# CREA CO2 Tracker Methodology

19 January 2025 - v1.1

**Authors**: Hubert Thieriot, Lauri Myllyvirta

## Overview

The CREA CO2 Tracker is an online tool estimating CO2 emissions from fossil fuels within the EU27 region.

**Approach -** The tracker derives fossil fuel consumption from EUROSTAT's Supply, Transformation, and Consumption data[1] and applies IPCC emission factors[2] to estimate associated CO2 emissions.

**Missing data imputation** - Missing data, whether due to gaps or reporting latency, is filled using proxy datasets (EMBER, ENTSO-E, ENTSO-G, industrial production indices), heuristics and statistical projection methods (detailed below).

**Weather correction** - The tracker provides weather-corrected estimates that adjust for the influence of meteorological conditions (temperature, wind, solar irradiance) on energy demand and renewable generation, allowing users to isolate underlying emission trends.

**Scope -** CREA CO2 Tracker covers CO2 emissions resulting from the combustion of fossil fuels. It does not include industrial process emissions not stemming from fossil fuels, nor does it include agricultural emissions or emissions related to Land use, land-use change, and forestry (LULUCF).

CREA

# 1. Fossil fuel consumption

Fossil fuel consumption is derived from EUROSTAT's Supply, Transformation, and Consumption data[1], separated into three categories: solid, liquid and gas.

## 1.1 Solid Fossil Fuels

We included the following solid fossil fuels:

- Hard coal
- Brown coal
- Coke oven coke
- Brown coal briquettes
- Peat
- Oil shale

During the **coking process**, hard coal is transformed into coke oven coke, with coke oven gas generated as a byproduct. To avoid double counting while still accounting for all emissions, we analyzed the relationship between hard coal inputs and resulting energy content. We estimate that coke oven gas emissions downstream represent approximately 8% of the equivalent hard coal emissions. Therefore, we retain 8% of the coal input to coking plants in our accounting to represent these coke oven gas emissions, while subtracting the remaining 92% to prevent double counting with the coke oven coke that is tracked separately in consumption statistics.

## 1.2 Liquid fossil fuels

The tracker accounts for these key oil products:

- Oil products (aggregate): The overall category encompassing all petroleum products
- Fuel oil: Heavy residual oil used in industry and shipping
- Heating gasoil: Oil used primarily for heating in buildings and industry
- Road diesel: Diesel fuel used in road transportation
- Motor gasoline: Gasoline used primarily in road transportation
- Jet fuel (kerosene): Fuel used in aviation
- Aviation gasoline: Specialized gasoline for small aircraft

We also track **biofuels** (biogasoline and biodiesel) to subtract them from fossil fuel totals and assume no associated emissions for those.

## 1.3 Gas fossil fuels

The tracker accounts for the following gaseous fuels:

- Fossil gas
- Coke oven gas.

# 2. Missing-data imputation

The tracker employs data imputation and projection methods to provide up-to-date emissions estimates despite reporting delays, missing values, and varying detail across timeframes.

Our approach follows a hierarchical structure, applying imputation method in the following order:

## 2.1 Sectorial attribution in monthly data

Monthly energy statistics sometimes lack the sectoral detail available in annual data. We allocate consumption to sectors using assumptions validated through statistical analysis of historical data, which show strong correlations and minimal year-to-year variation. These assumptions apply only to recent monthly data pending the release of detailed yearly figures (typically delayed 1–2 years).

### 2.1.1 Oil products and transport

- **Motor gasoline and diesel allocation**: A constant proportion of a country's gasoline and diesel supply (after excluding petrochemical use) is used for road transportation (a distinct proportion per fuel and country is applied).
- **Aviation fuel distribution**: The split between domestic and international aviation for kerosene-type jet fuel is assumed to be constant year-on-year for each country. Similarly, aviation gasoline is almost exclusively used for domestic flights, with this pattern holding steady over time.
- **Transport fuel composition**: Motor gasoline, road diesel, and kerosene together account for the vast majority (typically >95%) of transport sector oil consumption. This consistency allows us to estimate total transport emission trends from these major components.
- **Non-energy use**: The proportion of oil products used for non-energy purposes (like chemical feedstocks) relative to total deliveries is constant for each country.

### 2.1.2 Fossil gas

- **Non-energy use of gas**: For petrochemical feedstock use, we apply the historical ratio between non-energy use and total gas consumption from annual data.

## 2.2 Filling data gaps

### 2.2.1 Zero-Value Imputation

When a country reports consecutive small values (below the 5th percentile for that fuel x country) followed by missing data, we impute zeros. This approach is particularly useful for countries phasing out certain fuels (e.g., Slovenia's hard coal use for electricity) which stop publishing figures.

### 2.2.2 Interpolation for Stable Consumption Patterns

For countries with stable consumption patterns (coefficient of variation < 10%), we use interpolation to fill small gaps (up to 3 months). This method is applied for instance with coking coal in Finland.

### 2.2.3 Manual Fixes for Reporting Issues

Country-specific fixes are applied for manually detected reporting anomalies (e.g., France's intermittent reporting of hard coal for electricity, Greece wrongly declaring that none of its brown coal is used for electricity). These fixes are based on cross-validation with other data sources (e.g., ENTSOE power generation data).

## 2.3 EU Aggregation for the latest months

When EU-level data is missing but sufficient country-level data exists, we fit a linear model (without intercept) to project the EU total from available countries. The model is used only if $R^2 > 0.95$. This addresses cases where a single missing country prevents publication of EU figures, even when that country represents a small share of consumption.

## 2.4 Proxy-Based Projections

For the most recent months with incomplete reporting, we use various proxy indicators and associated imputation mechanisms:

### 2.4.1 Power Generation from EMBER and ENTSO-E

Electricity sector emissions are projected using monthly power generation data from EMBER as the primary source. For the most recent months not yet covered by EMBER, daily ENTSO-E data is aggregated to monthly totals and scaled to align with EMBER. Where ENTSO-E data is also incomplete, the latest EMBER value is held constant.

This composite power generation time series serves as the sole predictor for corresponding country-source power-sector emissions. Imputation of emissions data is only performed when the power generation trend and the emissions trend demonstrate a strong linear relationship ($R^2 > 0.9$).

### 2.4.2 Gas Consumption from ENTSO-G

Daily gas consumption is estimated from cross-border flows and storage changes reported by ENTSO-G (storage data from ENTSO-G or AGSI). This "apparent consumption" is validated against EUROSTAT historical data and only used for countries where it correlates strongly ($R^2 > 0.95$). Where validated, the apparent consumption is scaled to match EUROSTAT totals and used as a predictor for gas-sector emissions.

### 2.4.3 Industrial Production Indices

Industrial coal and oil use is projected using monthly industrial production indices. Sector-specific indices (e.g., iron and steel, chemicals) are matched to relevant fuel types, capturing economic activity impacts on emissions beyond the power sector.

## 2.5 Time Series Forecasting with Uncertainty

For sectors without reliable proxies, we apply statistical time series models that incorporate seasonal patterns, trends, and recent observations. Confidence intervals reflect increasing uncertainty for longer projection periods. Models are recalibrated as new data becomes available.

All imputation methods have been validated against historical data (see Section 4).

# 3. Weather Correction

## 3.1 Purpose and Approach

Weather conditions affect CO2 emissions through two mechanisms:

1. temperature-driven changes in heating and cooling demand;
2. variability in renewable generation due to wind, solar irradiance, and precipitation.

To isolate underlying emission trends from meteorological effects, the tracker provides weather-corrected estimates. The core principle preserves residuals (non-weather factors) while removing only the estimated weather effect:

$$\text{corrected} = \text{actual} + (\text{fitted\_normal\_weather} - \text{fitted\_actual\_weather})$$

where fitted values come from statistical models trained on weather variables, and "normal weather" refers to climatological averages for each day of year.

## 3.2 Demand Correction

### 3.2.1 Electricity Demand

Electricity demand is corrected using a linear regression model:

$$\text{demand} \sim \text{HDD} + \text{CDD} + \text{weekday}$$

where:

- HDD (Heating Degree Days): captures increased demand during cold periods
- CDD (Cooling Degree Days): captures increased demand during hot periods
- weekday: controls for weekly consumption patterns

Daily correction factors are computed by comparing fitted demand at actual weather versus fitted demand at climatological average weather (mean HDD/CDD for each day-of-year over the historical period).

### 3.2.2 Gas Demand Decomposition (Non-Power Sector)

Non-power gas demand is decomposed into heating and non-heating components using a Generalized Additive Model (GAM):

$$demand \sim s(date) + s(date, by{=}HDD) + day\_of\_week + HDD{:}day\_of\_week$$

where:

- **s(date)**: smooth trend capturing the non-heating baseline over time
- **s(date, by=HDD)**: time-varying sensitivity to heating degree days, allowing the heating response to evolve (e.g., due to efficiency improvements or changes in building stock)
- **day_of_week**: weekly pattern in baseline demand
- **HDD:day_of_week**: interaction allowing heating response to vary by day of week

The model separates:

- **Heating component**: HDD-driven terms ( s(date, by=HDD) × HDD + HDD:day_of_week )
- **Non-heating component**: residual baseline ( s(date) + day_of_week )

This decomposition enables separate analysis of temperature-driven versus structural changes in gas consumption. Weather correction is applied to the heating component only.

## 3.3 Renewable Generation Correction

### 3.3.1 Wind

Wind generation is corrected using a Gradient Boosting Machine (GBM):

$$generation \sim year + ws + ws^2 + ws^3 \times inv\_temp$$

where:

- **ws**: wind speed at 50m (from NASA POWER)
- **inv_temp**: inverse temperature (1/K), capturing air density effects on turbine output

- **year**: accounts for capacity additions over time

Weather variables are collected at turbine locations identified in the Global Energy Monitor wind tracker. Because the model is non-linear, weather-corrected generation is computed by averaging predictions across all historical weather conditions for each day-of-year, rather than predicting at average weather.

### 3.3.2 Solar

Solar generation is corrected using a GBM:

*generation ~ year + solar_radiation*

where **solar_radiation** is surface solar irradiance from ERA5 (direct + diffuse). The same averaging approach as wind is applied to handle non-linearity.

### 3.3.3 Hydropower

Hydro generation is normalized using historical capacity factors rather than a weather model. Unlike wind and solar, hydro output depends on factors operating over longer timescales: seasonal precipitation, snowmelt timing, glacier melt, reservoir management, and competing water demands (irrigation, cooling). These factors make daily weather-based correction impractical.

Instead, we normalize to average capacity factor over the historical period (since 2015):

*corrected = actual × (average_CF / actual_CF)*

This approach removes year-to-year hydrological variability but does not isolate specific meteorological drivers.

## 3.4 Power Sector $CO_2$ Correction

Weather-corrected renewable generation is converted to a $CO_2$ correction factor by assuming 1 MWh of additional renewable generation displaces 1 MWh of thermal generation:

*thermal_corrected = thermal_actual − (renewable_corrected − renewable_actual)*

The correction factor is the ratio of weather-corrected thermal share to actual thermal share:

$$correction\_factor = thermal\_share\_corrected \,/\, thermal\_share\_actual$$

To avoid unreasonable results in countries with very low thermal generation, corrected thermal generation is capped at the maximum observed since 2015.

## 3.5 Combined CO2 Correction

Weather corrections are applied to CO2 emissions as follows:

| Sector | Fuel | Correction Applied |
|---|---|---|
| Power | All | Demand correction (electricity) × Power mix correction |
| Non-power | Gas | Demand correction (gas) |

As a first-order approximation, interactions between demand changes and electricity mix are not modeled

# 4. CO2 Emissions

Emission factors are taken from the Intergovernmental Panel on Climate Change (IPCC) Emission Factor Database.[2] Table 4 below indicates the emission factors considered, expressed in tonne CO2 emitted per Terajoule (expressed in Net Calorific Value). Net calorific values (NCVs) used to convert physical fuel quantities to energy content are taken from the IEA World Energy Statistics conversion factors.

| Fuel | Value ($t_{CO2}$ / TJ) |
|---|---|
| Hard coal | 92.8 |
| Brown coal | 113.1 |
| Brown coal briquettes | 99.0 |
| Peat | 117.8 |
| Oil shale | 108.0 |
| Oil products | 72.3 |
| Fuel oil | 77.7 |
| Heating gasoil | 73.33 |
| Motor gasoline | 72.1 |
| Road diesel | 72.1 |
| Natural gas | 55.7 |
| Coke oven coke | 113.0 |
| Coke oven gas | 41.2 |
| Kerosene | 72.7 |
| Aviation gasoline | 70.6 |

Table 1 - **CO2 emission factors considered from the IPCC**

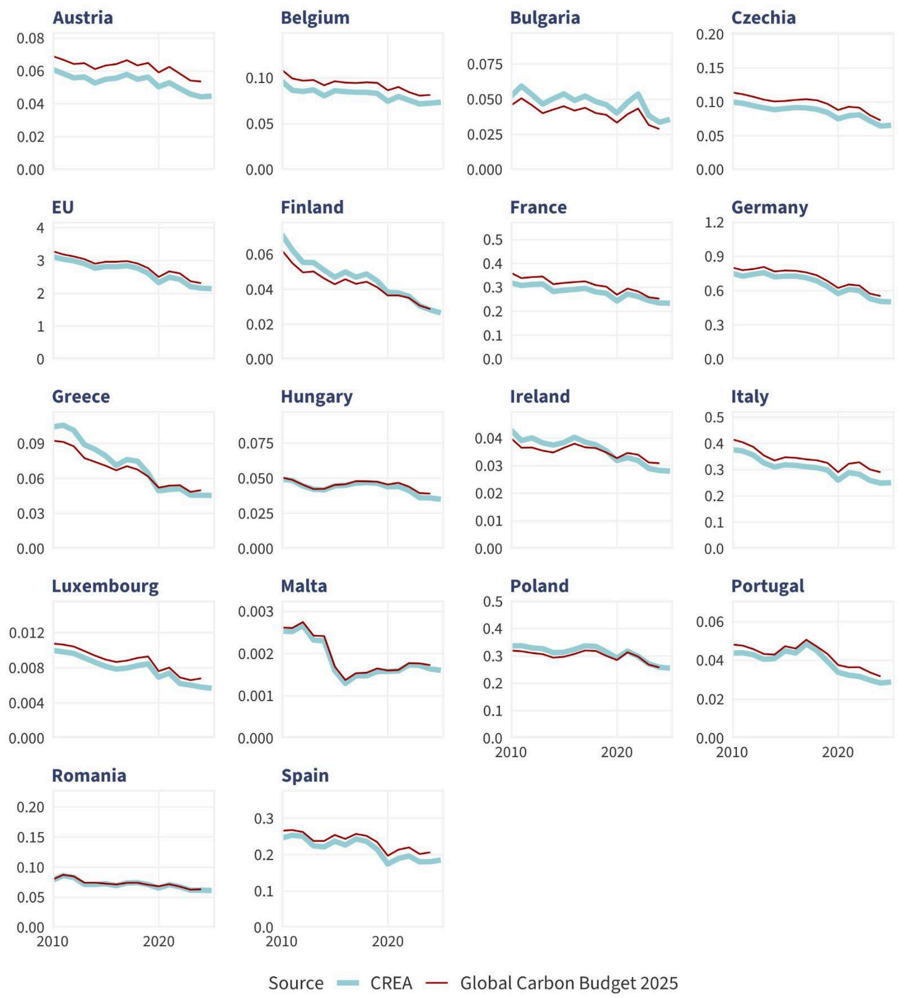
# 5. Validation

## 5.1 Comparison with Global Carbon Budget 2025

We validate national estimates values and trends against Global Carbon Budget 2025 datasets.[3] We only consider fossil emissions from coal, oil and gas and ignore emissions from the cement, flaring, and others categories.

Figure 1 - **CO2 emissions from fossil fuels - Comparison with Global Carbon Budget**

Billion tonne CO2 per year




**Source**: CREA analysis and Global Carbon Budget 2025 (Friedlingstein et al., 2025).

# Remaining limitations

## Monthly data limitations

The main application of the CO2 Tracker is to assess recent trends. To do so, we largely rely on Eurostat monthly data which comes with the previously mentioned limitations: monthly data offers fewer energy balance and fuel categories than yearly data.

To provide recent estimates, we have assumed that certain characteristics remains constant  since the latest available data point (see section 2. Missing-data imputation), namely:

- The share of motor gasoline, road diesel and aviation gasoline used by the transport sector is assumed constant in the last year.
- The non-energy use of liquid fossil fuels is directly proportional to the amount used by the petrochemical industry and the ratio remains constant.

These assumptions have been softly validated using historical data. Changes in industrial mix may therefore lead to overestimating or underestimating most recent trends in emissions.

# References

1. Statistics | Eurostat | Supply, transformation and consumption - commodity balances. https://ec.europa.eu/eurostat/databrowser/explore/all/envir?lang=en&subtheme=nrg.nrg_quant.nrg_quantm.nrg_cb_m&display=list&sort=category.

2. IPCC. IPCC Emission Factor Database. https://www.ipcc-nggip.iges.or.jp/EFDB/find_ef.php.

3. Friedlingstein, P., M. O'Sullivan, M. W. Jones, et al. "Global Carbon Budget 2025." Earth System Science Data Discussions 2025 (2025): 1–139. https://doi.org/10.5194/essd-2025-659.