

Reducing uncertainties in greenhouse gas emissions from chemical production

Received: 19 August 2023

Accepted: 24 February 2024

Published online: 26 March 2024

 Check for updatesLuke Cullen^{1,4}, Fanran Meng^{2,4}✉, Rick Lupton³ & Jonathan M. Cullen¹✉

Uncertainties in greenhouse gas emissions estimates for petrochemical production have lacked quantification globally, impacting emissions reporting and decarbonization policymaking. Here we analyze cradle-to-gate emissions of 81 chemicals at 37,000 facilities worldwide, assessing 6 uncertainty sources. The results estimate a 34% uncertainty in total global emissions of 1.9 ± 0.6 Gt of CO₂-equivalent emissions for 2020, and 15–40% uncertainties across most petrochemicals analyzed. The largest uncertainties stem from the inability to assign specific production processes to facilities owing to data limitations. Uncertain data on feedstock production and off-site energy generation contribute substantially, while on-site fuel combustion and chemical reactions have smaller roles. Allocation method choices for co-products are generally insignificant. Prioritizing facility-level process specification in data collection for just 20% of facilities could reduce global uncertainty by 80%. This underscores the necessity of quantifying uncertainty in petrochemical greenhouse gas emissions globally and outlines priorities for improved reporting. The dataset generated offers independent emissions factor estimates based on facility-specific information for 81 chemicals, supporting future analyses.

The petrochemical industry outputs nearly 1 billion tonnes of products annually¹, contributing to approximately 7% of global gross domestic product². Products include 420 ± 40 Mt of plastics and 190 ± 20 Mt of fertilizers^{1,3}. Petrochemical production is energy intensive, requiring 30% of final industrial energy use, including 14% of global oil demand and 9% of global natural gas demand⁴. Therefore, petrochemical production is a major cause of greenhouse gas (GHG) emissions, with the International Energy Agency (IEA) estimating annual GHGs emitted during petrochemical production, excluding external energy generation, at 1.30 Gt of CO₂-equivalent emissions (CO₂e) in 2020⁵. This is equivalent to approximately 14% of global industrial GHG emissions and 2.5% of all anthropogenic GHG emissions^{6,7}. Summary environmental assessments providing industry-wide figures describe the scale of action required to reduce GHG emissions in line with climate change mitigation goals. This sector substantially influences global GHG emissions and poses substantial challenges in decarbonization efforts. These have

motivated the industry to consider decarbonization as a priority^{8–15}, yet GHG emissions continue to rise year on year⁷. Given the complexity of decarbonizing petrochemical production, complete, accurate and detailed GHG emissions quantification is crucial to identifying opportunities for interventions, assessing their implications and setting credible emissions targets. However, the difficulty of monitoring the operations and emissions of the global petrochemical sector means this cannot be done precisely, and so quantifying the uncertainty of emission estimates is an essential component of a complete and transparent emission inventory¹⁶.

Existing environmental assessments use emission intensity factors (EFs) to convert material and energy flow data for individual materials or processes into GHG emissions estimates, which are used in compiling life-cycle analyses (LCA) and emissions-inventory databases¹⁷. EFs can be either obtained directly from commercial LCA databases^{18,19}, or calculated from reliable primary inventory data derived from rigorous

¹Department of Engineering, University of Cambridge, Cambridge, UK. ²Department of Chemical and Biological Engineering, Faculty of Engineering, University of Sheffield, Sheffield, UK. ³Centre for Sustainable Energy Systems, Institute for Sustainability, University of Bath, Bath, UK. ⁴These authors contributed equally: Luke Cullen, Fanran Meng. ✉ e-mail: f.meng@sheffield.ac.uk; jmc99@cam.ac.uk

process simulations²⁰ or proxy data²¹. These data present two main problems in the context of robustly quantifying global petrochemical sector emissions. Most obviously, uncertainties are often ignored or not recorded when reporting LCA results^{22–24}. The second problem is the diversity of manufacturing processes, feedstocks, plant locations or supply-chain routes for a product in the real world^{25,26}. Relying on generic or exemplar EFs may hide large differences in emissions between different manufacturing processes for the same product and can lead to large errors in emissions estimations.

Although previous studies have delved into the uncertainty of impact estimates resulting from practitioner choices in LCAs concerning global upstream crude oil refining^{27–29} or specific chemicals such as ethylene^{30,31} and ammonia^{32,33} at the country level, they have been limited in scope. These studies have not provided a comprehensive understanding of how to quantify and reduce uncertainty across the entire petrochemical supply chain. Effective policymaking, scenario mapping and assessment of decarbonization progress are dependent on reliable emissions estimates for the petrochemical sector. Therefore, there arises a critical need for a more extensive exploration that encompasses a comprehensive quantification of uncertainties and a deep-dive analysis of the diverse sources of uncertainty across all petrochemical production processes. Only then can the reliability of current estimates be judged, with suggestions on prioritization for future data collection to reduce emissions uncertainties leading to more accurate assessments.

Here we aim to quantify the extent of uncertainty at an aggregated (global) level in estimates of emissions from production of all widely-used petrochemicals, and what the most important sources of uncertainty are for each. In this Analysis, uncertainty reflects imperfect knowledge about the varied real operations of the world's petrochemical facilities, as well as ambiguity about how impacts should be calculated. To do this, we develop a process-based LCA model to produce all possible EFs and associated uncertainties for 81 chemicals, based on 2,043 types of chemical manufacturing process, informed by the IHS (now belongs to S&P Global) Process Economics Program (PEP) Yearbook³⁴. The model ensures mass and energy balances across processes. We then use data from the Independent Commodity Intelligence Services (ICIS) Supply and Demand Database¹ to assign possible EFs to 37,379 petrochemical plants worldwide.

We consider two sources of uncertainty arising from the most impactful modeling choices made by LCA practitioners. First, the choice of the allocation method used to divide the total EF calculated for a process between the co-products produced, referred to as 'allocation'. Second, several processes often exist for manufacturing the same product, and the precise process used at individual facilities is often not known, meaning that a unique match cannot be made between the facility and the relevant EF. This leads to a model uncertainty source referred to as 'process specificity'. Then, additional uncertainty arises owing to uncertainty in the data themselves. We split this uncertainty based on four types of emissions source, which taken together describe the cradle-to-gate EF for each process type: embedded in upstream feedstocks (referred to as 'feedstock'), off-site energy generation ('indirect energy use'), on-site fuel consumption ('direct energy use') and chemical reactions ('direct processes'). Although the 'allocation uncertainty' represents a different type of uncertainty from the other sources, resulting from practitioner choice rather than lack of knowledge or imprecise data, the 'correct' choice is often unclear and it valuable to know to what extent the choice is important, compared with other sources of uncertainty. We therefore include it here in the category of 'model uncertainty'.

First, this paper explores the two model uncertainty sources, allocation and process specificity, with case studies of ethylene and methanol production. Second, we incorporate the four data uncertainty sources to assess the overall impact of all six uncertainty sources on total emissions estimates, both at a process level for individual LCAs

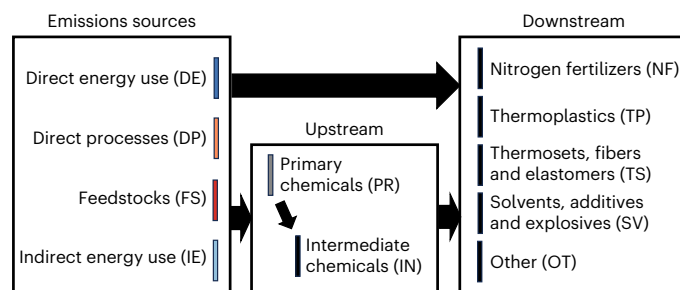


Fig. 1 | Schematic diagram of the four sources of emissions that apply directly to both upstream and downstream petrochemical production. Upstream products are used as input materials both for other types of upstream product and for downstream products as depicted by the black arrows. A collapsed version of this diagram is used as a key for other figures. See Supplementary Section 1.1 for product aggregation into groups.

and aggregated globally. Finally, we propose pathways to improve future emissions reporting through uncertainty reduction.

Results

To frame our discussion of uncertainty, Fig. 1 shows a schematic of emissions sources in the petrochemical industry and how emissions are embodied through upstream petrochemical production to final downstream products. We follow the naming conventions for chemical classification used by the IEA and refs. 35,36.

The first part of the results section addresses the two model uncertainty sources in estimating the EF of a production facility: the practitioner's choice of allocation method and the availability of data allowing for process specificity.

Uncertainties in EF estimation

Model uncertainty due to allocation. In the petrochemical industry, production processes often result in co-products. To measure the emissions due to any individual product, the total EF for a process must be split between the co-products. This can be done according to the output mass, economic or energy value of the co-products known as mass-, cost- and energy-based allocation, respectively. Although in the context of specific LCA studies there is often reason to choose a particular allocation method³⁷, in general, the appropriate choice is not always unambiguous or possible, and so the impact of different possible choices creates a source of modeling uncertainty³⁸. It is important to understand how important this choice is to the overall results, in aggregate and compared with other sources of uncertainty. Figure 2 shows the difference in EFs for a process resulting from different allocation methods.

Figure 2a shows that for most processes, the allocation method used makes little difference in the final EF calculated. Differences are seen between mass- and cost-based allocation for some processes for producing methanol, phenol and high-density polyethylene, shown by the black line. This is supported by Fig. 2b, with the case of methanol produced as a co-product of pure oxygen, where mass-based allocation yields an EF of 2.6 ± 0.5 kgCO₂e per kg, but cost-based allocation yields a higher EF of 5.3 ± 1.9 kgCO₂e per kg. The lower half of Fig. 2b shows that even the processes with the greatest difference between mass- and energy-based show little variation between these allocations. In general, the impact of uncertainty from allocation method choice is low, but in circumstances where co-products of significantly different economic values are produced, uncertainty stems from the difference between cost-based allocation on one side and mass- and energy-based allocation on the other. In practice, the chosen method for allocation between co-products is often based on available data^{37,39} and the bias this produces in emissions estimates should be considered in the allocation step.

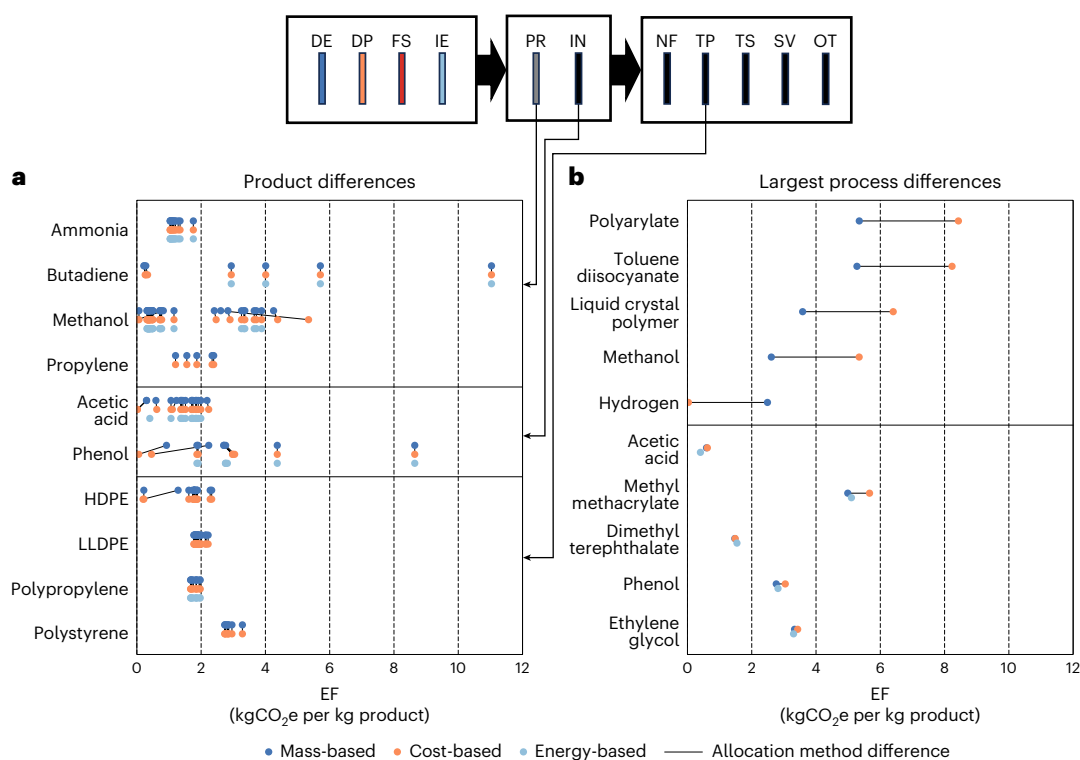


Fig. 2 | EFs from mass-, cost- and energy-based allocation. Each point represents the EF for one process using the allocation method specified by the point's color. Points representing EFs for the same process but with different allocation methods are linked by a black line to show the variation due to allocation methods. The range of possible EFs for a set of primary chemicals, intermediate chemicals and thermoplastics. **a**, The range of possible EFs for a set of primary chemicals, intermediate chemicals and thermoplastics. **b**, Processes, corresponding to the products on the y axis, with

the greatest difference due to allocation out of all processes considered. The top half shows those with the greatest difference between mass-based and cost-based, and the bottom half shows those with the greatest difference between mass-based and energy-based. HDPE: high density polyethylene, LLDPE: linear low density polyethylene. See Supplementary Section 2.1 for comparisons for other product groups.

Model uncertainty due to lack of process specificity. A range of processes, and therefore possible EFs, exist for the production of each product in the petrochemical industry. To accurately estimate emissions, each facility should be allocated the appropriate EF for the process being used. In practice, knowledge of the specific process being used at a facility is limited owing to data availability and the data can often be subject to industrial secrecy. Consequently, a range of EFs can be possible for a facility, resulting in 'process specificity' uncertainty. Figure 3 illustrates the range of possible EFs found using our model for primary chemicals and plastics, and hence the magnitude of possible process specificity uncertainties if the only information known about a facility is the product being produced.

Figure 3 shows that each product is subject to a wide range of possible EFs dependent on the production process used. The range of 0.2–11 kgCO₂e per kg for butadiene EFs is large compared with the standard global values of 1.20 kgCO₂e per kg and 1.56 kgCO₂e per kg offered by the LCA databases^{18,19}. It should be noted that not all possible process methods are commonly used, for instance, butadiene production through the *N*-methylpyrrolidone process, with an EF of 1.5 ± 0.3 kgCO₂e per kg through mass allocation, is far more widespread than the bio-based version indirectly produced via 1,3-butanediol with an EF of 11 ± 2 kgCO₂e per kg through mass allocation. An unweighted mean of possible EFs may therefore not reflect the true mean for global production. However, the wide range of possible EFs leaves room for considerable uncertainty in using generic LCA database values for the emissions estimation of a particular product or facility. Overall, Fig. 3 shows that primary chemicals are subject to a larger variety of processing methods than downstream thermoplastic production.

To understand the reasons for differences in EFs from different process methods in primary chemical production, we can look more closely at the contribution of each emissions source. Ethylene and methanol are two major chemicals with considerable variability and are taken as an example shown in Fig. 4. See Supplementary Section 2.3 for a breakdown of other primary chemicals and plastics and a summary of minimum and maximum EFs calculated for all products considered in this study.

Processes that mainly use coal for ethylene production have cradle-to-gate EFs ranging from 6.0 ± 1.1 to 7.3 ± 1.3 kgCO₂e per kg, which is substantially higher than those of processes mainly using naphtha, which range from 0.6 ± 0.1 to 1.3 ± 0.2 kgCO₂e per kg. Emissions from feedstocks can vary between processes with the same principal feedstock due to the quantities used in each process recipe, but overall processing technologies that share their primary feedstocks tend to have similar EFs. This shows the importance of knowing the embedded feedstock emissions when determining the final EF and presents an opportunity for limiting process specificity uncertainty through knowledge of a facility's primary feedstock, which will be quantified below. Exceptions occur in some cases such as electric arc processing via acetylene, with feedstock emissions of 0.47 ± 0.07 kgCO₂e per kg, similar to other acetylene-based processes, but an indirect energy use EF of 5.0 ± 0.8 kgCO₂e per kg renders this an emission intensive process with a total EF of 5.5 ± 0.9 kgCO₂e per kg, much higher than other acetylene-based processes in the 0.3–0.5 kgCO₂e per kg range. Contrasting EFs between feedstock types supports a transition of ethylene production away from coal and methanol towards ethane and naphtha. However, the case of acetylene-based processes shows that a thorough LCA of the chosen process must be conducted to avoid

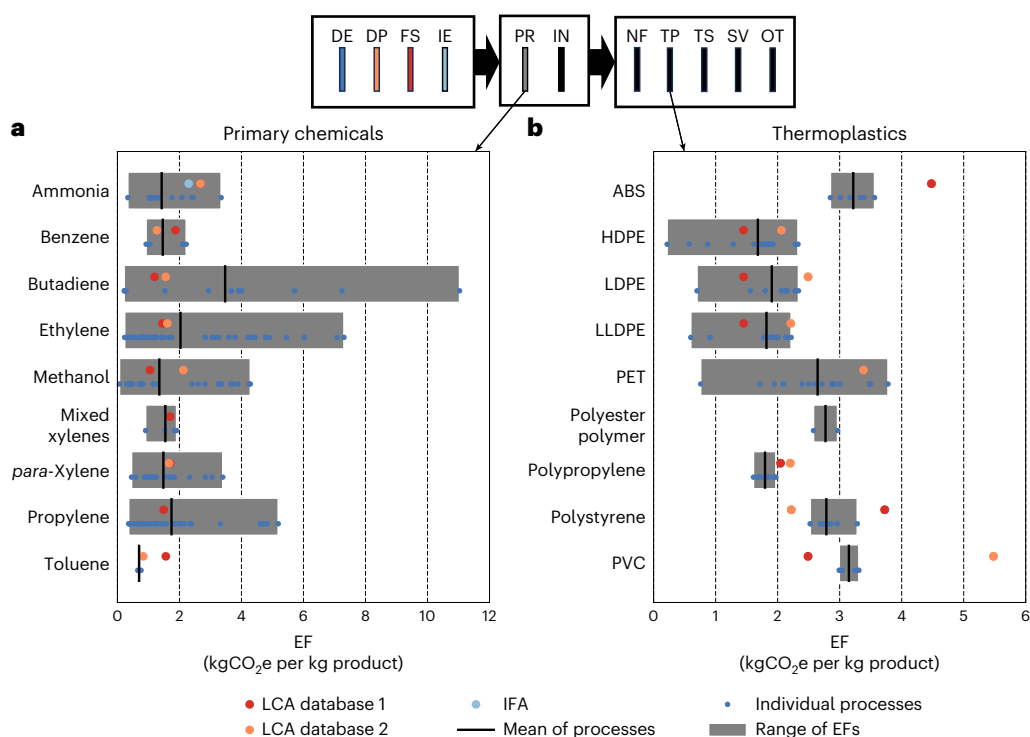


Fig. 3 | Cradle-to-gate EFs for selected chemicals and products using mass-based allocation. a, Primary chemicals. b, Thermoplastics. The gray bar indicates the range of possible EFs for each product due to different process types. Individual EFs are shown along the bottom of each bar with the mean value shown by the vertical black bar. In **b**, to isolate the impact of process specificity in downstream processes, uncertainty from process specificity in upstream chemicals is not propagated to downstream EFs. Comparisons with the ecoinvent

and CarbonMinds LCA databases^{18,19} and IFA EFs are shown where available. The key at the top of the figure refers to Fig. 1. ABS: acrylonitrile butadiene styrene, LDPE: low density polyethylene, PET: polyethylene terephthalate, PVC: polyvinyl chloride. See Supplementary Section 2.2 for other product groups and Supplementary Table 4 in Supplementary Section 2.3 for the EF ranges for all chemical products.

undermining the gains from reducing embedded feedstock emissions with losses from an increase in emissions from other sources including electricity use.

Lack of process specificity compared with data uncertainty. Figure 4 shows that when specific processes are not known, the process specificity uncertainty is larger than the sum of the four data uncertainty sources seen in the error bars for each process. When the specific process for a facility is defined, process specificity uncertainty is eliminated and only uncertainty from the four data sources and allocation methods remain. Extended Data Table 1 breaks down the average contribution of emissions from each source, the average uncertainty associated with the data source and the implied contribution of each data uncertainty source to the sum of data uncertainty for a process. On average, data uncertainties from feedstocks (56%) and indirect energy use (40%) are more significant than those from direct energy use (2.0%) and direct processes (1.5%).

This section has shown that process specificity, feedstocks and indirect energy use are the largest sources of uncertainty in calculating EFs in the petrochemical industry. Direct energy use, direct process and allocation uncertainties are less significant overall but can be important in some specific cases. The next section quantifies the average impact of EF uncertainties on process LCAs across different products and aggregated global estimates of GHG emissions from petrochemical production.

Aggregated impact of EF uncertainties

Uncertainties in EFs impact process emissions estimates for each petrochemical facility, as discussed in the previous section. To understand the average impact across different products in current LCAs, we propagate

uncertainties through to global emissions estimates. To quantify uncertainties at a global scale, we use facility-level data from the ICIS Supply and Demand Database¹ and associate every petrochemical manufacturing plant with EFs for possible specific processes used. Given the small impact of allocation uncertainty, we use mass-based allocation for all EFs. Where a facility may be employing one of multiple possible processes, process specificity uncertainty is assigned to each emissions source as discussed in Methods. Combining the assigned EFs for each facility with estimated facility-level production data derived from the United Nation Food and Agriculture Organization⁴⁰, the International Fertilizer Association (IFA)³ and the ICIS¹ databases, we estimate total global petrochemical cradle-to-gate emissions for 2020 as 1.9 ± 0.6 GtCO₂e. Emission uncertainties from upstream production are propagated to downstream products when primary or intermediate chemicals are used as inputs for the downstream production process. Figure 5 shows the average EF uncertainty for a process in each product group. The impact on global emissions uncertainty is shown at the top of each plot by multiplying the average EF uncertainty of product groups with their production mass.

At a process level, the highest average EF uncertainty amongst the primary chemicals seen in Fig. 5c is for butadiene at 2.45 kgCO₂e per kg, reflecting the wide range seen in Fig. 3. Together with the ‘thermosets’ downstream group, this suggests that fewer produced products are often associated with the highest level of average EF uncertainty. The distribution of inputs, and therefore data uncertainty, for low-production products is similar to high-production products, hence the increased uncertainty originates from the process specificity source where facility processes cannot be specified beyond a range of processes with large differences in EFs. Less produced products are therefore more liable to large uncertainties in process LCAs. The average uncertainties seen

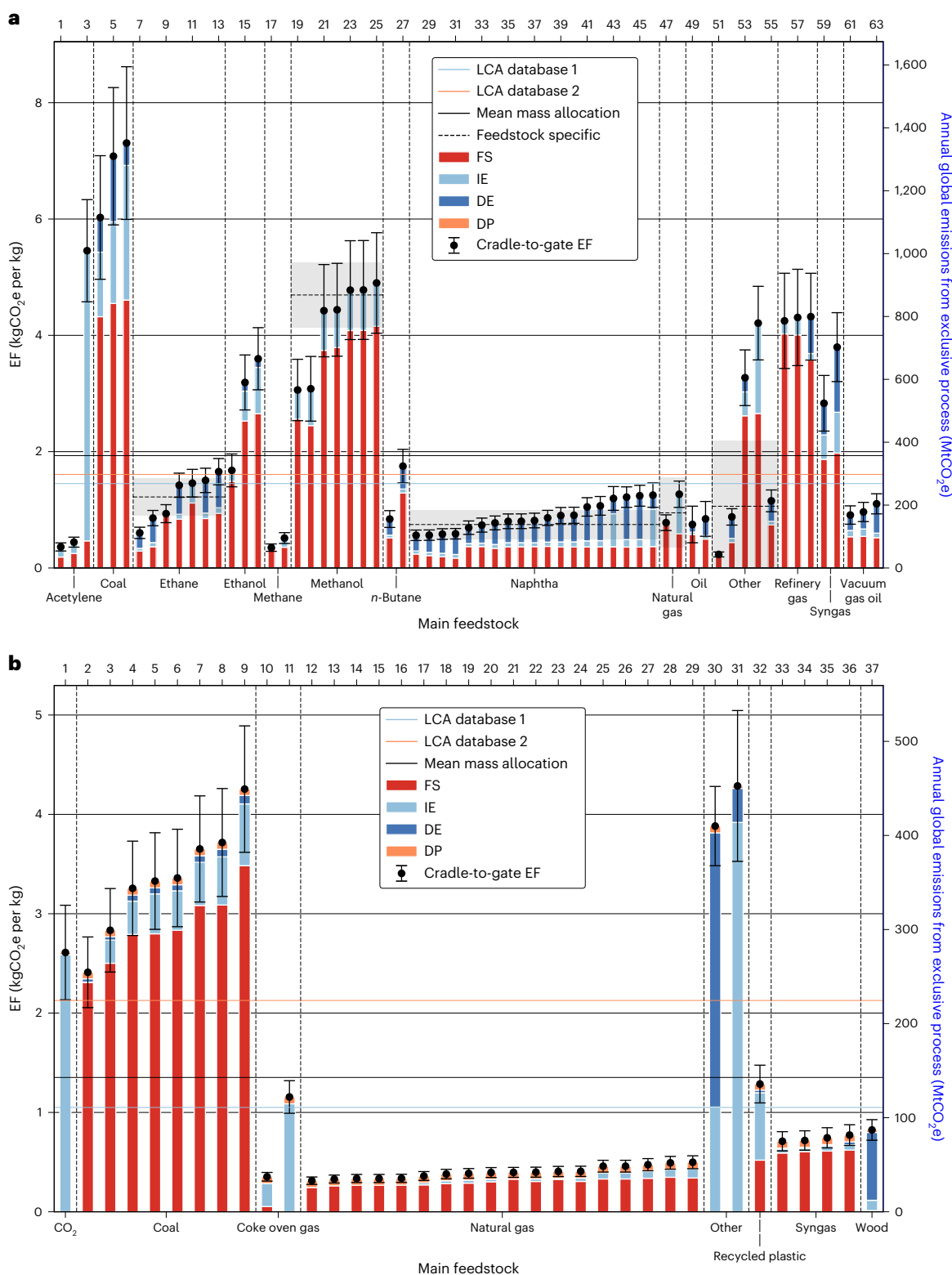


Fig. 4 | EFs for selected chemical manufacturing processes using mass allocation and grouped by principal feedstocks. a,b, Ethylene (a) and methanol (b) manufacturing processes using mass allocation and grouped by principal feedstocks. The contribution of each emissions source to the total cradle-to-gate EF is shown for each process. The cradle-to-gate EF shown by the black error bars is the sum of the four emissions sources. The uncertainty shown by the error bars corresponds to the 95% confidence interval as defined in Methods. The right axis shows the equivalent annual emissions from the product

if all the product was manufactured using each single process exclusively. The gray shaded areas represent the uncertainty bounds of the feedstock-specific EFs for ethylene that are used later in this study when weighting ethylene EFs according to facility feedstock information as detailed in Methods. The specific processes for each bar are detailed in Supplementary Section 2.3 and a dataset containing all calculated EFs is available at <https://doi.org/10.5281/zenodo.10532625>. Note that the numbers at the top of each figure represent the number of chemical manufacturing processes.

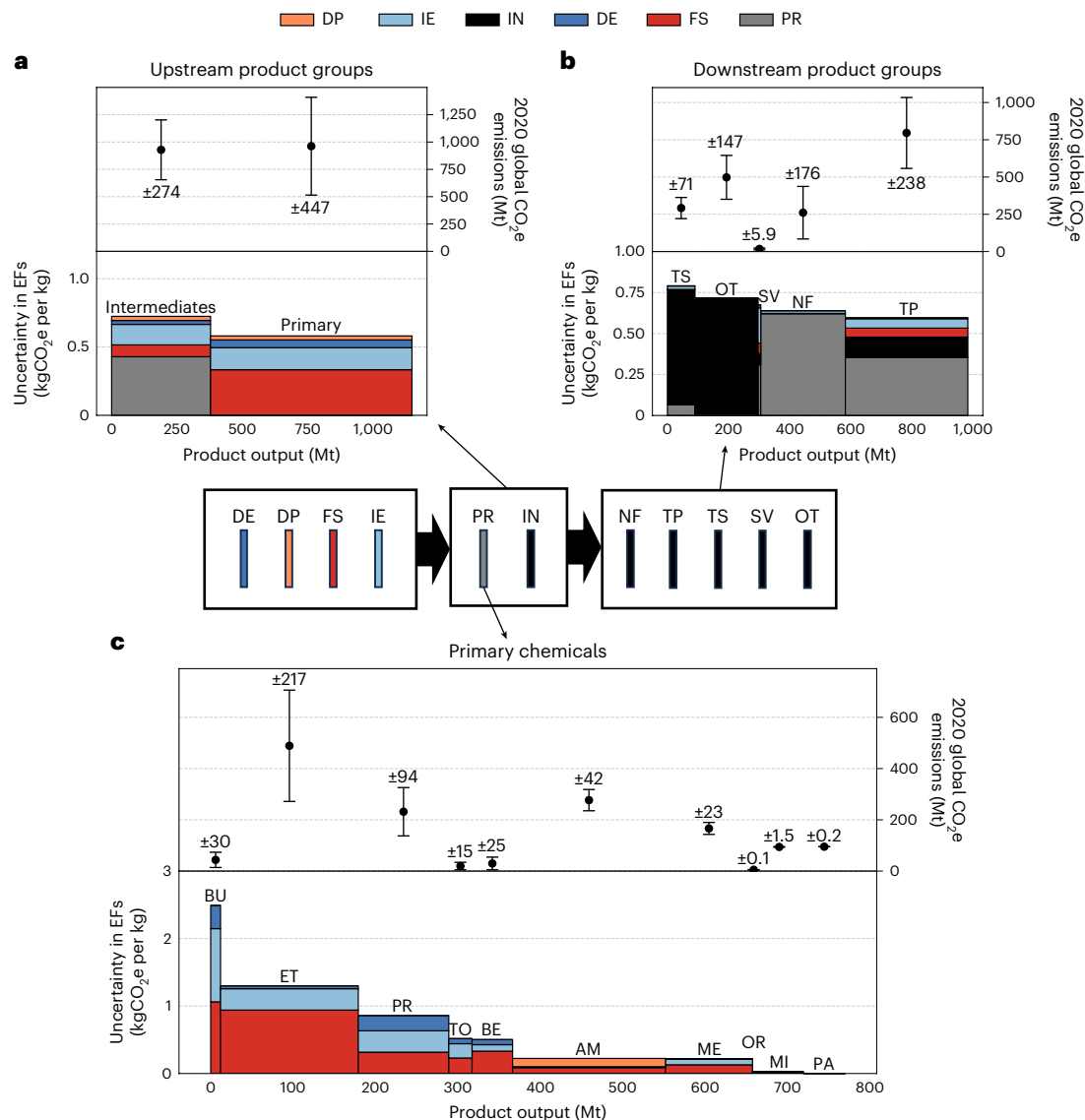


Fig. 5 | Impact of uncertainty in EFs on different product groups. a–c, Total emissions estimates are calculated by multiplying average EF uncertainties on the left axis with the output product mass on the x axis. The resulting absolute uncertainty in emissions estimates is represented by the area of the boxes and labeled in the plot of total emissions above each figure on the right axis. This uncertainty quantification represents the 95% confidence interval as defined

in Methods. For upstream product groups (a), downstream product groups (b) and primary chemicals (c). Full country-level EF data are available at <https://doi.org/10.5281/zenodo.10532625>. AM, ammonia; BU, butadiene; BE, benzene; ET, ethylene; ME, methanol; MI, mixed xylenes; OR, *ortho*-xylene; PA, *para*-xylene; PR, propylene; TO, toluene.

at the petrochemical production stage significantly affect LCA results for downstream products. For example, ref. 41 estimated the total emissions due to a 7.7 g high-density polyethylene grocery bag is 14.5 gCO₂e. If the uncertainty in ethylene production emissions of 0.5 kgCO₂e per kg is propagated through to high-density polyethylene, the actual emissions could be up to 35% higher, resulting in 14 ± 5 gCO₂e. For a car tire made of just 25% butadiene-based rubber, the production emissions of 334 kgCO₂e calculated in ref. 42 should incorporate an uncertainty of 31 kgCO₂e solely due to butadiene production. In line with results from Extended Data Table 1, data uncertainties originating from feedstocks and indirect energy use are more significant as a proportion of total uncertainties than those from direct energy use and direct processes for most products. An exception is ammonia where on average 55% of emissions are due to direct processes and therefore a higher proportion of uncertainty is due to direct processes.

Figure 5a shows that the total uncertainty for annual global primary chemical production is 459 MtCO₂e, which corresponds to 24% of

total GHG emissions from the petrochemical industry. This significant uncertainty propagates downstream to intermediate chemicals and downstream products seen in Fig. 5b. Figure 5b shows that thermoplastics are the downstream product group with the largest uncertainty with 238 MtCO₂e, largely due to high production volume. For downstream products, 85% of uncertainty originates from uncertainties in upstream production emissions, which propagate downstream to their use as inputs to downstream processes. To reduce uncertainties throughout the industry, the most valuable target is therefore upstream chemicals and in particular primary chemicals, where owing to high production volumes, ethylene, propylene and ammonia have the largest absolute uncertainties.

The first two results sections have explored the origins and impacts of uncertainty in EFs in the petrochemical industry. At a process level, low-production-volume products, including butadiene and thermosets, are the most susceptible to high uncertainties. At a global level, primary chemicals including propylene, ethylene and ammonia have

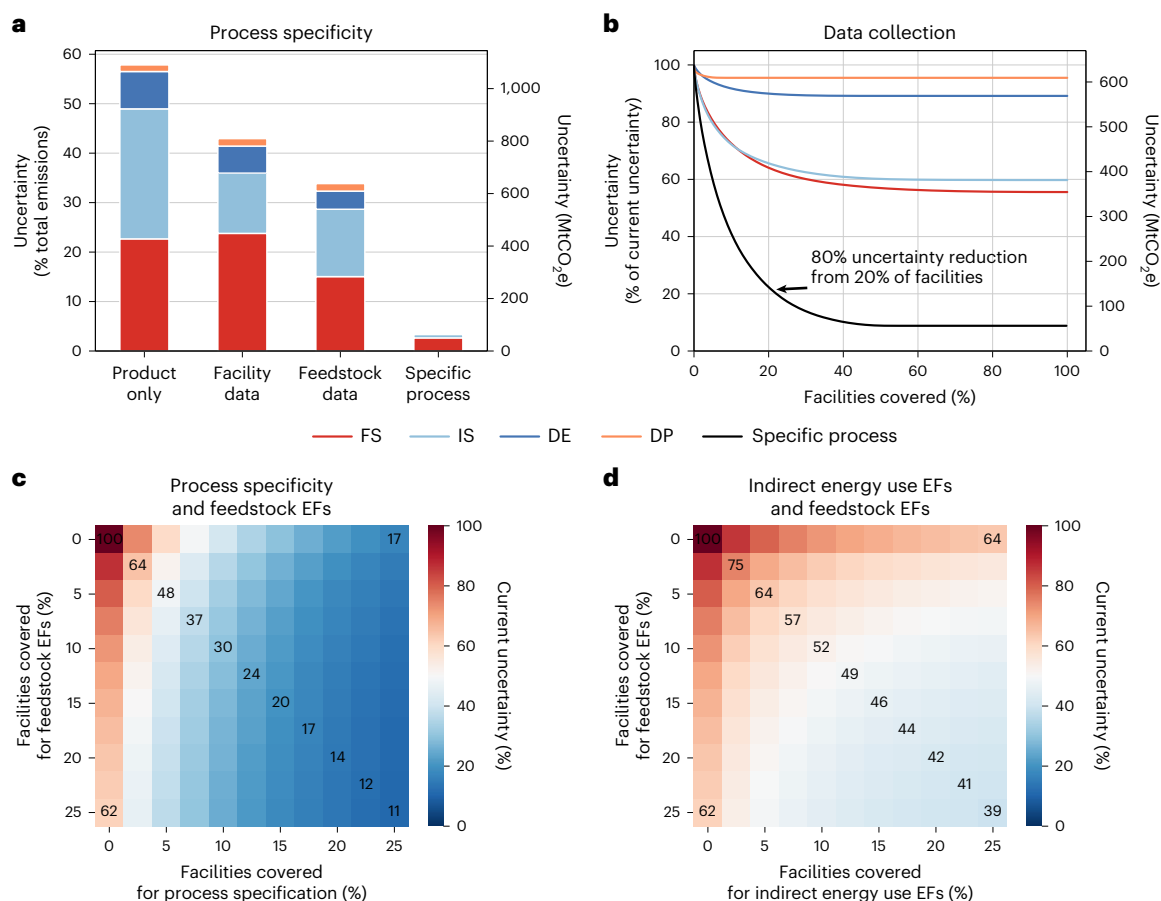


Fig. 6 | Uncertainty-reduction scenarios. a, Total uncertainty reduction from process specificity improvements. **b**, Total uncertainty reduction when targeting facilities in order of highest uncertainty values and eliminating uncertainty for each source. **c**, Uncertainty reduction as a proportion of current overall

uncertainty by combining data gathering for process specificity and feedstock EFs. **d**, Uncertainty reduction as a proportion of current overall uncertainty by combining data gathering for feedstock EFs and indirect energy use EFs.

the highest absolute uncertainty and a knock-on effect on downstream product uncertainties, making them the priority targets for uncertainty reduction. To understand the potential gains from future data collection across the four sources of data uncertainty and process specificity we will now establish uncertainty-reduction scenarios.

Uncertainty-reduction scenarios

Emissions uncertainty can be reduced by collecting additional data, via either LCAs or simply defining which exact process among a range of possible processes is being used at a facility. This section quantifies the potential of reducing uncertainty by considering the drop in uncertainty that would result from the collection of data to the point that an uncertainty source is eliminated for a given number of facilities. We consider the four data uncertainty sources and process specificity, as allocation uncertainty is less significant and cannot be directly targeted through further data collection. Figure 6a details the effects of reducing process specificity uncertainty for 100% of facilities by showing the average uncertainty ratio of process LCA emissions estimates at one facility and the equivalent absolute uncertainty when aggregated to the global level. Four levels of process specificity are considered: ‘product only’, where only the product of the facility is known and the mean EF of all processes is used; ‘facility data’, where details of the facility from the ICIS Supply and Demand Database are used to filter possible processes, and the mean EF of the remaining processes is used; ‘feedstock data’, where a weighted mean of possible processes is taken according to feedstock information for ethylene and ammonia, as described in Methods; and ‘specific process’, where hypothetical uncertainty if a

specific process is known for the facility. Figure 6b shows the drop in total uncertainty given the elimination of each uncertainty source individually at a given number of facilities. Facilities are ranked by the highest level of uncertainty and prioritized accordingly. Figure 6c,d shows the drop in uncertainty when combining data collection efforts across multiple uncertainty sources.

Figure 6a shows the significant uncertainty reduction that can be achieved through improving the specific knowledge of the facility process. In this study, we have used the ICIS Supply and Demand Database to improve uncertainties from the level of ‘product only’ to ‘facility data’ and furthered this by using feedstock weightings from the ICIS and the IFA to achieve an average facility-level uncertainty of 34% as denoted by the ‘feedstock data’ column. First, these data are not readily available in the public realm, which makes uncertainty reduction challenging. Second, we are still well above the hypothetically possible average uncertainty of 4% in a scenario where specific processes are known for all facilities. In the ‘specific process’ scenario, the remaining uncertainty is due to data uncertainties and allocation only. Third, weighting EFs by feedstock is effective for reducing uncertainty for ethylene and ammonia where processes can be easily grouped into types and data concerning input feedstocks exist, but it is not implementable for all products.

Figure 6b shows that by targeting the facilities with the highest overall uncertainty for future data collection, global uncertainty could be reduced by 80% by assigning specific processes to just 20% of facilities, for example, total uncertainty from yearly ethylene production emissions could be reduced from 217 Mt to 44 Mt by making the specific

process data available from 217 facilities. Other important opportunities lie in improving feedstock and indirect energy use uncertainty values which, leaving all other uncertainties constant, offer global uncertainty reductions of 36% and 34% respectively, with improved data from just 20% of facilities. Although this is promising, we must note that over 37,000 production facilities exist globally so covering 20% of plants is no trivial endeavor. Combining data gathering of process specificity and feedstock EFs would be the most efficient way to reduce overall uncertainty, as seen in Fig. 6c. However, specific process information can be sensitive data and may be challenging to obtain in some cases. Figure 6d shows the improvement to uncertainties that could be made independent of process specification, with a maximum uncertainty reduction of 61% if precise information is obtained for the feedstock and indirect energy use inputs to 25% of plants.

Emissions uncertainty at individual plants is not only an issue for process LCAs but also accumulates to create considerable uncertainty at a global scale. This section shows that there is potential to significantly reduce emissions uncertainty across the petrochemical industry through data collection and improved transparency, which would allow for process specification.

Discussion

EFs, essential in LCA and mandated by the United Nations Framework Convention on Climate Change framework⁴³, form the foundational basis for credible emissions reporting. Assigning accurate EFs to petrochemical production processes is challenging owing to the complexity of the industry, with numerous production processes for each type of product. The average uncertainty in process-level emissions estimates is 34% of total emissions, which aggregates to 0.6 GtCO₂e of the 1.9 ± 0.6 GtCO₂e annual global emissions from petrochemical production. LCA studies of common petrochemical products, including plastic bags, bottles and films, could be regularly inaccurate by up to 40% due to primary chemical production uncertainties and over 100% inaccurate if supply chains include uncommon production methods. Average uncertainties across downstream petrochemical EFs range from 15% to 40%. Therefore, while initial estimates facilitated by generic LCA database factors can be useful as policy guidance, they fall short in detailed comparative studies and decarbonization scenario analyses.

This study critically examines the origins of uncertainties, urging a move toward precision at the facility level in emissions assessments. The foremost source of uncertainty emerging in this study is the detail of specific production methods employed at individual facilities, which are largely unavailable in the public domain. This highlights the challenges posed by industrial confidentiality, which is a hurdle to comprehensive emissions estimation. While the choice of allocation method is the least impactful of the six uncertainty sources evaluated, inconsistent practices can hinder cross-study emissions comparisons. Proposing a standardized LCA allocation method for petrochemical emissions similar to encouraged practices in other industries, such as construction⁴⁴, can foster uniformity and transparency in emissions accounting.

Data inputs are responsible for the remainder of uncertainty, after process specificity and allocation. Upstream emissions from the production of feedstocks and off-site energy generation each contribute about half of the remaining uncertainty, with on-site fuel combustion and chemical reaction emissions making small contributions. Precise knowledge of emissions from chemical reactions reduces uncertainties related to direct process emissions. Previous studies have shown that variability in upstream feedstock sources can lead to large uncertainties³⁰. Planned improvements to the material-specific uncertainty quantification in ecoinvent will allow for more detailed uncertainty assessment of specific products but this is unlikely to significantly impact the results detailed in this study, across the industry as a whole. Indirect energy use uncertainties stem largely from electricity production and may be the easiest source of uncertainty reduction given

widespread data availability in this sector. Additional granularity for both upstream feedstock and indirect energy use sources could be combined with this study to provide a more holistic life-cycle uncertainty assessment, which could extend to use and end-of-life phases.

Uncertainties due to primary chemicals account for 70% of total uncertainties in petrochemical production emissions, which propagates throughout downstream products owing to the widespread use of primary chemicals as inputs. Addressing uncertainties linked to primary chemicals such as propylene, ethylene and ammonia emerges as a priority for researchers aiming to reduce overall emissions uncertainty. Strategic data collection is key for effective uncertainty reduction and Fig. 6 shows that global uncertainty in emissions can be substantially reduced by targeting just 20% of production facilities. Although this is a considerable challenge given the scale of the global petrochemical industry across over 37,000 facilities, the reward for implementing such a data-driven strategy could be effective decarbonization strategies grounded in a reliable assessment of current and future GHG emissions. To implement this, the meticulous performing of LCAs must become an intrinsic part of a chemical engineer's education.

In the era of intensified scrutiny of GHG emissions and the rapid growth of net-zero commitments, recalibrating the approach to uncertainties within the petrochemical sector is crucial. Generic EFs, while convenient, inadequately capture the diversity of processes and production methods used. Through enhanced data transparency, technological innovation and the pursuit of facility-level precision, chemical engineers have the potential to lead the push for accurate GHG emissions estimation. Engineers dedicated to uncertainty reduction should prioritize primary chemical production and facilities that account for the largest uncertainties. Subsequent investigations should build on this foundation, incorporating uncertainties from petrochemical use-phase and end-of-life scenarios to establish a comprehensive life-cycle understanding, thereby pinpointing the targets for GHG emissions uncertainty reduction.

Methods

This study develops a process-based LCA model to generate a cradle-to-gate EF estimate for 2,043 petrochemical production processes broken down into four sources of emissions: feedstocks, indirect energy use, direct energy use and direct processes. Emissions are also released from the use phase of some petrochemicals (for example, fertilizers) and from end-of-life product treatment, both of which are excluded from this study. Other environmental impacts can occur from sources other than GHGs, including fertilizer run-off contributing to eutrophication, bioaccumulation of toxic chemicals in organisms, and plastic waste in the world's oceans harming sea life, but are outside the scope of this study. In this section, we first discuss the calculation process and allocation methods applied to obtain process cradle-to-gate EFs. Second, we discuss the uncertainty sources and the propagation of uncertainties through each step of our calculations. Finally, we discuss the disaggregation of country-level production mass data to the inventory of facilities to establish the overall impact of uncertainties on industry-wide emissions estimates.

EF calculation

EFs are estimated for 2,043 petrochemical production processes using the mass–energy balances between inputs and outputs for each type of process, known as the ‘process recipes’, obtained from the IHS PEP yearbook³⁴. The IHS database contains process simulations and datasets that have been verified by industrial experts. From the output of individual GHGs, CO₂e global warming potentials are calculated following the 100-year horizon published by the Intergovernmental Panel on Climate Change (IPCC)⁴⁵. Given the cradle-to-gate focus, biogenic emissions are not distinguished and are included as part of overall emissions. The overall EF for an individual process is

calculated as an addition of the four emissions sources following equation (1).

$$EF_{\text{process}} = EF_{\text{feedstocks}} + EF_{\text{indirect energy use}} + EF_{\text{direct energy use}} + EF_{\text{direct process}} \quad (1)$$

Upstream chemicals can be used as inputs to downstream chemical production, in which case the emission factor $EF_{\text{intermediate}}$ is added to equation (1) for the calculation of the downstream processes' EF. Ignoring uncertainty propagation, which is covered above, the EFs for each emissions source are calculated as follows:

- (1) 'Feedstock' emissions, defined in equation (2), originate from the sum of GHG emissions embedded in the supply chains of each feedstock f for a total of F feedstocks used for a particular process. This is dependent on the EF of each feedstock, specific to the region of use extracted from ecoinvent 3.8¹⁸, the quantity Q_f of each feedstock used according to IHS process recipes and the total mass m of all output products produced following the recipe, typically 1 kg.

$$EF_{\text{feedstocks}} = \sum_{f=1}^F \frac{EF_f \times Q_f}{m} \quad (2)$$

- (2) 'Indirect energy use' includes any emissions embodied in energy generation, including electricity, undertaken off-site. This is region specific, depending on the energy mix in each region when attributed to individual facilities. This is taken into account by employing the relevant EF_e coefficients from ecoinvent 3.8 (ref. 18), with Q defined in units of energy for each energy source e from all energy sources E , and total mass m of all output products.

$$EF_{\text{indirect energy use}} = \sum_{e=1}^E \frac{EF_e \times Q_e}{m} \quad (3)$$

- (3) 'Direct energy use' represents any CO₂e emissions that originate from the on-site combustion of fuels to generate heat. This is calculated as the sum of combustion emission factors, sourced from the IPCC and US Department of Energy^{46,47} for each energy source e and calculated according to equation (4) with factors as defined for equation (3).

$$EF_{\text{direct energy use}} = \sum_{e=1}^E \frac{EF_e \times Q_e}{m} \quad (4)$$

- (4) 'Direct process' emissions originate from the chemical reactions involved in production. Stoichiometric ratios determine the output quantity of GHG molecules released from a reaction compared with the output quantity of the desired atoms used in a process products (for example, carbon and nitrogen). These are based on equations obtained from the IPCC⁴⁸ and are a combination of the molecular masses M of the GHG being analyzed and the chemical product. Direct process emissions resulting from the oxidation of input chemicals are calculated on a stoichiometric basis assuming all carbon is fully oxidized to CO₂ and all nitrogen is emitted as NO₂. Data on other potential GHG emissions (methane) are not available and are assumed to be negligible. In equation (5), the stoichiometric ratio is $C = M_{\text{GHG}}/M_{\text{product}}$

$$EF_{\text{direct process}} = \frac{m_{\text{input}}}{m_{\text{product}}} C \quad (5)$$

Chemical production processes often yield co-products alongside the product under consideration. To avoid double counting of emissions, total process emissions are allocated between co-products. To investigate the effect of the choice of allocation method on overall

emissions, we calculate three separate EFs for each process by using mass, energy and economic allocation. In each case, the emissions allocated to a product from a facility are proportional to its ratio of the mass, energy or cost relative to the entirety of the co-products (see illustration in Supplementary Fig. 2). Equation (6) defines the EF of a co-product c from the following allocation according to property X from the total process EF.

$$EF_{\text{co-product}} = \frac{X_c}{\sum_{c=1}^C X_c} EF_{\text{process}} \quad (6)$$

Process recipes are defined by the IHS as mass balances; therefore, for energy allocation, product masses are converted to equivalent energy using conversion factors from the 1996 IPCC guidelines³⁶. Similarly, mass is converted to cost by using cost factors published by the IHS for the year 2020⁴⁹. Energy and economic allocation are only calculated for co-products where the conversion data are available. After allocation, we have EF_{product} values across the four emissions sources for every process, following each allocation method. The mean of values from each allocation method is taken to obtain a single EF for each emission source corresponding to each process.

To calculate facility-specific EFs, we implement an automated algorithm matching each of 37,379 facilities to possible production processes based on each facility's product, route and technology information from the ICIS⁵. In instances when a unique match was not found and multiple possible processes p exist for a facility, the mean of EFs of all possible processes for that facility is used as stated in equation (7) where P is the total number of matching processes.

$$EF_{\text{facility}} = \frac{\sum_{p=1}^P EF_p}{P} \quad (7)$$

In the exceptional cases of ethylene and ammonia production, we go beyond attribution using the facility data alone, by incorporating additional feedstock ratio information from the ICIS⁵ and the IFA³. The feedstock ratios are used to improve accuracy by taking a weighted mean of possible processes. Processes p are grouped into their principal feedstock categories (for example, naphtha and methanol), and weighted according to ratio r of each input feedstock category f , as shown in equation (8).

$$EF_{\text{facility}} = \sum_{f=1}^F \left(r_f \times \frac{\sum_{p=1}^P EF_p}{P} \right) \quad (8)$$

The output of the facility attribution step leads to individual cradle-to-gate EFs for each facility broken down into the four emissions sources for each product. These EFs can be combined with production statistics to estimate total GHG emissions as seen above. The next section details the aggregation and propagation of uncertainties through the calculations in this section.

Uncertainty in EF calculations

In this study, we use an analytical approach to uncertainty estimation for a fully transparent and exhaustive quantification of the contributions from different uncertainty sources. Equations (2)–(5) define four emissions sources that correspond to four sources of data uncertainty. We follow the intermediate recommendation of ref. 50, by characterizing data uncertainty for each term in equations (2)–(5) as a normal probability distribution. Uncertainties throughout the study will therefore be expressed as the extent of the 95% confidence interval (CI) of the distribution, equivalent to 1.96 times the standard deviation of the distribution. Uncertainty associated with each data source is summarized in Extended Data Table 2 where CIs are written as a percentage of the mean value.

In most cases, uncertainty is unspecified fromecoinvent and in that case, we follow the basic uncertainty variance for CO₂ emissions of 0.0006 proposed in ref. 51. Following the uncertainty estimation methodology in ref. 52, and assuming their default pedigree matrix rankings (2, 2, 1, 5, 1), this results in a CI of 10% attributed to the unit EF_f or EF_e. The values in IHS process recipes are subject to up to a 5% CI⁴⁹, which is attributed to the quantities Q_f, Q_e and *m*. Molecular masses are known precisely, and uncertainty is deemed negligible for M_{product} and M_{GHG}. Similarly, chemical reactions are well understood, and combustion is optimized in industry, but a 1% CI is attributed to *C* to account for process losses, following optimal yield rates for primary chemicals in ref. 36. Where uncertainty is not explicitly stated, a 1% CI is assumed for the conversion from mass to energy in X_i as these ratios are consistent and well established. Finally, uncertainty is not published for the source of economic value to mass ratios X_i, but as IHS records indicate that product prices can vary by up to 10% within a year⁴⁹, we use 10% as a CI for cost values.

Uncertainty is propagated through each calculation following the standard Taylor series method for uncertainty propagation⁵³. Hence, for a value *V* calculated from variables *A*, *B*, ..., *N*, the posterior distribution of *V* is obtained from calculating the posterior standard deviation σ(*V*) according to equation (9) for multiplications, such as equations (2)–(5), and according to equation (10) for additions such as equation (1).

$$\sigma(V) = |V| \times \sqrt{\left(\frac{\sigma(A)}{A}\right)^2 + \left(\frac{\sigma(B)}{B}\right)^2 + \dots + \left(\frac{\sigma(N)}{N}\right)^2} \quad (9)$$

$$\sigma(V) = \sqrt{\sigma(A)^2 + \sigma(B)^2 + \dots + \sigma(N)^2} \quad (10)$$

Therefore, the data uncertainties considered in equations (2)–(5) for each process are:

- The uncertainty in feedstock EFs σ(EF_f), the uncertainty in the quantity of each feedstock used during the process in question to make the mass *m* of the product σ(Q_f).
- The uncertainty in indirect energy EFs σ(EF_e), the uncertainty in the quantity of indirect energy (that is, primarily electricity) used during the process in question to make the mass *m* of the product σ(Q_e).
- The uncertainty in direct EFs σ(EF_d), the uncertainty in the quantity of direct energy (that is, natural gas and oil combusted for energy) used during the process in question to make the mass *m* of the product σ(Q_d).
- The uncertainty in the stoichiometric ratio σ(*C*).

Beyond the four sources of data uncertainty introduced in equations (2)–(5) and propagated through subsequent calculations, an element of model uncertainty is introduced due to the three choices of allocation method possible; this is the fifth uncertainty source and will be known as ‘allocation uncertainty’. Where economic or energy conversions are available, two or three values with associated uncertainty distributions result from equation (6). To take into account the uncertainty distributions associated with the results of each allocation, we compare two values: (1) the mean of the standard deviations associated with the EFs from each allocation method, and (2) the standard deviation of the means for the EFs from each allocation method. The greater of the two values is taken as the standard deviation for EF_{product}. Equation (6), therefore, requires the input of the uncertainty associated with the total of the data uncertainties resulting from equation (1) σ(EF_{process}), and the uncertainties σ(X_i) corresponding to each co-product for each allocation method considered.

A second element of model uncertainty is introduced during the averaging of possible processes attributed to each facility; this is the sixth uncertainty source and will be known as ‘process non-specificity’.

If the exact process used at a facility is known this step is avoided and only five sources of uncertainty exist. The method for uncertainty propagation in this step is the same as with allocation uncertainty, where the greater of the mean of the standard deviations, and the standard deviation of the means is used as the standard deviation of EF_{facility}. The only difference is in the calculation of the standard deviation of the means. When more than three types of process are possible at a facility, processes with EFs lying beyond three standard deviations of the mean were flagged. We proceeded to research these processes individually and excluded them from the sample if they had not yet been rolled out beyond demonstration plants that did not correspond to the facility in question. This is a measure that avoids bias in facility EFs from very new low-emission bio-based processes. As a result, the only input uncertainty to equations (7) and (8) is the uncertainty associated with the EF calculated for each process σ(EF_p). The identification of six uncertainty sources, four data uncertainties and two model uncertainties allows us to analyze the impact on overall emissions uncertainties of different parts of emissions calculations and to identify the greatest opportunities for uncertainty reduction.

GHG emission estimation

Production mass data for 81 large-volume chemicals and fertilizers in 2020 were obtained from the ICIS⁵ and the IFA³. Capacity data for the 37,379 petrochemical manufacturing facilities were extracted from the ICIS Supply and Demand Database⁵. To attribute country-level production to individual facilities, an equal capacity utilization ratio is assumed per country and product; see Supplementary Section 1.2 for a diagram of this attribution. Uncertainties for facility capacity and regional production are not explicitly stated from the data sources, but the ICIS methodology states that uncertainties can be up to 10% for facility capacity⁵. Previous carbon budget studies and the IPCC guidelines for activity data suggest 7% uncertainty^{48,54} for production data. Combining these with the assumption of a uniform utilization rate, we define a 95% CI of 10% for overall facility production. This source of uncertainty is not part of the EF calculation process but must be considered when considering total emissions quantities rather than emissions intensity factors.

Given the production mass at each facility and the EF from above of each facility and product, the corresponding GHG emissions can be simply calculated according to equation (11), with uncertainties propagated according to equation (9).

$$\text{emissions}_{\text{facility,product}} = m_{\text{production}} \times \text{EF}_{\text{facility,product}} \quad (11)$$

In the chemical industry, downstream processes often use upstream products as inputs. To avoid double counting in considering the emissions of the whole petrochemical industry, emissions from the production of upstream chemicals that are then used in downstream processes are deducted from the total.

Overview and limitations. This study considers six sources of uncertainty in EF estimation: feedstocks, indirect energy use, direct energy use, direct processes, allocation and process specificity. Further uncertainties from production estimates are incorporated when calculating total GHG emissions estimates. Another source of uncertainty not explicitly included is the choice of model boundaries including: a system boundary other than cradle-to-gate used for EFs in this study could be considered, the temporal resolution of data, the technology readiness level of processes considered and the presence of paywalls for industrial data that may lead to missing parts of the industry. Displacement (system expansion) is an alternative system boundary but might not be a suitable option for well-established industries and products that are unlikely to replace chemical production elsewhere. This practice is also observed in commercial databases¹⁹ and existing literature¹⁰. In the next stage of this analysis, adopting the system expansion method could be considered.

This study is limited by the scope of analyzing 81 chemicals, which does not cover all petrochemical products. The analysis focuses on the largest volume of petrochemicals and any chemicals excluded are likely to be associated with higher levels of uncertainty than reported in this paper due to the variability of production at small scale. A further limitation is the omission of uncertainties from manufacturing processes further downstream than those considered, the use phase and end-of-life emissions. Future studies could address these limitations and provide a full life-cycle understanding of the impact of uncertainties on petrochemical emissions. Nonetheless, the major conclusions about the largest sources of emissions and prioritization should not be significantly affected by this. First, missing smaller products should have a small overall effect on absolute uncertainty. Second, issues that affect the whole system, such as the system boundary, will tend to have a similar effect on all results and a smaller effect on comparisons.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data are publicly available. However, some cases require a user license from IHS Markit to access the underlying process recipes. To gain access, IHS Markit (now belonging to S&P Global) can be contacted via the following website: <https://www.spglobal.com/commodity-insights/en/ci/products/chemical-technology-pep-index.html>. The data shown in the figures are available at <https://doi.org/10.6084/m9.figshare.23618862.v3> (ref. 55) and full versions of calculated emissions intensity factors for each process and per country are available at <https://doi.org/10.5281/zenodo.10532626> (ref. 56). An interactive beta version dashboard detailing the full global emissions and uncertainties from this study can be accessed at <https://9e4z.short.gy/c-thru-petrochemical-emissions>.

Code availability

The code used for the analysis was generated using Python and is available at <https://github.com/Resource-Efficiency-Collective/chemical-uncertainties> (ref. 57).

References

1. ICIS Supply and Demand Database (ICIS, 2021); <https://www.icis.com/explore/services/analytics/supply-demand-data/> (2021).
2. International Council of Chemical Associations *The Global Chemical Industry: Catalyzing Growth and Addressing Our World's Sustainability Challenges* (Oxford Economics, 2019); <https://icca-chem.org/wp-content/uploads/2020/10/Catalyzing-Growth-and-Addressing-Our-Worlds-Sustainability-Challenges-Report.pdf>
3. IFASTAT Databases (IFA, 2021); <https://www.ifastat.org/databases>
4. Meng, F. et al. Planet-compatible pathways for transitioning the chemical industry. *Proc. Natl Acad. Sci. USA* **120**, e2218294120 (2023).
5. Industry (IEA, 2022); <https://www.iea.org/reports/industry>
6. Chemicals. IEA <https://www.iea.org/reports/chemicals> (2023).
7. IPCC *Climate Change 2023: Synthesis Report* (eds Core Writing Team, Lee, H. & Romero, J.) (IPCC, 2023); <https://www.ipcc.ch/report/sixth-assessment-report-cycle/>
8. Barton, J. L. Electrification of the chemical industry. *Science* **368**, 1181–1182 (2020).
9. Gabrielli, P., Gazzani, M. & Mazzotti, M. The role of carbon capture and utilization, carbon capture and storage, and biomass to enable a net-zero-CO₂ emissions chemical industry. *Ind. Eng. Chem. Res.* **59**, 7033–7045 (2020).
10. Meys, R. et al. Achieving net-zero greenhouse gas emission plastics by a circular carbon economy. *Science* **374**, 71–76 (2021).
11. Falter, W., Langer, A., Wesche, F. & Wezel, S. Decarbonization strategies in converging chemical and energy markets. *J. Bus. Chem.* **2**, 20–40 (2020).
12. D'Angelo, S. C. et al. Planetary boundaries analysis of low-carbon ammonia production routes. *ACS Sustain. Chem. Eng.* **9**, 9740–9749 (2021).
13. Kätelhön, A., Meys, R., Deutz, S., Suh, S. & Bardow, A. Climate change mitigation potential of carbon capture and utilization in the chemical industry. *Proc. Natl Acad. Sci. USA* **116**, 11187–11194 (2019).
14. Galán-Martín, Á. et al. Sustainability footprints of a renewable carbon transition for the petrochemical sector within planetary boundaries. *One Earth* **4**, 565–583 (2021).
15. Stegmann, P., Daioglou, V., Londo, M., van Vuuren, D. P. & Junginger, M. Plastic futures and their CO₂ emissions. *Nature* **612**, 272–276 (2022).
16. Gregory, R., Satterfield, T. & Boyd, D. R. People, pipelines, and probabilities: clarifying significance and uncertainty in environmental impact assessments. *Risk Anal.* **40**, 218–226 (2020).
17. Hellweg, S., Benetto, E., Huijbregts, M. A. J., Verones, F. & Wood, R. Life-cycle assessment to guide solutions for the triple planetary crisis. *Nat. Rev. Earth Environ.* **4**, 471–486 (2023).
18. *The ecoinvent Database Version 3* (ecoinvent, 2021); <https://www.ecoinvent.org/database/database.html>
19. *LCA Database for Chemicals and Plastics* (Carbon Minds, 2022).
20. Parvatker, A. G. & Eckelman, M. J. Simulation-based estimates of life cycle inventory gate-to-gate process energy use for 151 organic chemical syntheses. *ACS Sustain. Chem. Eng.* **8**, 8519–8536 (2020).
21. Parvatker, A. G. & Eckelman, M. J. Comparative evaluation of chemical life cycle inventory generation methods and implications for life cycle assessment results. *ACS Sustain. Chem. Eng.* **7**, 350–367 (2019).
22. Herrmann, I. T., Hauschild, M. Z., Sohn, M. D. & McKone, T. E. Confronting uncertainty in life cycle assessment used for decision support: developing and proposing a taxonomy for LCA studies. *J. Ind. Ecol.* **18**, 366–379 (2014).
23. Marsh, E., Allen, S. & Hattam, L. Tackling uncertainty in life cycle assessments for the built environment: a review. *Build. Environ.* **231**, 109941 (2022).
24. Leow, W. R. et al. Electrified hydrocarbon-to-oxygenates coupled to hydrogen evolution for efficient greenhouse gas mitigation. *Nat. Commun.* **14**, 1954 (2023).
25. *Rulemaking Notices for GHG Reporting* (EPA, 2022); <https://www.epa.gov/ghgreporting/rulemaking-notices-ghg-reporting>
26. *Guidance on How to Measure and Report Your Greenhouse Gas Emissions* (DEFRA, 2009); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/69282/pb13309-ghg-guidance-0909011.pdf
27. Jing, L. et al. Carbon intensity of global crude oil refining and mitigation potential. *Nat. Clim. Change* **10**, 526–532 (2020).
28. Ma, S., Lei, T., Meng, J., Liang, X. & Guan, D. Global oil refining's contribution to greenhouse gas emissions from 2000 to 2021. *The Innovation* **4**, 100361 (2023).
29. Lei, T. et al. Adaptive CO₂ emissions mitigation strategies of global oil refineries in all age groups. *One Earth* **4**, 1114–1126 (2021).
30. Chen, Q., Dunn, J. B. & Allen, D. T. Mapping greenhouse gas emissions of the U.S. chemical manufacturing industry: the effect of feedstock sourcing and upstream emissions allocation. *ACS Sustain. Chem. Eng.* <https://doi.org/10.1021/ACSSUSCHEMENG.2C00295> (2022).
31. Ghanta, M., Fahey, D. & Subramaniam, B. Environmental impacts of ethylene production from diverse feedstocks and energy sources. *Appl. Petrochem. Res.* **4**, 167–179 (2014).

32. Liu, X., Elgowainy, A. & Wang, M. Life cycle energy use and greenhouse gas emissions of ammonia production from renewable resources and industrial by-products. *Green Chem.* **22**, 5751–5761 (2020).
33. Rafiqul, I., Weber, C., Lehmann, B. & Voss, A. Energy efficiency improvements in ammonia production—perspectives and uncertainties. *Energy* **30**, 2487–2504 (2005).
34. *Process Economics Program* (IHS Markit, 2021); <https://ihsmarkit.com/products/chemical-technology-pep-index.html>
35. *The Future of Petrochemicals: Towards More Sustainable Plastics and Fertilisers* (IEA, 2018); <https://www.iea.org/data-and-statistics/charts/direct-co2-emissions-by-scenario-2017-2050>
36. Levi, P. G. & Cullen, J. M. Mapping global flows of chemicals: from fossil fuel feedstocks to chemical products. *Environ. Sci. Technol.* **52**, 1725–1734 (2018).
37. Valente, A., Iribarren, D. & Dufour, J. Life cycle assessment of hydrogen energy systems: a review of methodological choices. *Int. J. Life Cycle Assess.* **22**, 346–363 (2017).
38. Björklund, A. E. Survey of approaches to improve reliability in LCA. *Int. J. Life Cycle Assess.* **7**, 64–72 (2002).
39. Luo, L., van der Voet, E., Huppel, G. & Udo de Haes, H. A. Allocation issues in LCA methodology: a case study of corn stover-based fuel ethanol. *Int. J. Life Cycle Assess.* **14**, 529–539 (2009).
40. FAOSTAT (FAO, 2021); <http://www.fao.org/faostat/en/#home>
41. Lewis, H., Verghese, K. & Fitzpatrick, L. Evaluating the sustainability impacts of packaging: the plastic carry bag dilemma. *Packag. Technol. Sci. Int. J.* **23**, 145–160 (2010).
42. Katarzyna, P., Izabela, P., Patrycja, B.-W., Weronika, K. & Andrzej, T. LCA as a tool for the environmental management of car tire manufacturing. *Appl. Sci.* **10**, 7015 (2020).
43. *Report of the Conference of the Parties on Its Nineteenth Session, Held in Warsaw from 11 to 23 November 2013 Addendum Part Two: Action Taken by the Conference of the Parties at Its Nineteenth Session Contents* (UNFCCC, 2014).
44. Gelowitz, M. & McArthur, J. Comparison of type III environmental product declarations for construction products: material sourcing and harmonization evaluation. *J. Clean. Prod.* **157**, 125–133 (2017).
45. Stocker, T. F. et al (eds). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).
46. IPCC 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories (eds Calvo Buendia, E. et al) (IPCC, 2019).
47. *GREET: The Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation Model* (Department of Energy, 2019); <https://www.energy.gov/eere/bioenergy/articles/greet-greenhouse-gases-regulated-emissions-and-energy-use-transportation>
48. *IPCC 2000 Revision to 1996 IPCC Guidelines for National Greenhouse Gas Inventories* (eds Houghton, J. et al) (IPCC/OECD/IEA: UK Meteorological Office, 2000).
49. *Chemical Economics Handbooks (CEH)* (IHS Markit, 2021); <https://ihsmarkit.com/products/chemical-economics-handbooks.html>
50. Igos, E., Benetto, E., Meyer, R., Baustert, P. & Othoniel, B. How to treat uncertainties in life cycle assessment studies? *Int. J. Life Cycle Assess.* **24**, 794–807 (2019).
51. Weidema, B. P. et al. *Overview and Methodology: Data Quality Guideline for the Ecoinvent Database Version 3* (St. Gallen: The ecoinvent Centre, 2013).
52. Ciroth, A., Muller, S., Weidema, B. & Lesage, P. Empirically based uncertainty factors for the pedigree matrix in ecoinvent. *Int. J. Life Cycle Assess.* **21**, 1338–1348 (2016).
53. Coleman, H. W. & Steele, W. G. *Experimentation, Validation, and Uncertainty Analysis for Engineers* (John Wiley & Sons, 2018).
54. Friedlingstein, P. et al. Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
55. Cullen, L. Reducing greenhouse gas emissions uncertainties to improve decarbonisation strategies for petrochemical production. *figshare* <https://doi.org/10.6084/m9.figshare.23618862.v3> (2023).
56. Cullen, L. Reducing greenhouse gas emissions uncertainties for chemical production. *Zenodo* <https://doi.org/10.5281/zenodo.10532625> (2024).
57. Cullen, L. Repository for code relating to paper entitled “Reducing greenhouse gas emissions uncertainties for chemical production”. *GitHub* <https://github.com/Resource-Efficiency-Collective/chemical-uncertainties> (2024).

Acknowledgements

We acknowledge support from C-THRU: carbon clarity in the global petrochemical supply chain (www.c-thru.org).

Author contributions

Conceptualization: L.C., F.M. and J.M.C. Methodology: L.C., F.M. and J.M.C. Formal analysis: L.C., F.M., R.L. and J.M.C. Writing—original draft: L.C. and F.M. Writing—review and editing: L.C., F.M., R.L. and J.M.C. Supervision: F.M. and J.M.C. Funding acquisition: R.L. and J.M.C.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s44286-024-00047-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44286-024-00047-z>.

Correspondence and requests for materials should be addressed to Fanran Meng or Jonathan M. Cullen.

Peer review information *Nature Chemical Engineering* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Extended Data Table 1 | Emissions sources ranked by average contribution to EF uncertainty across all petrochemical production processes

Emissions source	Contribution to EF	Data uncertainty	Contribution to EF uncertainty
Feedstock	51%	15%	56%
Indirect energy use	35%	16%	40%
Direct energy use	2.9%	9.7%	2.0%
Direct process	11%	2.0%	1.5%

Contribution to EF and data uncertainty are also an average across all processes used in this study.

Extended Data Table 2 | Input parameters for all emissions sources and associated uncertainties

Input parameters	Data Type	Source	Uncertainty
EF_f, EF_o (FS, IE, DE)	Emission factors (kg CO _{2e} /kg)	ecoinvent & IPCC ^{18,45}	10%
Q_f, Q_o (FS, IE, DE)	Quantity of input (kg/kg product)	IHS ⁴⁹	5%
m, m_{input} (FS, IE, DE, DP)	Masses of output product and input gas for DP (kg)	IHS ⁴⁹	5%
$M_{GHG}, M_{product}$ (DP)	Molecular masses (kg/mol)	Lide et al. ⁵⁵	0%
C (DP)	Process reaction equations (kg GHG/kg)	Levi and Cullen ³⁶	1%
X_i - energy (Allocation)	Energy intensities (kJ/kg)	Levi and Cullen ³⁶	1%
X_i - economic (Allocation)	Cost values(\$/kg)	IHS ⁴⁹	10%

The sources in which each parameter are involved are in brackets in the left column: FS - Feedstock, IE - Indirect energy use, DE - Direct energy use, DP - Direct process.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection All code for collecting data from sources stated in the data availability statement are provided at <https://github.com/Resource-Efficiency-Collective/chemical-uncertainties>.

Data analysis All analysis code is publicly available at <https://github.com/Resource-Efficiency-Collective/chemical-uncertainties>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data are publicly available. However, some cases require a user license from IHS Markit to access the underlying process recipes. To gain access, IHS Markit can be contacted via the following website: <https://ihsmarkit.com/products/chemical-technology-pep-index.html>. Data shown in the figures are available at <https://doi.org/10.6084/m9.figshare.23618862.v3> and full versions of calculated emissions intensity factors for each process and per country are available at <https://>

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Use the terms *sex* (biological attribute) and *gender* (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.
Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Reporting on race, ethnicity, or other socially relevant groupings

Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status). Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)
Please provide details about how you controlled for confounding variables in your analyses.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.

Data exclusions

Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.

Replication

Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.

Randomization

Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.

Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).

Research sample

State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic

Research sample	<i>information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>
Randomization	<i>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</i>

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.</i>
Research sample	<i>Describe the research sample (e.g. a group of tagged <i>Passer domesticus</i>, all <i>Stenocereus thurberi</i> within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.</i>
Sampling strategy	<i>Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.</i>
Data collection	<i>Describe the data collection procedure, including who recorded the data and how.</i>
Timing and spatial scale	<i>Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Reproducibility	<i>Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.</i>
Randomization	<i>Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.</i>
Blinding	<i>Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.</i>

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.