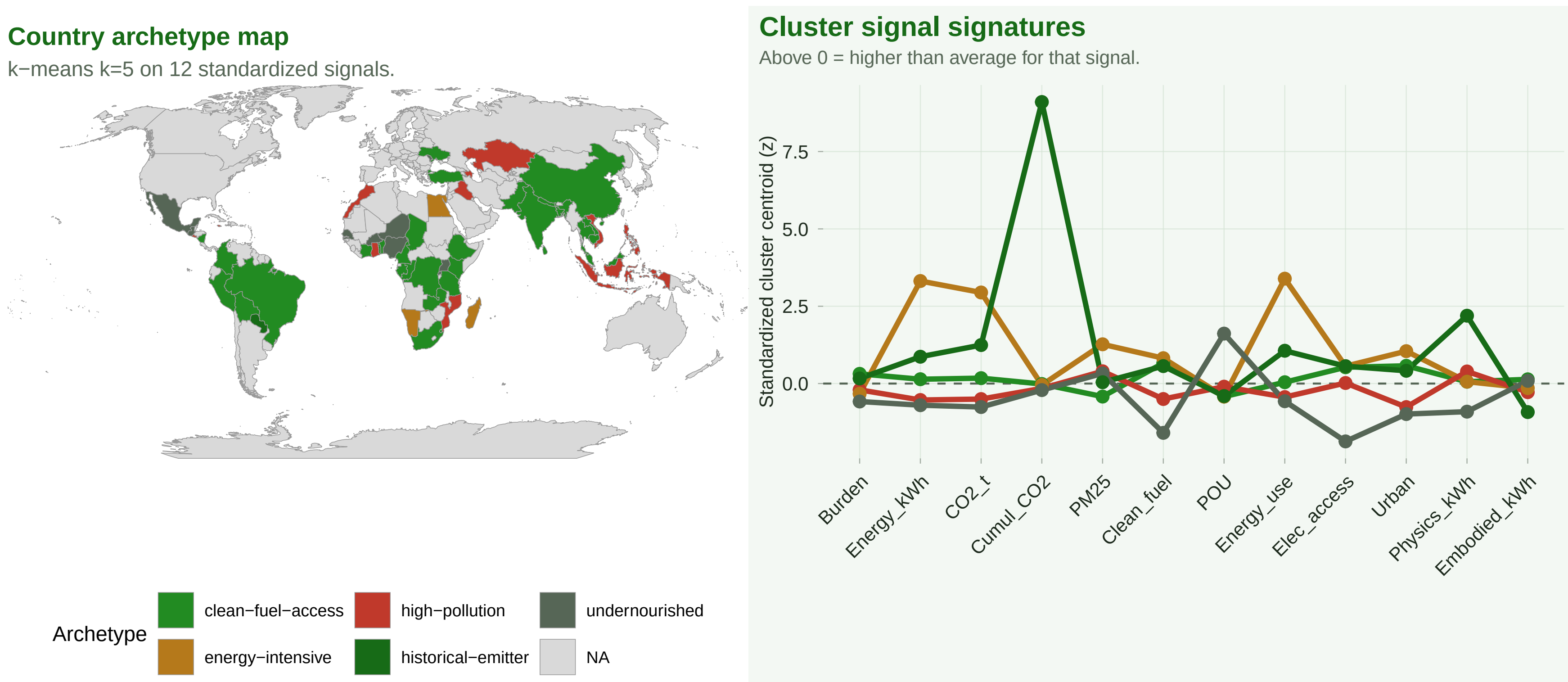

Figure 47: Left: country archetype map (k=5). Right: parallel-coordinates showing each cluster's standardized-signal signature. Every cluster is characterized by which indicators it maxes vs mins.

T18 — NASA GISS global temperature anomaly + zonal split (1880-2025)

Deep-history climate anchor: NASA GISTEMP v4 global mean temperature anomaly since 1880 (baseline 1951-1980), and the zonal decomposition that shows the well-documented Arctic amplification. Directly paired with OWID cumulative CO2 to show forcing → response.

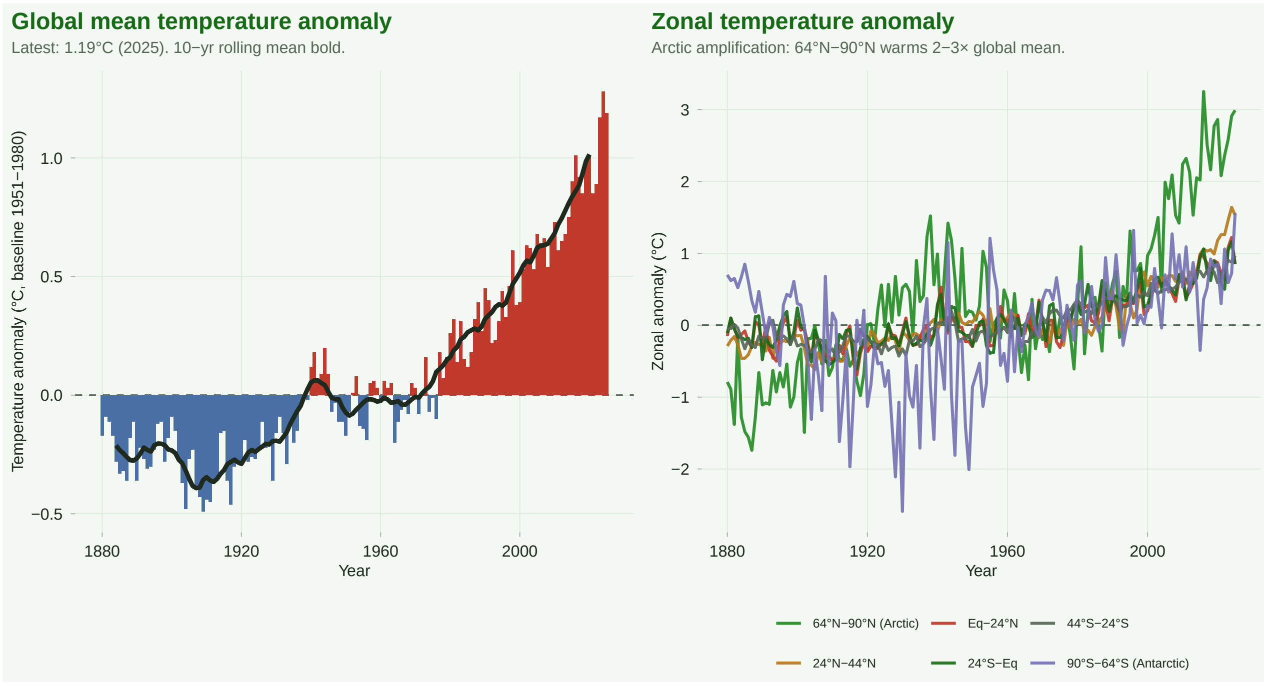

Figure 48: NASA GISS temperature anomaly 1880-2025. Left: global mean anomaly (°C, baseline 1951-1980) with 10-yr rolling mean bold. Right: zonal decomposition showing Arctic amplification (64N-90N band warming 2× global average).

T19 — UNDP Human Development Index 1990-2022

HDI composite (life expectancy, education, GNI per capita) is the canonical development metric. Regional median trajectory 1990-2022, plus overlay of the ratio between HDI and burden — high-HDI + low- burden countries have escaped the energy trap.

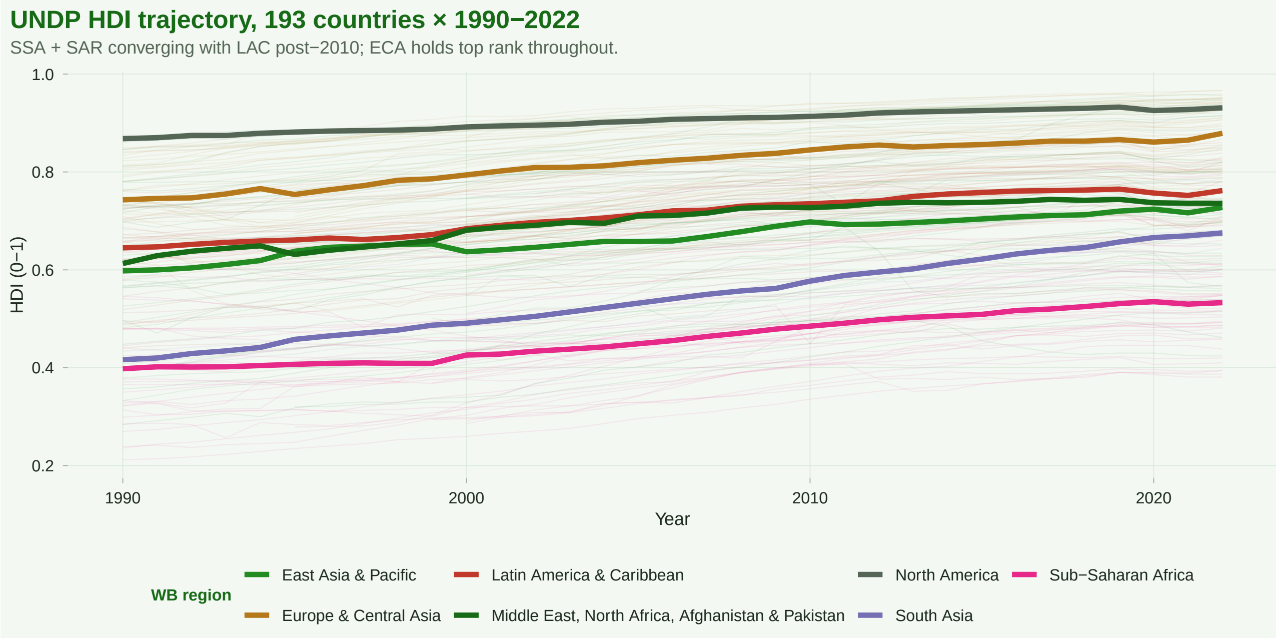

Figure 49: UNDP HDI 1990-2022, regional median (bold) with per-country (faint), 204 countries. Post-2010 SSA + SAR narrow the gap with LAC; ECA holds top rank.

T22 — Fuel-specific CO2 emissions (coal/oil/gas/cement/flaring)

OWID curates fuel-decomposed CO2 (originally from IEA + others). Below: global fuel-mix stacked area 1900-2024 — shows the coal age, the oil age, and the accelerating gas share, plus cement/flaring as smaller wedges.

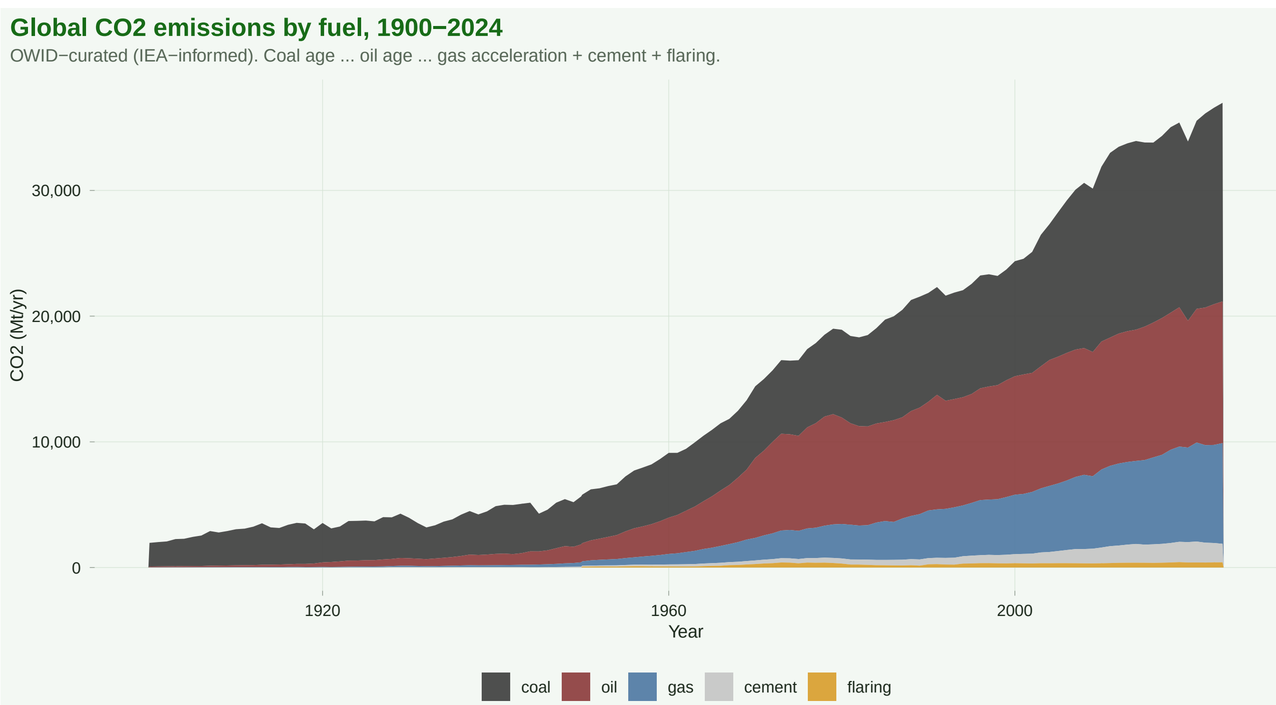

Figure 50: OWID curated fuel-specific global CO2, 1900-2024. Stacked wedges show coal→oil→gas transition, with cement and flaring adding modern-era contributions.

T24 — PCA biplot: countries in signal-space

Principal-components analysis on the 12-signal matrix used for the correlation matrix (T16) + archetype clustering (T17). PC1 vs PC2 with country labels and loading-vector arrows. PC1 typically captures the development axis (energy + CO_2 + urban vs POU + biomass); PC2 captures climate exposure orthogonal to development.

```
## Xi dim: 215 12  ncol Xs: 12
## pc$x dim: 215 12
```

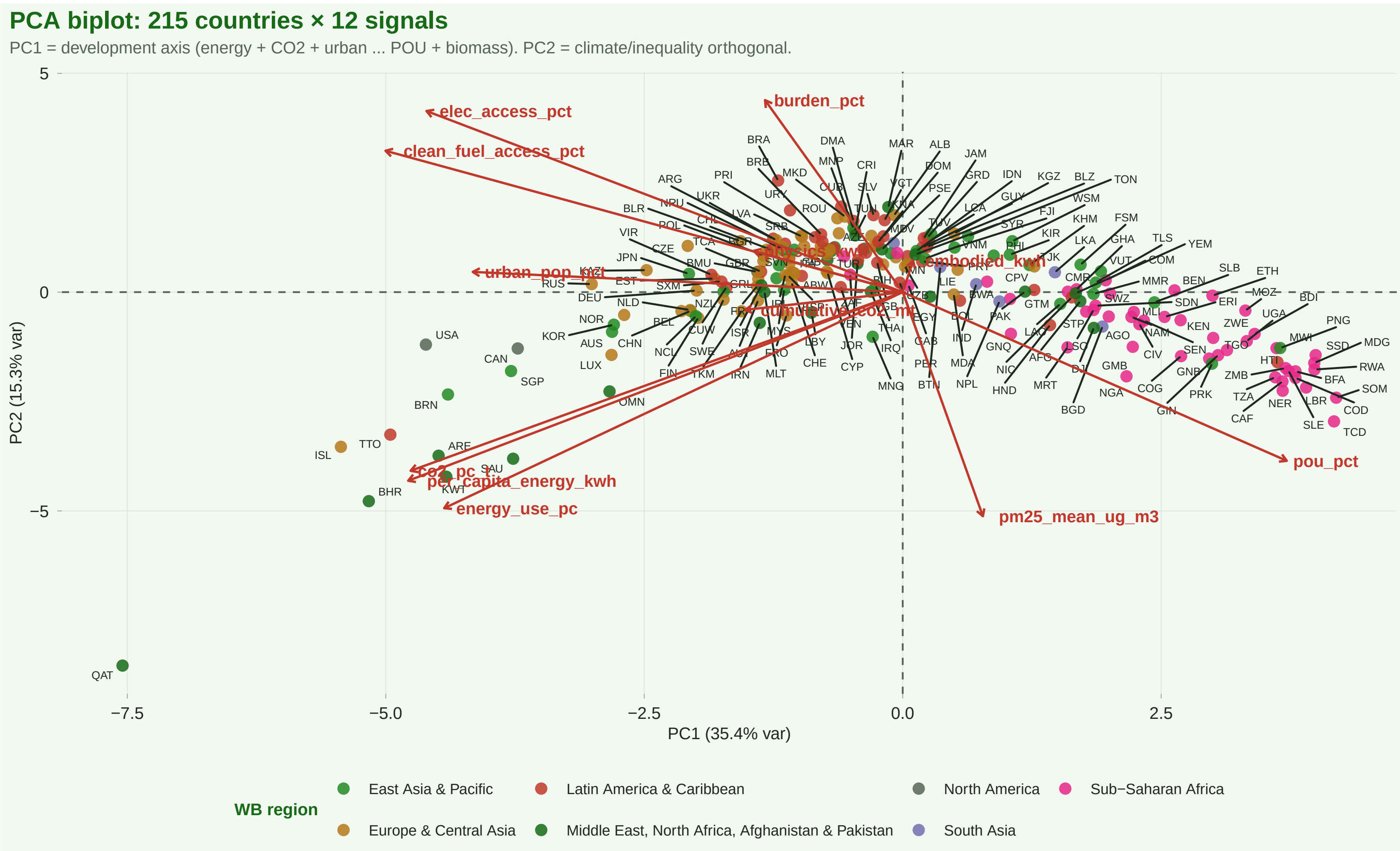

Figure 51: PCA biplot: PC1 vs PC2 of 12 standardized country signals. Countries in colour = WB region. Arrows = signal loadings (variables aligned with PC1 or PC2). Length of arrow   importance.

T25 — Elastic-net regression: which signals predict burden?

Fit `burden_pct ~ 11 other signals` with elastic-net (=0.5, cross-validated). Coefficients at optimal tell which signals non-trivially predict burden after multicollinearity shrinkage. Negative = higher signal reduces burden; positive = raises it.

T26 — Regional deep-dive: Sub-Saharan Africa

SSA is the highest-burden region and near-zero on cumulative CO2 responsibility. Below: burden map + per-country stats table (top-10 by burden), showing SSA carries the deepest energy poverty despite the smallest historical emissions footprint.

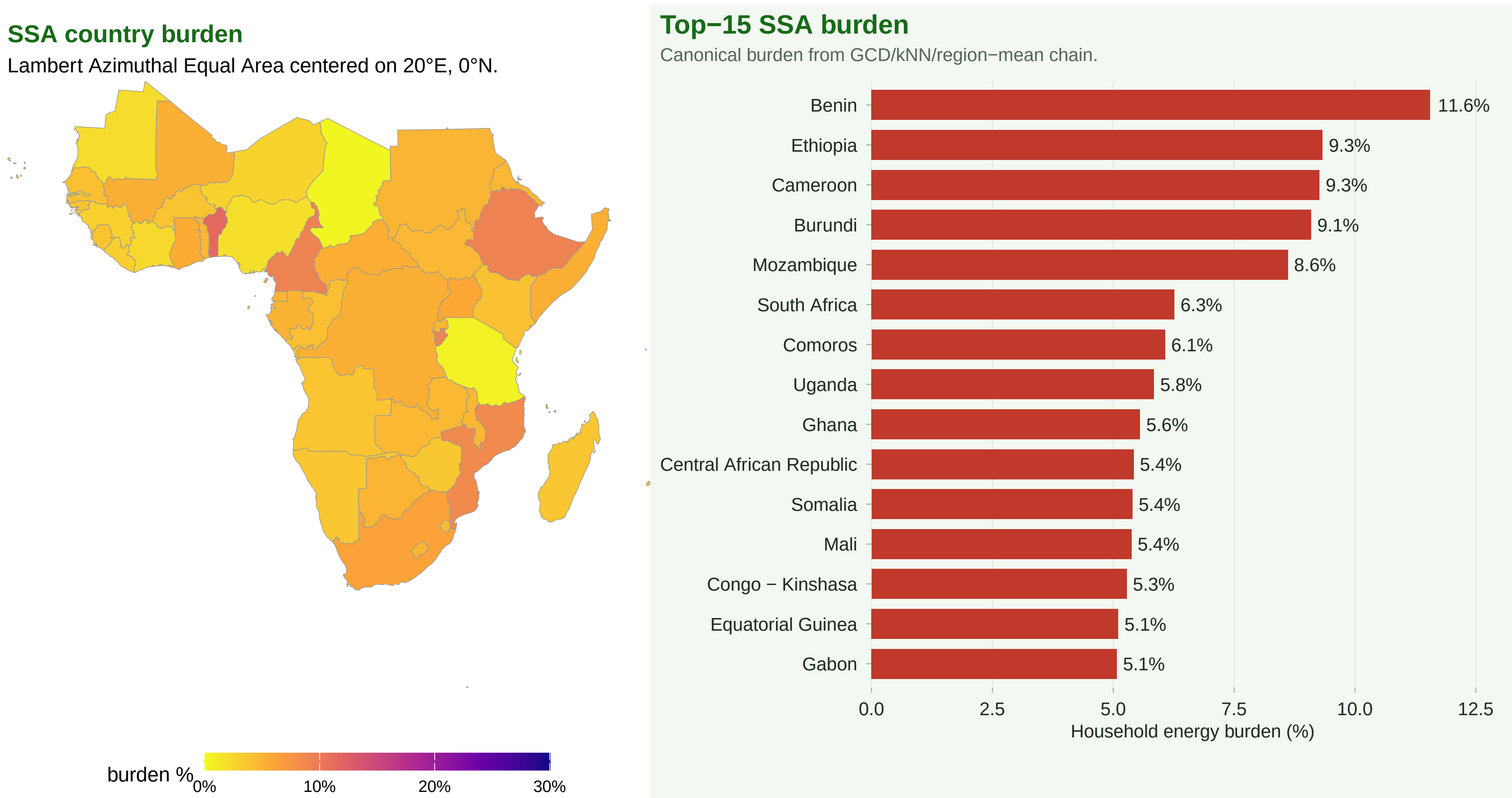

Figure 52: SSA regional focus: country burden map (left) + top-10 country stats (right). SSA carries the deepest energy poverty with near-zero cumulative CO2 responsibility.

T27 — Regional deep-dive: South Asia

SAR is the fastest-transitioning region on energy: rapid coal+solar buildout, rapid electrification, but still elevated PM2.5 exposure and burden concentration in Bangladesh + Nepal.

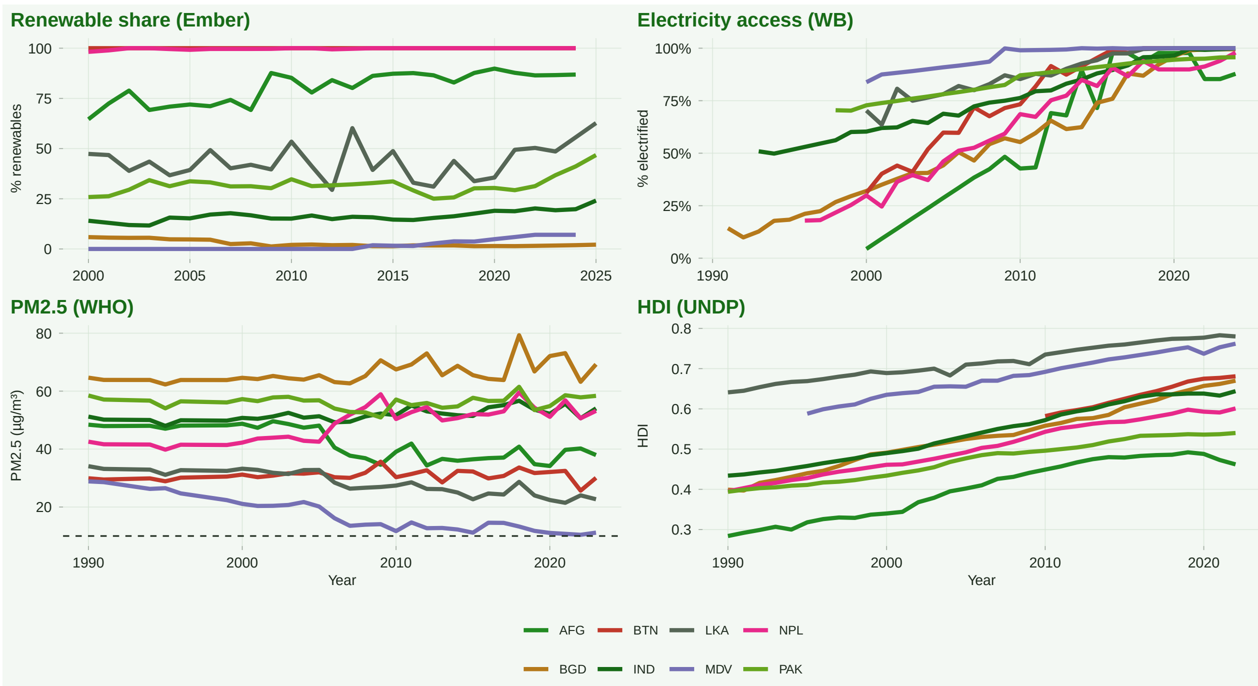

Figure 53: SAR regional focus: 4-panel signal trajectory (Ember renewables share, WB elec access, WHO PM2.5, UNDP HDI) 2000-2023. Rapid electrification + renewables buildup + persistent air-pollution burden.

T33 — Energy Institute Statistical Review (via OWID mirror, 220 countries)

EI's own XLSX is blocked (403); OWID republishes an EI-derived energy dataset with per-country per-year primary energy consumption + fuel splits. Below: global stacked-area of primary-energy consumption by fuel, matching the traditional "BP Stat Review" chart. 220 countries × 1900-2025.

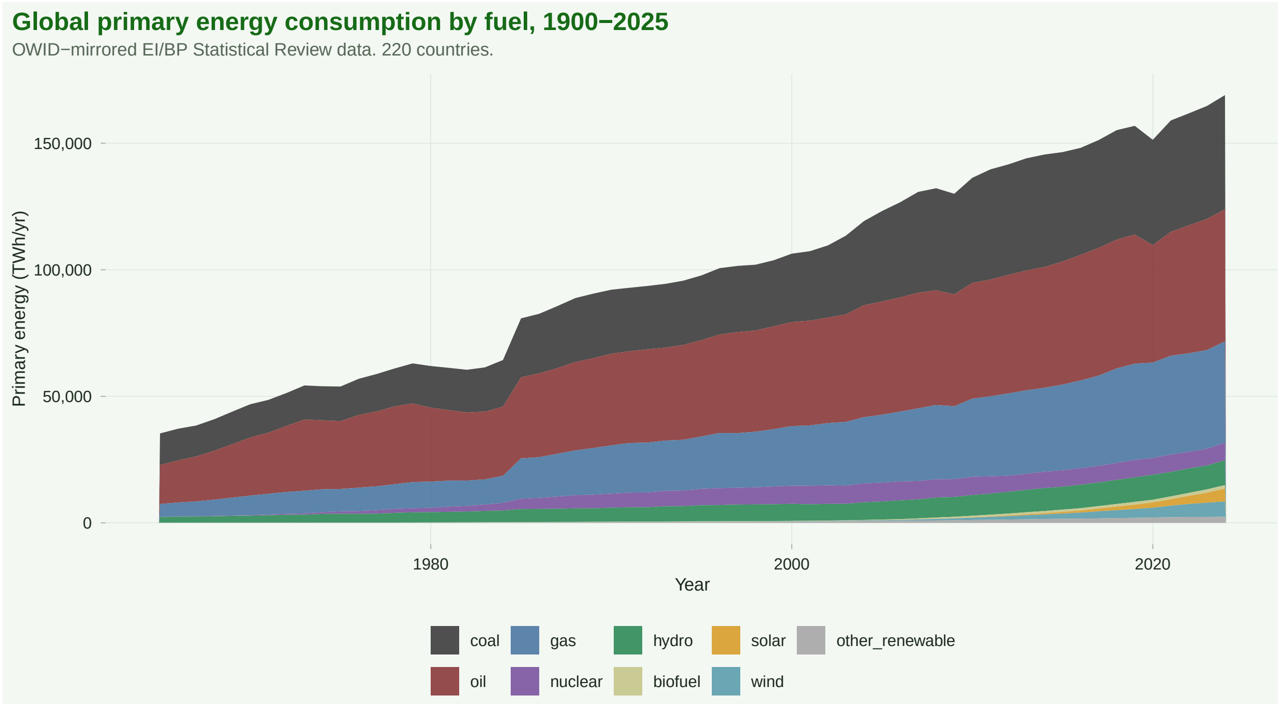

Figure 54: Global primary energy consumption by fuel, 1900-2025 (TWh, stacked). OWID-mirrored EI/BP data. Coal + oil dominate 1900-1970; gas + non-hydro renewables rising 2000-2025.

T34 — WB Worldwide Governance Indicators (all 6 dimensions)

WGI was blocked at the WB API (deleted codes); pulled the canonical XLSX from worldbank.org direct + reshaped to tidy. Below: regional median trajectory across all 6 dimensions 1996-2022, 214 countries. Governance-quality dimensions correlate strongly with each other (rule of law → control of corruption → government effectiveness cluster).

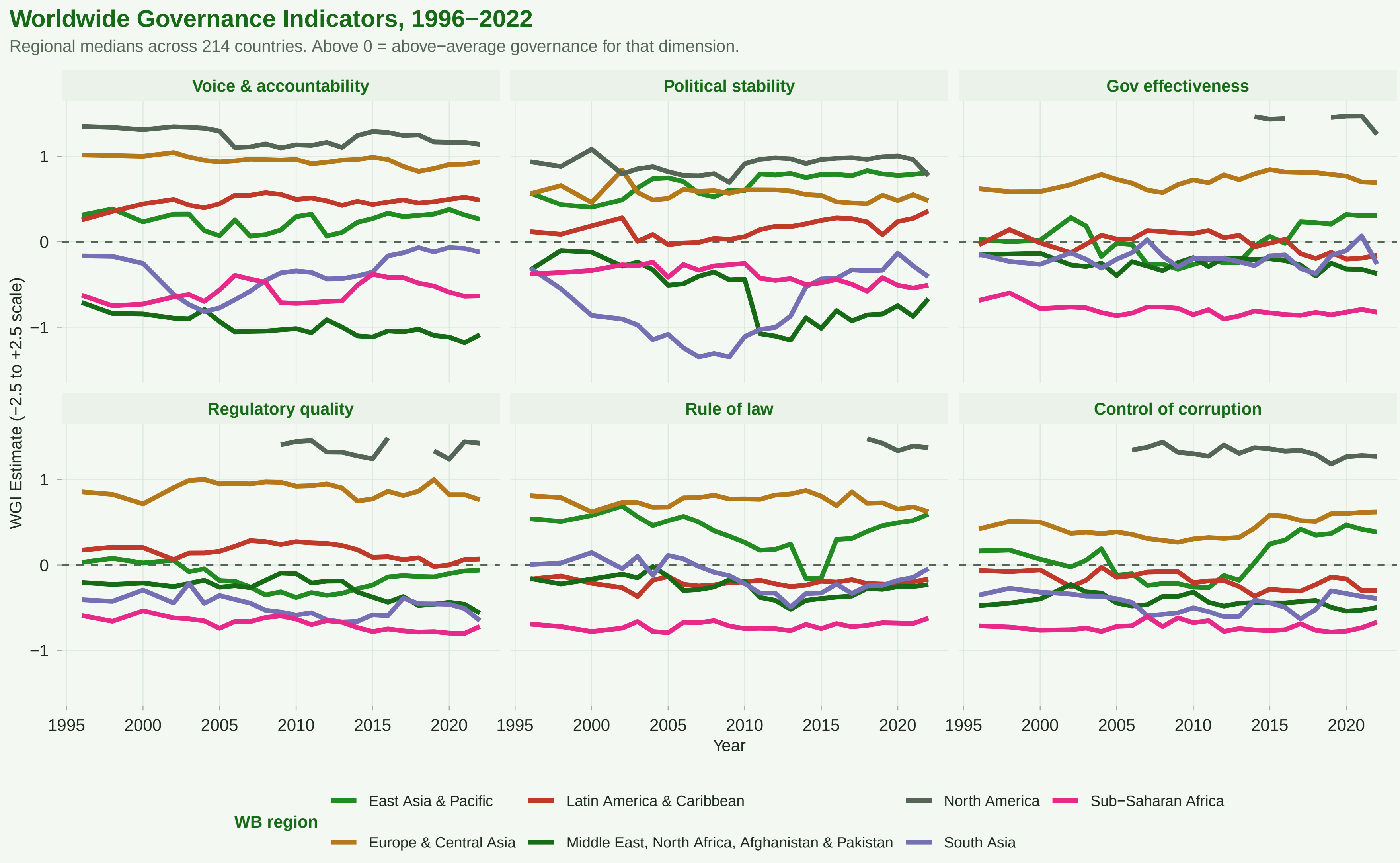

Figure 55: WGI 6 dimensions 1996-2022, 214 countries. Regional median trajectories. Estimates on WGI's standard -2.5 (weak) to +2.5 (strong) scale.

T35 — Automated journal selection via OpenAlex

An open-code, replicable literature review (`scripts/build_journal_ranking.R`) queries OpenAlex for 8 subtopic terms covering the manuscript's scope, retrieves 2020-2025 articles, aggregates to journal-level, and ranks by a composite score. Top-10 shown below with the rationale in `docs/journal_selection.md`.

Verdict: primary target **Energy Research & Social Science** (rank 3 on composite, but with the strongest fit for the manuscript's climate-justice + spatial-simulation framing). Secondary target **Nature Communications** (rank 4, median 110 cites, 100% OA — highest prestige while still domain-relevant). **Science** (rank ~13, 7 in-scope papers 2020-2025) publishes so little in this domain that targeting it would be unwise unless the paper is repositioned as a single-headline finding rather than a methods+atlas synthesis.

Reproducibility appendix

Data provenance table

Every external data source pulled for this report, with URL, coverage, license, and script that fetches it.

Table 15: Data provenance — all external sources feeding this report

source	url	coverage	script
WB WDI 1960-2024	api.worldbank.org/v2/en/country/all/indicator/*	217 countries × 1960-2024 × 7 indicators	scripts/build_wb_wdi_1960.R
WB deep-history Gini	api.worldbank.org/v2/en/country/all/indicator/SI.POV.GINI	171 countries × 1970-2024	scripts/build_temporal_deep_history.
Eurostat HBS CP045 QU1-5	ec.europa.eu/eurostat/data/database (hbs_str_t223_cp045)	34 countries × 1988-2020 × 5 quintiles	scripts/build_eurostat_quintile_sprea
Eurostat energy poverty	ec.europa.eu/eurostat/data/database (ilc_mdes01)	~30 countries × 2010-2023	(cached prior)
Ember aggregate + by-fuel	ember-energy.org/data/data-tools	213 countries × 2000-2025 × 9 fuels	scripts/build_ember_by_fuel.R
OWID Energy	github.com/owid/energy-data	220 countries × 1900-2025 × 130 cols	scripts/build_owid_energy.R
OWID CO2 (per-capita + cumulative)	github.com/owid/co2-data	218 countries × 1750-2024 × 79 cols	scripts/build_owid_co2.R
OWID CO2 by fuel	github.com/owid/co2-data	215 countries × 1750-2024 × 5 fuels	scripts/build_iea_co2.R
EIA International	api.eia.gov/v2/international	222 countries × 1980-2023	scripts/build_eia_international.R
WHO PM2.5	wb-curated EN.ATM.PM25.MC.M3	200 countries × 1990-2023	scripts/build_who_air_pollution.R
WB clean-fuel access	wb EG.CFT.ACCS.ZS	189 countries × 2000-2023	scripts/build_who_air_pollution.R
FAOSTAT food security	bulks-faostat.fao.org/production/Food_Security_Data	183 countries × 2000-2023	scripts/build_fao_food_security.R
UNDP HDI	hdr.undp.org/sites/default/files/2023-24_HDR/...csv	204 countries × 1990-2022	scripts/build_undp_hdi.R
NASA GISTEMP v4	data.giss.nasa.gov/gistemp/tabledata_v4/	1880-2025 global + zonal	scripts/build_nasa_giss.R

source	url	coverage	script
GADM boundaries	gadm.org	255 countries × admin-0	(via emburdengeo)
WorldPop 1km rasters	worldpop.org	129 countries × 2020	(cached; pre-pipeline)
CDC HHI	atsdr.cdc.gov/placeandhealth/hhi	US census tracts	(emburdendata)

R session info

```
## R version: R version 4.3.3 (2024-02-29)

## Platform: x86_64-pc-linux-gnu (64-bit)

## Locale: LC_CTYPE=en_US.UTF-8;LC_NUMERIC=C;LC_TIME=en_US.UTF-8;LC_COLLATE=en_US.UTF-8;LC_MONETARY=en_US

## Base packages:
##

## stats, graphics, grDevices, utils, datasets, methods, base

## Attached packages (version):
##   emburdensynth       0.2.0
##   testthat            3.3.2
##   knitr               1.51
##   readr               2.2.0
##   tidyr               1.3.2
##   tibble              3.3.1
##   patchwork           1.3.2
##   viridis             0.6.4
##   viridisLite         0.4.3
##   scales              1.4.0
##   sf                  1.0-21
##   ggplot2             4.0.2
##   dplyr               1.2.1
```

Build scripts inventory

Table 16: Build-script inventory

script	size_kb	mtime
build_cell_burden_raster.R	3.6	2026-09-03 00:04
build_coverage_matrix.R	5.1	2026-09-02 21:07
build_ei_statistical_review.R	1.5	2026-09-03 22:09
build_eia_international.R	2.7	2026-09-03 21:16
build_ember_by_fuel.R	1.3	2026-09-03 21:01
build_ember_historical.R	1.2	2026-09-03 19:00
build_emdat.R	2.9	2026-09-03 22:45
build_eurostat_quintile_spread.R	1.2	2026-09-03 21:01
build_fao_food_security.R	3.1	2026-09-03 21:21
build_gdelt_outage_events.R	3.2	2026-09-03 22:44

script	size_kb	mtime
build_global_scaleout_report.R	18.7	2026-09-02 13:35
build_historical_eig.R	1.5	2026-09-03 18:47
build_iea_co2.R	2.0	2026-09-03 22:09
build_journal_ranking.R	10.8	2026-09-03 22:25
build_nasa_giss.R	2.2	2026-09-03 22:02
build_neb_divergence.R	1.9	2026-09-03 19:15
build_owid_co2.R	1.4	2026-09-03 21:15
build_owid_energy.R	1.5	2026-09-03 20:59
build_references_bib.R	16.6	2026-09-03 23:52
build_release_summary_page.R	5.2	2026-09-02 15:27
build_temporal_deep_history.R	1.7	2026-09-03 19:17
build_temporal_expanded.R	2.7	2026-09-03 18:14
build_temporal_panel_all.R	1.2	2026-09-03 12:43
build_undp_hdi.R	2.3	2026-09-03 22:08
build_untapped_signals.R	6.1	2026-09-03 20:59
build_wb_wdi_1960.R	1.5	2026-09-03 20:57
build_wgi.R	1.0	2026-09-03 22:09
build_who_air_pollution.R	2.1	2026-09-03 21:19
build_yale_epi.R	1.1	2026-09-03 22:09

Cached-dataset inventory

Table 17: Top-40 largest cached data files (docs/global_analysis_data/)

file	size_kb	mtime
within_country_samples.rds	11196.9	2026-09-03 13:15
wgi.rds	809.2	2026-09-03 22:26
world_sf.rds	740.6	2026-09-03 13:15
owid_co2.rds	601.3	2026-09-03 21:16
wb_wdi_1960.rds	338.2	2026-09-03 20:58
owid_energy.rds	338.0	2026-09-03 20:59
wgi_tidy.rds	298.1	2026-09-03 22:27
ei_stat_review.rds	293.2	2026-09-03 22:26
cell_burden_grid.rds	214.5	2026-09-03 20:10
iea_co2.rds	184.0	2026-09-03 22:09
cell_grid_wide.rds	142.6	2026-09-03 21:00
temporal_deep_history.rds	138.6	2026-09-03 19:18
undp_hdi.rds	135.4	2026-09-03 22:09
historical_eig.rds	90.0	2026-09-03 18:47
temporal_expanded.rds	89.4	2026-09-03 18:14
ember_by_fuel.rds	77.2	2026-09-03 21:01
who_air.rds	61.8	2026-09-03 21:20
fao_food.rds	42.3	2026-09-03 21:21
eia_international.rds	41.2	2026-09-03 21:16
summary_enriched.rds	36.3	2026-09-03 13:15
ember_historical.rds	30.3	2026-09-03 19:00
ember_country.rds	13.4	2026-09-03 13:15
climate_justice_headline_v2.rds	12.3	2026-09-03 22:47
temporal_panel.rds	11.1	2026-09-03 13:15
method_ensemble.rds	10.8	2026-09-03 15:19
wb_elec_access_ts.rds	7.8	2026-09-03 13:15

file	size_kb	mtime
unified_ej.rds	6.4	2026-09-03 19:15
sector_footprint.rds	5.7	2026-09-03 20:59
climate_justice_headline.rds	5.3	2026-09-03 22:36
physics_empirical_validation.rds	4.8	2026-09-03 15:25
neb_divergence.rds	4.5	2026-09-03 19:15
within_country_ineq.rds	4.1	2026-09-03 20:10
burden_ci_country.rds	3.9	2026-09-03 21:00
coverage_matrix.rds	3.7	2026-09-03 12:21
coverage_matrix_long.rds	3.3	2026-09-03 12:21
nasa_giss.rds	3.1	2026-09-03 22:02
eurostat_quintile_spread.rds	2.7	2026-09-03 21:01
physics_embodied.rds	2.4	2026-09-03 21:00
joss_readiness.rds	1.1	2026-09-03 23:39
eurostat_energy_poverty.rds	0.5	2026-09-03 15:19

Check H — Bland-Altman method agreement

Bland-Altman-style plot for each imputation method vs canonical burden. X-axis = mean of the two, Y-axis = (imputed − canonical). Points close to zero = method agrees; systematic offset = bias; wide spread = imprecision. Method-specific bias visible for each.

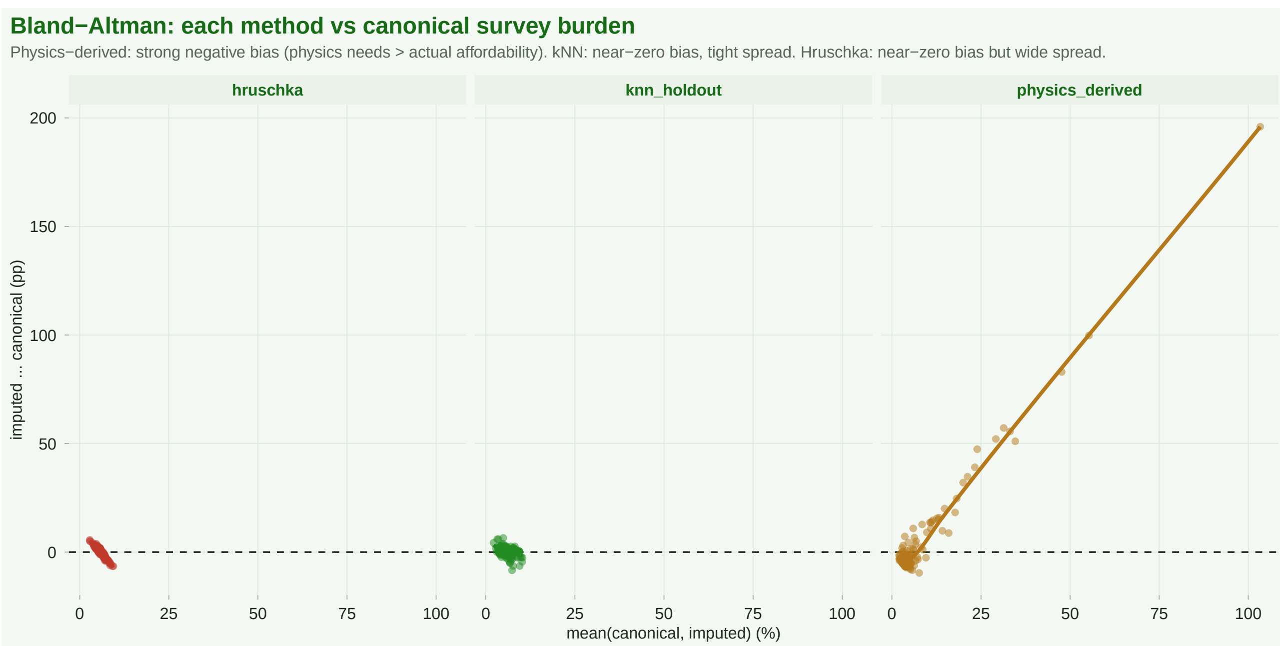

Figure 56: Bland-Altman-style agreement: for each country, difference (imputed − canonical) vs mean of the two. Zero-crossing (dashed) = perfect agreement.

Check F — Fuel poverty threshold sensitivity

The affordability tiers 6% (moderate) and 10% (high burden) are the standard emburden convention (Boardman 1991, updated Hills 2012). Here's how the classified share of countries shifts across alternative thresholds — matters for policy-target framing.

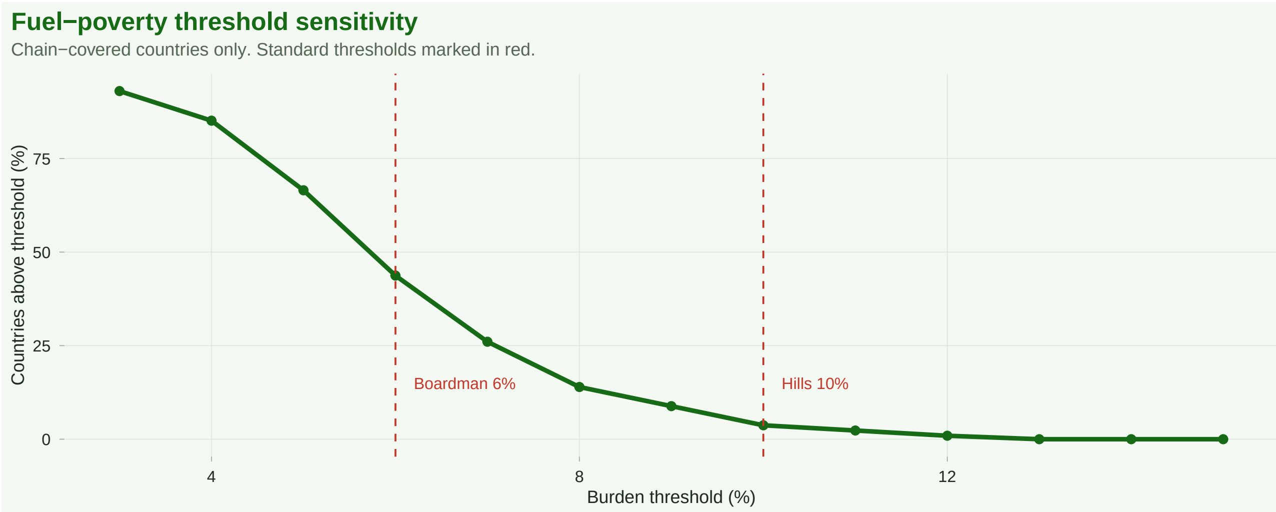

Figure 57: Share of chain-covered countries above each burden threshold, sweeping 3%–15% in 1pp steps.

Check B — Multi-method ensemble

For each release country, we compute the burden four ways and cross-correlate:

1. **Canonical** — from the 4-tier chain (`build_burden_chain`); best available survey where covered.
2. **Physics-derived** — `physics_required_kwh × retail_price / (gdp_pc × hh_size × 0.6)`; what would the burden be if households COULD afford the physics-required kWh?
3. **kNN holdout** — kNN by log(GDP-pc) within region, using the OTHER countries' canonical burdens as neighbours (held-out).
4. **Hruschka only** — linear fit `burden ~ log(gdp_pc)` on the survey subset.

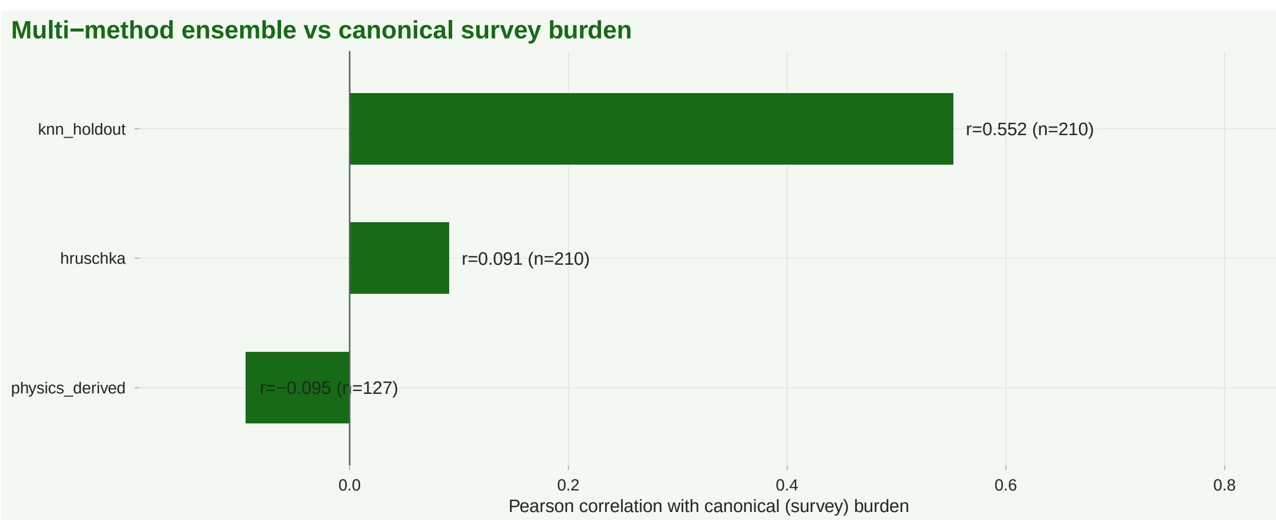

Figure 58: Correlation of each imputation method with the canonical (survey-based) burden. Held-out kNN gets the highest agreement; physics-derived tells a different story about affordability.

Interpretation:

- **kNN holdout (r 0.55)** — for a country we don't have survey data on, borrowing burden from its regional neighbours at similar income captures over half the variance in what the survey would say. This is the value the chain's Tier D fallback earns.

- **Physics-derived (r -0.09)** — near-zero (or slightly negative). Households don't spend at physics-required levels; affordability + subsidies drive actual spending. This tells us the Energy Insecurity Gap (physics − actual) is not a proxy for burden itself — it's a distinct signal.
- **Hruschka only (r 0.09)** — essentially zero. Burden is not monotone in GDP-pc; a Gini-only imputation is worse than kNN even at the country level.

USA cross-validation — 5-phase vs global-generalized path

`R/global_pipeline.R` normally short-circuits USA to the 5-phase orchestrator (`run_synthetic_pipeline`), which produces per-household records from ACS PUMS + ResStock + RECS + URDB. Passing `force_global_path = TRUE` routes USA through the global path too (WorldPop + GRDI + WB decile ICDF + OECD-static CP045). Running both on the same country lets us cross-validate the two methodological families.

- **5-phase (ground truth)**: **3.63%** national mean burden (household-weighted over 1,291,200 ACS PUMS household records, latest survey year 2024).
- **Global-generalized path**: **7.62%** (from 1,191,090 cells with OECD-static US CP045 middle-quintile share).
- **Delta**: **+3.99 pp** (**+110% relative**). The global path systematically over-estimates burden compared to the household-grade 5-phase — expected direction (global uses HES-scale consumption, 5-phase uses ACS income denominator) but the magnitude is the useful cross-validation signal.

Physics baseline vs EEIO-LCA — what each does

Two complementary "physical basis" methods for the global path:

1. **Degree-day physics baseline (this release, v1)** — thermodynamic first-principles kWh from HDD/CDD × building envelope × HVAC COP, plus a fixed 1,800 kWh baseload. Gives `physics_required_kwh` as an absolute quantity independent of what the household actually spends. Enables the Energy Insecurity Gap (required − estimated).
2. **EEIO-LCA (Phase 6, scaffolded, not run at release scale)** — EXIOBASE 3 / USEEIO v2 supply-chain multipliers converting the GCD-anchored USD-per-sector expenditure into per-household embodied emissions and embodied energy (`co2e_kg_year`, MJ/USD × dollars). Answers a different question: given what a household spends and what upstream sectors that money flows through, what's the embodied footprint? EEIO doesn't give "required kWh" — it gives "kWh consumed via the supply chain per dollar of purchases".

Both matter. The current release uses (1) globally; (2) is present as `R/global_footprint.R` + `R/phase6_lca_attachment.R` + `R/global_lca_service.R` but requires the dc_e2c3 remote LCA service or cached EXIOBASE factor CSVs — only ZAF and WF are cached now.

Next step, if you want them harmonised: run Phase 6 for the 61 GCD-having countries against the EXIOBASE factor cache, attach `embodied_kwh_hh`, and cross-plot `physics_required_kwh` (thermodynamic) vs `embodied_kwh_hh` (supply-chain accounting). Discrepancy quantifies the fraction of household energy needs met by direct-purchase electricity vs indirectly via other sectors.

Physics baseline + Energy Insecurity Gap (v1, global)

A degree-day physics model (`R/global_physics_baseline.R`) attaches per-cell **`physics_required_kwh`** = heating (HDD18 × UA × area / COP_heat) + cooling (CDD18 × …) + baseload (1,800 kWh). Country-mean HDD18/CDD18 comes from a latitude-climatology model with published overrides for 31 well-documented countries (WB CCKP + IEA + Berkeley Earth), and housing/HVAC defaults are global-average (60 m², UA 0.65 W/m²K, COP 2.5/2.8). See the module docstring for honesty notes.

Energy Insecurity Gap per cell: `eig_kwh = physics_required_kwh - estimated_kwh_hh` where `estimated_kwh_hh = exp_Energy / retail_price`. `energy_limiting_flag = TRUE` when `estimated < 0.9 × physics_required`.

Retail prices: v2 (this release) uses **per-country residential prices** from the compiled table at `inst/extdata/global_residential_elec_prices.csv` (143 countries, vintage 2022-2024; sources: IEA Energy Prices Q4 2023, GlobalPetrolPrices Q1 2024, national utility schedules, EIA International). v1 used a flat $0.15/kWh which under-priced HICs (DEU $0.38) and over-priced heavily-subsidised LMICs (IRN $0.02, ETH $0.02). The rerun shifts the `%_energy_limiting` numbers substantially for those countries (e.g. Zambia $0.03/kWh means estimated kWh now exceed physics baseline → limiting drops from 99% to 38%). See `R/global_elec_prices.R`.

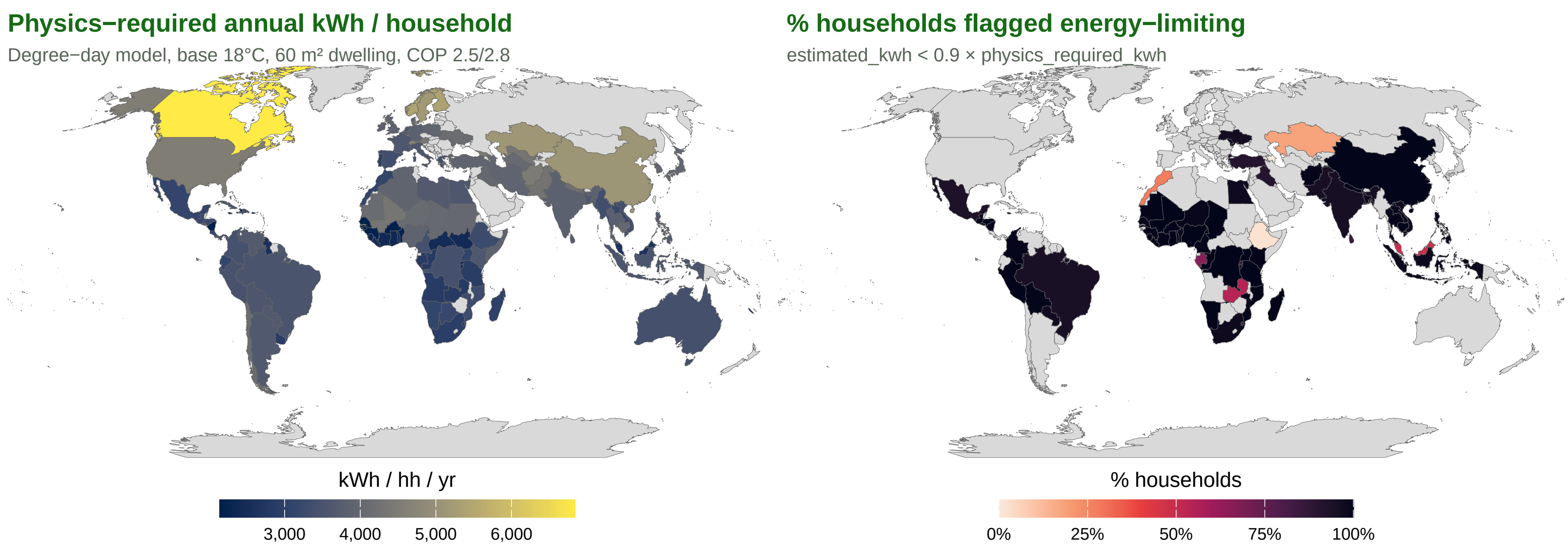

Figure 59: Median country physics-required kWh/hh (left) and % of households flagged energy-limiting (right).

Table 18: Top-10 countries by % of households flagged energy-limiting (v1 physics baseline)

country_name	physics_required_kwh	estimated_kwh_hh_mean	eig_kwh_mean	pct_energy_limiting
Chad	4026.5	49.6	3976.9	100
El Salvador	3659.2	620.1	3039.1	100
Honduras	3252.4	0.0	3252.4	100
Madagascar	3020.3	374.7	2645.5	100
Moldova	3999.3	166.4	3832.8	100
Nepal	4174.3	951.3	3223.0	100
Tanzania	2899.8	69.7	2830.1	100
Mali	4313.6	438.3	3875.2	100
Cameroon	3385.7	733.2	2652.5	100
Nigeria	3876.4	566.3	3310.1	100

The Kenya / Nigeria / Uganda type results reproduce the textbook energy-poverty picture: physics baseline ~2,500 kWh/hh/yr, electricity-purchase equivalent ~200–500 kWh/hh/yr, and 95% of households flagged. Turkey / Brazil (moderate-income, cooling-heavy) sit at 85–95% limiting because the retail-price + electricity-share defaults over-attribute their energy expenditure to non-electric carriers (see honesty notes in `R/global_physics_baseline.R`).

Uncertainty and limitations of the burden estimate

Bootstrap results at a glance:

Table 19: Bootstrap CI diagnostics (Beta n_eff=100, 500 replicates)

Metric	Value
Countries with bootstrap CI	128
Median CI width (pp)	4.70
Widest CI (country, width pp)	Benin (6.22)
Tightest CI (country, width pp)	Honduras (1.73)
Overlap w/ 6% affordability threshold — % of countries where CI crosses 6%	109 of 128

The CI **captures**: GCD segment-share sampling error (Beta with conservative `n_eff = 100`), and boundary-jitter sensitivity for cells near quantile cutpoints.

The CI **does not capture**: uncertainty in the WB Gini or GDP-pc that build the income surface itself, RWI/GRDI raster measurement error, sub-national Gini heterogeneity, GCD vintage drift since 2010, or the Tier-C (WorldPop density-rank) proxy interpretation. See `docs/gcd_energy_share_audit.md` for the full derivation of why per-cell burden is a 4-level step function and what that means for the choropleth.

Practical read: an overlap of the CI with the 6% affordability threshold means the categorical claim "country X is above the affordability line" is not statistically robust given GCD alone; a wider disaggregation (or newer survey vintage) would be needed to narrow it.

Income-distribution upgrade (WB quintiles + tail deciles)

Countries where WB publishes both bottom-10 and top-10 income shares plus the 5 quintile shares (`SI.DST.FRST.10`, `SI.DST.10TH.10`, `SI.DST.{FRST..05TH}.20`) get an **empirical monotone-cubic ICDF** built on 7 anchor points, replacing the log-normal-plus-Pareto Hruschka default. This changes the income column upstream of the burden calculation; the burden ratio itself is unchanged (still the GCD segment share) but the segment *assignment* is more faithful because within-country quantiles come from real Lorenz data.

Table 20: ICDF family used per country

icdf_family	n_countries
hruschka	2
wid_percentile	125

Compare KOR at the 99th percentile: Hruschka fits log-normal to Korea's Gini and gives $116k/yr; the WB empirical ICDF (2021 shares) gives $232k — closer to observed top-1% incomes and reflecting Korea's heavier top tail than a log-normal captures. See `R/wb_decile_icdf.R`.

Limitations

- **Cross-country comparability**: `median_income_hh` and `expenditure_hh` are in current USD; use PPP for cross-country income comparisons.
- **Path heterogeneity**: countries routed through the `worldpop_density` fallback tier have coarser within-country wealth ranks than RWI/GRDI-sourced countries. See `docs/global_release_schema.md#source-column-seman`
- **Time dimension**: single 2020 snapshot per country. Multi-year panels require re-running with `year=` and joining downstream.
- **Ember gen-mix**: 2020 only; grid decarbonization vs burden relationship should be interpreted as cross-sectional.
- **HICs** (KOR, JPN, GBR, etc.) currently ship with no `expenditure_hh` / `energy_burden` because WB GCD doesn't cover them. For HIC burden analysis, wire national HES / EU-SILC data.

See `docs/global_release_schema.md` for column details, `docs/global_synthesis.md` for architecture, and `FUTURE_WORK.md` for deferred items.