DOI: 10.1111/jiec.13399

JOURNAL OF INDUSTRIAL ECOLOGY WILEY

RESEARCH ARTICLE

Expert elicitation and data noise learning for material flow analysis using Bayesian inference

Jiayuan Dong | Jiankan Liao | Xun Huan | Daniel Cooper

Mechanical Engineering, University of Michigan, Ann Arbor, Michigan, USA

Correspondence
Daniel Cooper, Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA.
Email: drcooper@umich.edu

Editor Managing Review: Wei–Qiang Chen

Funding information
National Science Foundation, Grant/Award Number: #2040013

Abstract

Bayesian inference allows the transparent communication and systematic updating of model uncertainty as new data become available. When applied to material flow analysis (MFA), however, Bayesian inference is undermined by the difficulty of defining proper priors for the MFA parameters and quantifying the noise in the collected data. We start to address these issues by first deriving and implementing an expert elicitation procedure suitable for generating MFA parameter priors. Second, we propose to learn the data noise concurrent with the parametric uncertainty. These methods are demonstrated using a case study on the 2012 US steel flow. Eight experts are interviewed to elicit distributions on steel flow uncertainty from raw materials to intermediate goods. The experts' distributions are combined and weighted according to the expertise demonstrated in response to seeding questions. These aggregated distributions form our model parameters' informative priors. Sensible, weakly informative priors are adopted for learning the data noise. Bayesian inference is then performed to update the parametric and data noise uncertainty given MFA data collected from the United States Geological Survey and the World Steel Association. The results show a reduction in MFA parametric uncertainty when incorporating the collected data. Only a modest reduction in data noise uncertainty was observed using 2012 data; however, greater reductions were achieved when using data from multiple years in the inference. These methods generate transparent MFA and data noise uncertainties learned from data rather than pre-assumed data noise levels, providing a more robust basis for decision-making that affects the system.

KEYWORDS
Bayesian inference, Bayes factor, data noise, material flow analysis, prior elicitation and aggregation, uncertainty quantification

1 | INTRODUCTION

Material flow analysis (MFA) is a foundational tool of industrial ecology research and characterizes how a given material is transported and transformed through a supply chain. MFAs are key to identifying potential resource efficiency improvements (e.g., increased recycling), and to evaluating the upstream and downstream system impacts of local interventions; for example, the potential to reduce greenhouse gas (GHG) emissions released during material production by improving downstream manufacturing process yields (Cullen & Cooper, 2022). MFAs have been used to help set environmental policies and goals by national governments (e.g., justifying Japan's reduce, reuse, and recycling laws), local governments (e.g., remedial action taken against toxic releases into New York City harbor), and companies (e.g., Toyota's corporate MFA was used to set company goals for emissions and recycling) (Graedel, 2019). The proliferation of MFA, however, is hindered by at least two major challenges. First is the long timeline for creating and updating detailed MFAs, currently taking months or even years. Second is the lack of uncertainty quantification (UQ) in most MFA results. The data on mass flows collected to conduct an MFA are often incomplete, noisy, conflicting, and out-of-date. Poor data quality necessitates data reconciliation, and alongside data sparsity, is at the root of MFA uncertainty. Without UQ, MFA results provide limited insight into the impacts, risks, and unintended consequences of system interventions. It is increasingly accepted that UQ must be included in MFA results if they are to be meaningful and able to support informed decision- and policy-making (Graedel, 2019; Schwab & Rechberger, 2018). Bayesian methods help address these challenges of UQ and laboriousness in MFA, as explained below.

1.1 | Previous work on data reconciliation and Bayesian inference in MFA

An MFA practitioner must create mass-balanced networks from unbalanced data. Manual reconciliation is common, as used in Cullen et al.'s seminal work on global steel and aluminum flows (Cullen & Allwood, 2013; Cullen et al., 2012). This requires the practitioner to use his or her judgment to adjust the data to achieve mass balance. Otherwise, more formal approaches to MFA data reconciliation include (non)linear least squares optimization (Kopec et al., 2016; Zhu et al., 2019; Cencic, 2016) and Bayesian inference (Gottschalk et al., 2010; Lupton & Allwood, 2018).

The popular "STAN" open-source software performs MFA least squares data reconciliation and uses error propagation to quantify the uncertainty in the final mass flow results if the collected data are input into the software as probability distributions (Cencic, 2016). However, defining these distributions is troublesome given the MFA data being collected are rarely recorded with uncertainty information (Cencic, 2016). Elsewhere, Meylan et al. (2017) investigate the reliability of MFA results by either assigning reliability indicators to the collected data or by observing the degree of adjustment that the collected data undergo during the reconciliation process. A lower degree of data adjustment reflects fewer data conflicts and hints at higher reliability (Meylan et al., 2017) but such metrics do not quantify the uncertainty in the MFA flow parameters.

Bayesian inference is a probabilistic approach to uncertainty quantification that can be used to reconcile MFA data by adjusting the MFA variable estimates based on combining prior knowledge with collected material flow data (Jaynes, 2003; Probability Theory: The Logic of Science, 2008; Berger, 1985; Von Toussaint, 2011). The prior information is typically a combination of fact-based knowledge with subjective impressions based on experience (Moyé, 2008). The prior belief about an MFA variable, such as the mass flow between two processes in a factory, is expressed as a probability density function (PDF); for example, a Gaussian PDF could be used to represent a prior belief that a mass flow is expected to be 10 t with a variance of 1 t^2. MFA data are subsequently collected and the "noise" in the collected data—for example, due to the error in a mass sensor reading—is also expressed as PDFs. In Bayesian inference, the collected MFA data are combined with the priors to generate an updated posterior belief, represented as updated conditional PDFs.

Bayesian inference presents several benefits for MFA data reconciliation. First, it allows a rigorous quantification of result uncertainty (not proxies) via the probability and statistics formalism, and allows flexible probability distributions able to capture high-order non-Gaussian and correlation effects. For example, the prior knowledge of an MFA variable might be best represented as a uniform rather than normal distribution if nothing is known other than the upper and lower bounds on the variable; that is, with no knowledge to favor a higher probability region. Second, Bayesian inference is particularly applicable to MFA because it is well suited to handling sparse and noisy data, and can incorporate multiple data streams simultaneously. Third, compared with other approaches such as Gaussian error propagation (Bader et al., 2011), the Bayesian framework provides a natural entryway to inject domain knowledge, such as by using historical data or opinions from subject matter experts to form the prior distribution of the MFA variables (Wang & Romagnoli, 2003). Bayesian inference can also be "chained" together, to perform sequential learning that iteratively assimilates new data as they become available. Even when little data is available, the Bayesian approach can provide a practitioner with an MFA with associated uncertainties.

Bayesian inference was first used in MFA in 2010 by Gottschalk et al. (2010) to study nano-TiO_2 mass flows. They formed uniform and triangular prior distributions centered on values of historical data and performed Bayesian inference using a Metropolis sampling algorithm with simulated instead of measured data. In 2018, Lupton and Allwood (2018) introduced additional MFA prior forms (e.g., Dirichlet priors) and conducted a case study on deriving the global steel flow. Their case study highlights some of the challenges of applying Bayesian inference to MFA:

- **Assigning proper and rigorously justified prior distributions**. Lupton and Allwood's steel flow analysis (Lupton & Allwood, 2018) used previous results from Cullen et al. (2012) as the basis of their priors. However, historical data may not be always available and even when it is (and deemed relevant) it remains unclear how to form a probability distribution that properly reflects the uncertainty. Assuming a prior variance without justification can introduce bias to the posterior results. Alternatively, non-informative or weakly informative priors may be used; for example, assigning a wide uniform PDF for a mass flow between 0 and 200 Mt. However, they will likely require more MFA data to be collected in order to decrease uncertainty to desirable levels.

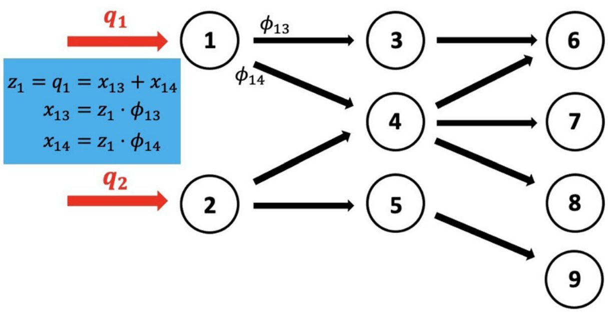


FIGURE 1 A graphical representation of an material flow analysis network structure.

- **Assigning noise to collected MFA data**. MFA relevant data are typically published without accompanying uncertainty information; for example, no error bars are given for the commodity mass flow data reported by the U.N. Comtrade Database (United Nations Comtrade Database, 2012) or from trade associations such as the World Steel Association (World Steel, 2012).

  How then to model the data noise? Assumptions can be made; for example, Lupton and Allwood (2018) assume the data noise in their collected data follows a Gaussian distribution with a standard deviation equal to 10% of the observed value. Such assumptions can introduce bias into the final posterior MFA results and provide a false sense of confidence if the assumed noise overestimates the data quality.

  Elsewhere, there are qualitative methods to categorize data into uncertainty levels based on features such as the perceived data source quality and specificity (Bonnin et al., 2013), and semi-quantitative approaches such as using a pedigree matrix or confidence score (Lloyd & Ries, 2007; Zhu et al., 2019) which translates an uncertainty level into a probability distribution or numerical value. However, the strict quantitative identification of MFA data noise has been lacking so far (Nga et al., 2014). The Bayesian framework offers the opportunity to quantify and *learn* the data noise.

1.2 | Scope and structure of this article

This paper explores how to: (a) Form informative prior distributions for MFA variables by eliciting information from industrial experts, and (b) Learn the data noise from collected data by incorporating data noise as random variables for inference. In Section 2, we first introduce the MFA problem mathematically using the conservation of mass principle (Section 2.1) and then formulate Bayesian inference for learning MFA model parameters and the collected data noise (Section 2.2). The two crucial components needed for solving a Bayesian inference problem—the likelihood and prior—are then presented in Sections 2.3 and 2.4, respectively. In particular, Section 2.4 reviews existing methods for expert elicitation and discusses their use for the MFA framework. In Section 3, we apply these methods to derive the US steel flow for 2012. Finally, in Section 4 we discuss the lessons learned from the case study.

2 | FORMULATION

2.1 | A mathematical representation of MFA

An MFA can be represented via a directed graph as shown in Figure 1. The nodes of the graph (indexed by $1, 2, \ldots, n_p$) represent different processes, products, or locations. Each directed edge connecting two nodes represents the mass flow of material from one process to another.

At the core of MFA is the conservation of mass, which requires the total mass of material flows into each node (total input) to equal the total mass of material flows out of each node (total output). We denote the total input (equivalently, total output) flow for node i by z_i. The flow along an edge out of node i (e.g., to node j) is then equal to $\phi_{ij} z_i$, where $\phi_{ij} \in [0, 1]$ is the allocation fraction of node i's total outflow going into node j ($\phi_{ij} = 0$ if there is no flow from node i to node j). Hence,

$$\sum_{i=1}^{n_p} \phi_{ij} z_i = z_j. \quad (1)$$

Furthermore, for each node, its output allocation fractions need to sum to unity:

$$\sum_{j=1}^{n_p} \phi_{ij} = 1. \quad (2)$$

We recommend working with the allocation fractions (ϕ_{ij}) as model parameters instead of working directly with the mass flow values due to the increased ease of performing the calculations. The allocation fractions offer a convenient method of expressing and allowing the mass balance relationships for the entire MFA to be assembled into a linear system as proposed by Gottschalk et al. (2010). For instance, the mass balance equations for the simple MFA shown in Figure 1 can be expressed as:

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\phi_{13} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\phi_{14} & -\phi_{24} & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\phi_{25} & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\phi_{36} & -\phi_{46} & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\phi_{47} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -\phi_{48} & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -\phi_{59} & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbb{I}-\Phi^T} \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \\ z_9 \end{bmatrix}}_{z} = \underbrace{\begin{bmatrix} q_1 \\ q_2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{q} \tag{3}$$

where in Equation (3), $\mathbb{I}$ is the $n_p \times n_p$ identity matrix, $\Phi \in \mathbb{R}^{n_p \times n_p}$ is the adjacency matrix where entries are the allocation fractions ϕ_{ij}, $z \in \mathbb{R}^{n_p}$ depicts all the nodal mass flows, and $q \in \mathbb{R}^{n_p}$ represents the external inflows to the network. The material inflows (q_i) originate from outside the network. For example, a material inflow to the US aluminum material flow network could be imports of aluminum billet.

For a scenario with given Φ and q, we can compute the model prediction for all nodal mass flows via:

$$z = (\mathbb{I} - \Phi^T)^{-1} q. \tag{4}$$

Subsequently, from the values of $\{z, \Phi, q\}$, other common MFA quantities of interest (QoIs) can be predicted, such as mass flows for each edge ($\phi_{ij} z_i$), and sums, products, and ratios of mass flows. We express these QoIs as a function, $G(\phi, q)$, where ϕ denotes a flattened vector containing all ϕ_{ij}'s.

2.2 | Bayesian parameter inference

Given an MFA model, the set of all unknown MFA parameters of interest $\theta \in \mathbb{R}^{n_\theta}$ and the collected data $y \in \mathbb{R}^{n_y}$ are treated as random variables and associated with a joint PDF $p(\theta, y)$. Here, we use θ to denote *all* MFA model parameters we are interested to learn, which may encompass ϕ, q, σ, as well as other parameters; y is the flattened set of collected and noisy MFA data records that correspond to the model prediction QoIs $G(\phi, q)$. Bayes' rule then directly follows from the axioms of probability, stating:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{5}$$

where $p(\theta)$ is the prior PDF representing the initial belief in the MFA parameters θ before having collected any data; $p(y|\theta)$ is the likelihood PDF (see Section 2.3); $p(\theta|y)$ is the posterior PDF representing the updated belief in the MFA parameters θ after having collected the data y; and $p(y)$ is the model evidence (marginal likelihood) and acts as a normalizing constant for the posterior PDF. Performing Bayesian parameter inference entails computing or characterizing the posterior $p(\theta|y)$ while accessing the likelihood and prior.

2.3 | Modeling the likelihood $p(y|\theta)$

The likelihood computes the probability of having collected MFA data y if the model parameters had the value equal to θ; that is, it provides a probabilistic measure on the mismatch between observation y and model prediction $G(\phi, q)$. There are many ways in which the collected data y may relate to $G(\phi, q)$. For example, the discrepancy may be viewed as an additive noise:

$$y_k = G_k(\phi, q) + \epsilon_k, \tag{6}$$

where k indicates the kth data component. Equation (6) is appropriate for MFA data where the error is insensitive to the scale of the measurement; for example, in the case of a mass sensor which has a sensitivity of $\pm$ 10 g across its measurement range. However, oftentimes the data error increases with the scale of the measurement and can be modeled as a relative noise in the form:

$$y_k = G_k(\phi, q)(1 + \epsilon_k). \tag{7}$$

In either case, modeling the noise ϵ_k as a Gaussian centered around zero is a sensible assumption describing our belief that the distribution of noise should typically be symmetric around zero without biasing to a particular direction. For its standard deviation σ_k, we do not presume its magnitude and will instead learn these values from the data. In the absence of information to the contrary, the noise associated with different pieces of data can be modeled as independent. Therefore, if adopting the relative error form in Equation (7) and with the unknown model parameters $\theta = \{q, \phi, \sigma\}$, the likelihood PDF is:

$$p(y|\theta) = p_\epsilon\left(\frac{y}{G(\phi, q)} - 1\right) = \prod_{k=1}^{n_y} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left\{-\frac{\left(\frac{y_k}{G_k(\phi,q)} - 1\right)^2}{2\sigma_k^2}\right\}. \tag{8}$$

2.4 | Expert prior elicitation for MFA

The goal of expert prior elicitation is to extract pertinent knowledge for θ from subject matter experts in a form that can used as a proper Bayesian prior PDF. There are two main challenges in expert prior elicitation. An expert does not typically have a preexisting quantification of her belief in the form of a PDF (Winkler, 1967). Therefore, the first challenge is how to elicit and synthesize a single expert's knowledge into a "quantified belief prior" (Winkler, 1967). Next, when there are multiple experts available, it is desirable to utilize all their opinions to have the prior capture the full diversity of background knowledge. However, Bayesian inference requires a single PDF for a parameter as the prior rather than multiple distributions from several experts. The second challenge is, therefore, how to combine and weight the beliefs of multiple experts into a single prior for each MFA variable of interest.

Expert prior elicitation has been widely applied in medicine (Azzolina et al., 2021) and has been used in environmental assessments to forecast future wind energy costs (Wiser et al., 2016) and regional climate change (Dessai et al., 2018). There has been some preliminary work on expert elicitation in MFA: Montangero and Belevi (2007) use expert elicitation to describe the uncertainty regarding the flow of nitrogen and phosphorus in a septic tank. In their work, multiple experts provide uncertainty information as quantiles and their opinions are combined using equal weights. Elsewhere, Mathieux and Brissaud (2010) conduct expert elicitation to understand where aluminum is being used in commercial vehicles. In their work, experts are brought together to determine collectively a single distribution for each estimate. In this work, we examine methods for eliciting, weighting, and aggregating multiple experts' beliefs under the Bayesian framework.

2.4.1 | Eliciting a prior from an expert

Prior elicitation is typically performed using surveys conducted either remotely (e.g., by mail or online), in-person, or via a hybrid format where the expert is assisted by telephone or video conference (Johnson et al., 2010). While the variable of interest θ can be multivariate (e.g., a joint distribution on all the allocation fractions leaving a node), eliciting multi-dimensional PDFs directly is very challenging; therefore, multivariate elicitation typically involves eliciting and then combining univariate marginal distributions (Daneshkhah & Oakley, 2010; O'Hagan et al., 2006). The common methods to elicit univariate PDFs are either a **variable interval method** or a **fixed interval method** (O'Hagan et al., 2006; Oakley, 2010). In a variable interval method, the expert provides estimates of the quantiles; for example, estimate a and b such that $\mathbb{P}(\theta \leq a) = 0.25$, $\mathbb{P}(a < \theta \leq b) = 0.5$, and $\mathbb{P}(b > \theta) = 0.25$ (Murphy & Winkler, 1974; Garthwaite et al., 2005). In fixed interval methods, the expert estimates the probability of θ within given fixed intervals (e.g., estimate $\mathbb{P}(a < \theta \leq b)$ for some given a and b) (O'Hagan, 1998). While it is unclear which set of methods yield more accurate representations of the expert's true belief (Abbas' findings contradicting those of Murphy and Winkler (Abbas et al., 2008; Murphy & Winkler, 1974)) it does appear that participants find the fixed interval method easier to complete (Abbas et al., 2008). Given that many MFA industry experts might be unfamiliar with statistical concepts such as quantiles, we recommend using the fixed interval method for MFA prior elicitation where possible; for example, for eliciting allocation fractions (ϕ) which are bounded within [0, 1]. Elsewhere, when eliciting external inflows (q) or data noise parameters (σ_k), there is no upper bound a priori so that it is appropriate to either ask the expert to specify an upper bound before defining the intervals used in the fixed interval method or else use a variable interval method (see Supporting Information Section 1).

2.4.2 | Prior aggregation from multiple experts

The typical methods for combining multiple experts' knowledge into a single proper prior PDF are behavioral aggregation and mathematical aggregation (O'Hagan et al., 2006) (see Supporting Information Section 2.4).

Behavioral aggregation

Experts collaborate to define agreed upon priors (O'Hagan et al., 2006). Thus, very "informed" experts have the chance to share their knowledge. However, it can be difficult to find a common time when all the experts are available and there are potential issues with strong personalities

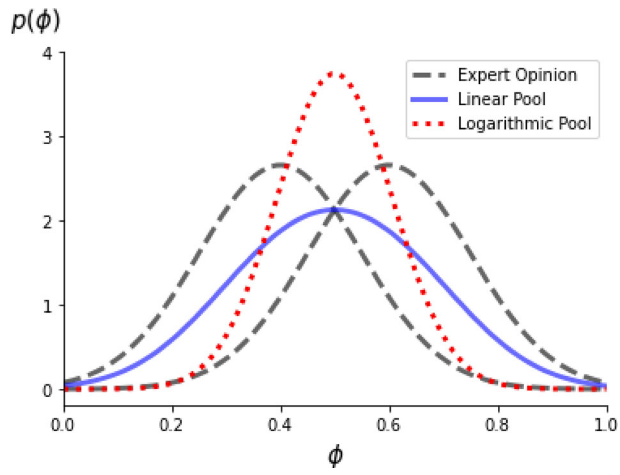


FIGURE 2 Linear versus logarithmic pooling of two equal weight elicited priors for an allocation fraction (ϕ). In general, the logarithmic pool result is more "concentrated" than the linear pool result (O'Hagan et al., 2006).

dominating the decision-making (O'Hagan et al., 2006), the risk of overconfidence in the group (Gigone & Hastie, 1993), some experts concealing their true views (Plous, 1993; Sniezek, 1992), group polarisation (Plous, 1993; Sniezek, 1992), and individuals with unique information being ineffective at sharing (Stasser & Titus, 1985). Therefore, methods such as the Delphi method (Linstone & Turoff, 1975; Rowe & Wright, 1999) and its variants (Degroot, 1974; Delbecq et al., 1986) have been proposed where direct interaction between the experts is restricted to prompt the experts to explain their views rather than relying on reputation or personality (O'Hagan et al., 2006); for example, experts may share their views anonymously, adjust their views based on the received information, and iterate until convergence on a distribution.

Mathematical aggregation

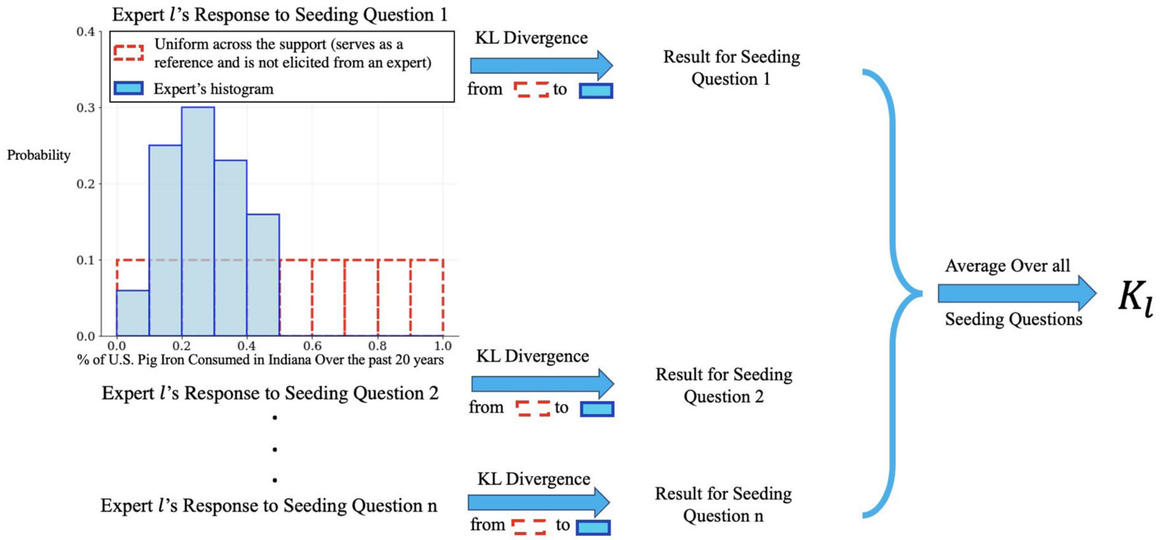
A distribution is elicited from each of the n_e experts independently, yielding n_e PDFs $\{p_1(\theta), \ldots, p_{n_e}(\theta)\}$. An aggregated PDF $p(\theta)$ is then calculated using either linear or logarithmic pooling (O'Hagan et al., 2006; Genest & Zidek, 1986; Clemen & Winkler, 1999):

$$\text{(linear)} \qquad p(\theta) = \sum_{\ell=1}^{n_e} w_\ell p_\ell(\theta), \qquad\qquad \text{(logarithmic)} \qquad p(\theta) = \frac{1}{Z} \prod_{\ell=1}^{n_e} p_\ell(\theta)^{w_\ell}, \tag{9}$$

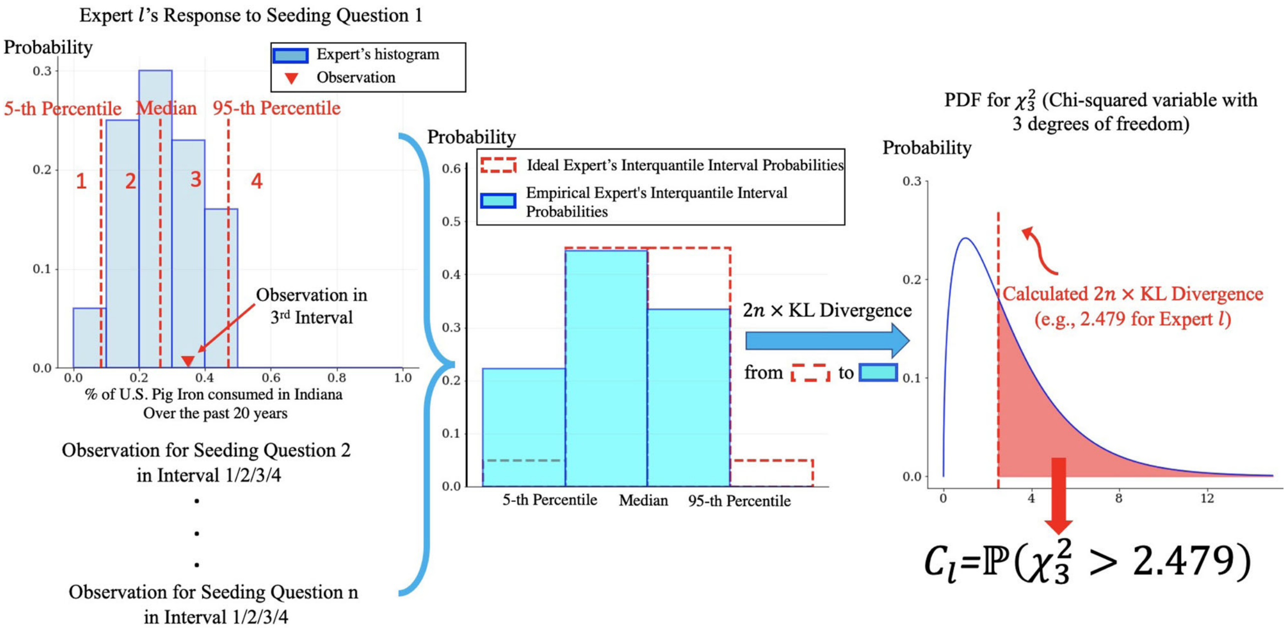
where w_ℓ is the weight associated with expert ℓ, and Z is a normalization constant. In logarithmic pooling, the resulting prior $p(\theta) = 0$ whenever any of the experts believe $p_\ell(\theta) = 0$. In contrast, the prior generated by linear pooling includes any value considered plausible by any of the experts and is therefore more conservative in terms of not ruling out experts' beliefs (see Figure 2).

Equal weights could be assigned to all experts such that $w_\ell = 1/n_e$; however, it is often desirable to allocate greater weighting to more informed experts, typically using either social network weighting or methods requiring questions on *seeding variables*, also referred to as *seeding questions* from hereon. Social network weighting usually assigns weights based on each expert's number of citations (Cooke et al., 2008) or a consensus among the experts on whose opinion should receive the most weight (Aspinall & Cooke, 2011). Such methods have been criticized for often excluding experts with predominately industry rather than academic experience and for resulting in prior PDFs with low accuracy (Cooke et al., 2008; Aspinall & Cooke, 2011; Colson & Cooke, 2018). Seeding questions assess each expert's expertise by comparing the experts' responses to collected seeding variable observations. One seeding variable example for a steel MFA study might be on the fraction of US pig iron consumed in Indiana within a certain time period. The experts' responses are compared to the observations from a credible source (e.g., USGS). Typically, Cooke's method (Cooke, 1991) (summarized in Figure 3) is used to convert expert seeding question responses to expert weights where, for expert ℓ, the weight $w_\ell \propto C_\ell K_\ell$ with C_ℓ being the calibration score and K_ℓ the information score. The calibration score measures the accuracy of the expert's responses, and the information score penalizes experts weakly informative (i.e., unsure) responses that approach the uniform distribution across the support. The Kullback–Leibler (KL) divergence is used in calculating both scores. The KL divergence is a measure of how two PDFs differ. It is non-symmetric and non-negative with a KL divergence being zero between two identical distributions, and a larger KL divergence implying a greater difference between two distributions (Kullback, 1997). For discrete variable X taking values in $\{1, 2, \ldots, m\}$, and two probability mass functions $P(x) = p_x$ and $Q(x) = q_x$, the KL divergence from Q to P is given as:

$$D_{\mathrm{KL}}(P||Q) = \sum_{i=1}^{m} p_i \ln\left(\frac{p_i}{q_i}\right). \tag{10}$$



(a) Calculating the information score for an expert.



(b) Calculating the calibration score for an expert.

FIGURE 3 (a) Each expert's information score K_ℓ is calculated as the Kullback–Leibler (KL) divergence from the expert's elicited histogram to the uniform distribution averaged across all seeding question responses. (b) Each expert's calibration score C_ℓ is calculated by first counting the number (fraction) of actual observations of the seeding questions in each inter-quantile interval across all seeding question responses from an expert, then computing the KL divergence from the ideal expert's inter-quantile interval probabilities to the empirical expert's inter-quantile interval probabilities, and lastly obtaining the likelihood ratio statistic with the corresponding p-value (i.e., $\mathbb{P}(\chi^2_3 > 2.479)$ in the above example) being the calibration score.

Each expert's information score K_ℓ is calculated as the KL divergence from the expert's elicited distribution to the uniform distribution, averaged across all the seeding questions (see Figure 3a). In order to calculate an expert's calibration score C_ℓ, each response to the seeding questions from expert ℓ is split into inter-quantile intervals. Typically, four inter-quantile intervals (three degrees of freedom) are adopted with a corresponding probability vector $Q = \{0.05, 0.45, 0.45, 0.05\}$. Each seeding variable observation is then compared to the inter-quantile intervals derived from the expert's response; for example, the USGS records that 34.9% of US pig iron was consumed in Indiana from 2002 to 2016, falling into the 50%–95% interval of the response from expert ℓ shown in Figure 3b. Then, let $P_\ell = \{p_1, p_2, p_3, p_4\}$ denote the fraction of all n seeding variable observations that fall into each of the four intervals elicited from expert ℓ. Cooke (1991) states that for a well-informed "ideal" expert (i.e., where seeding variable observations appear as independently drawn from a distribution consistent with the expert's quantiles) then P_ℓ tends to Q, and $D_{KL}(P||Q)$ tends to zero. Cooke defines the calibration score C_ℓ as the probability that a random variable following a Chi-square distribution (with three degrees of freedom if using four inter-quantile intervals) is greater than the likelihood ratio statistic ($2 \times n \times D_{\mathrm{KL}}(P||Q)$), see Figure 3b; therefore, if expert ℓ's knowledge differs from the seeding variable observations to a large extent, the associated KL divergence would be high and C_ℓ becomes small. While not discussed previously, we believe that Cooke's method of calculating C_ℓ is only appropriate when there is uncertainty in the seeding variable observation (see Supporting Information Section 2.5 for a detailed discussion). Eggstaff et al. (2014) state that there is no definitive minimum number of seeding questions; however, Cooke suggests that 8–10 seeding questions are sufficient for a substantial improvement compared to assigning equal weights (Cooke, 1991).

Once the expert weights are computed, they can then be used to carry out the mathematical aggregation in Equation (9) to complete the prior construction.

2.5 | Distributions for modeling the MFA priors

The next step is fitting the histograms elicited from the MFA experts to a family of parameterized PDFs. The distribution and corresponding hyperparameters are typically fitted and selected via a least squares procedure (McBride et al., 2012; R Core Team, 2017).

2.5.1 | Distributions for allocation fraction priors

Lupton and Allwood (2018) proposed using a Dirichlet distribution as the prior for the allocation fractions $\{\phi_{S,d_1}, \ldots, \phi_{S,d_S}\}$ from source node S, ensuring that all the allocation fractions remain in [0, 1] and sum to unity. The Dirichlet PDF for $\phi_{S,d_1}, \ldots, \phi_{S,d_S}$ given hyperparameters $\alpha_D = \{\alpha_{d_1}, \ldots, \alpha_{d_S}\}$ is

$$p(\phi_{S,d_1}, \ldots, \phi_{S,d_S} | \alpha_D) = \frac{\Gamma\left(\sum_{i=d_1}^{d_S} \alpha_i\right)}{\prod_{j=d_1}^{d_S} \Gamma(\alpha_j)} \prod_{k=d_1}^{d_S} \phi_{S,k}^{\alpha_k - 1} \tag{11}$$

and for each $\phi_{S,i}$, its marginal PDF follows a Beta distribution characterized by

$$p(\phi_{S,i} | \alpha_D) = \frac{\Gamma\left(\sum_{j=d_1}^{d_S} \alpha_j\right)}{\Gamma(\alpha_i)\Gamma\left(\sum_{k \neq i} \alpha_k\right)} \phi_{S,i}^{\alpha_i - 1} (1 - \phi_{S,i})^{(\sum_{l \neq i} \alpha_l) - 1}. \tag{12}$$

Another benefit of using a Dirichlet distribution is that eliciting each marginal distribution on $\phi_{S,i}$ (Equation 12) from an expert is sufficient to fully construct the Dirichlet joint distribution (Equation 11) (O'Hagan et al., 2006). After eliciting each marginal histogram of $\phi_{S,i}$ from the experts, we fit for the optimal Dirichlet hyperparameters α_D^* by minimizing the squared differences between the probability of each Beta marginal that corresponds to the Dirichlet joint distribution and the marginal weighted histograms with n_b intervals from the experts, summed across all d_S allocation fractions emanating from a source node S (Zapata-Vázquez et al., 2014):

$$\alpha_D^* = \underset{\alpha_D}{\operatorname{argmin}} \sum_{d_s=d_1}^{d_S} \sum_{i=1}^{n_b} \left[F(\theta_{d_s,i+1} | \alpha_D) - F(\theta_{d_s,i} | \alpha_D) - \mathbb{P}(\theta_{d_s,i}) \right]^2, \tag{13}$$

where F is the Beta CDF.

Other methods exist that may also be used to model the allocation fraction priors (Lupton & Allwood, 2018; Gelman et al., 1996). Softmax transformations offer extra flexibility compared to using Dirichlet priors; for example, they can incorporate a strong belief that $\phi_{S,d_1} = \phi_{S,d_2}$ while the relationship between other allocation fractions is unknown. However, the increased complexity of the procedure complicates the elicitation process because, for example, the CDF in Equation (13) cannot be evaluated in a closed form.

2.5.2 | Distributions for data noise parameter priors σ_k

Since σ_k is the standard deviation of Gaussian distributed ϵ_k, the prior must take positive support; that is, $p(\sigma_k < 0) = 0$. A common choice of prior for the standard deviation of a Gaussian distribution is the inverse gamma distribution, which enables an analytical evaluation of the posterior when the measured data is linear in the parameters of interest (Gelman, 2006). However, the inverse gamma and other commonly used positive support distributions such as the log-normal distribution place negligible probability of σ at regions close to zero, greatly reducing the possibility that the

MFA data is of high quality. We believe that generally in MFA there should be a moderate probability that the collected data is clean and high quality. Consequently, other prior distributions that place non-negligible probability at regions close to zero should be considered. Such distributions include the half-Cauchy distribution, uniform distribution, and truncated normal distribution.

2.5.3 | Distributions for mass inflow priors q

The mass inflows q to the network must be positive. Therefore, uniform and truncated normal distributions can be used. When there is little information about q, the upper bound can be set to a large value for both distributions. Alternatively, both the half-Cauchy and truncated normal distribution can be used without an upper bound, and set such that the shape of $p(q)$ is close to being flat.

Section 3 of the Supporting Information reviews the properties of different PDFs to aid selection of prior PDFs for data noise and mass inflow priors.

2.6 | Posterior sampling

Once the prior and likelihood are established, the Bayesian inference problem can be solved as stated in Equation (5), updating our knowledge about our model parameters through the posterior distribution. Attempting to compute the posterior PDF would entail evaluating the denominator (model evidence) in Bayes' rule: $p(y) = \int p(y|\theta)p(\theta)\, d\theta$, a task that is generally intractable to perform even numerically except for very low (e.g., < 3) dimensions of θ. Instead of computing the PDF, a major alternative strategy is to generate samples of θ from the posterior distribution. To that end, Markov Chain Monte Carlo (MCMC) algorithms (Gilks et al., 1996; Andrieu et al., 2003; Robert & Casella, 2004; Brooks et al., 2011) have become the predominant methods for computational Bayes in moderate θ dimensions (e.g., up to ~ 100), which iterates a Markov chain to generate samples that are consistent with the targeted posterior distribution. The more scalable MCMC methods include Hamiltonian Monte Carlo (HMC)-based samplers such as the No-U-Turn (NUTs) algorithm, which explore the parameter space efficiently leveraging the posterior gradient and Hamiltonian energy principles (Betancourt, 2018). However, HMC suffers from divergence in its time integration step when encountering neighborhoods of high posterior curvature (Betancourt, 2018; Livingstone et al., 2016); this difficulty was indeed observed in our study when incorporating the data noise parameters σ into the Bayesian inference. Therefore, we opt to use sequential Monte Carlo (SMC) (Doucet et al., 2001) to sample the posterior, which is based on the idea of iteratively re-weighing the samples using a tempered likelihood $[p(y|\theta)]^{\beta}$ at each stage where β is a tempering parameter that gradually increases from 0 to 1. We describe the SMC algorithm in Supporting Information Section 5.

3 | CASE STUDY ON THE US STEEL FLOW

The advances in the Bayesian inference approach to MFA discussed above are tested by mapping the US annual flow of steel, where we take the MFA network structure from Zhu et al.'s (2019) analysis of US steel flows in 2014. The case study demonstrates rigorous prior development based on expert elicitation and Bayesian inference to produce the posterior uncertainty for the MFA collected data noise and MFA flow parameters. All the data and code used in this case study are available online (see the Supporting Information).

In this case study, the parameters requiring prior formulation are the allocation fractions (ϕ), the external inputs (q), and the data noise standard deviation (σ) associated with data noise (ϵ).

3.1 | Constructing priors for allocation fractions (ϕ) and input flows (q)

Since the elicitation of informative priors for all MFA variables of interest can be time prohibitive, we only elicit expert priors for the upstream allocation fractions (ϕ) and external inputs (q), while using weakly informative priors elsewhere. Experts on US steel flows were identified by conducting a literature search on steel flows and recycling and by contacting US steel companies (e.g., US Steel and Nucor). All the experts had more than 5 years of working or research experience in the steel industry. Eight experts agreed to an emailed request to take part in the study (see Table 2 in the Supporting Information).

For prior construction, the experts independently completed surveys online that contained a total of 32 questions: 23 elicitation questions for allocation fractions associated with import, export, production, and consumption of ferrous raw materials, and an extra 9 seeding questions whose actual observations are taken from USGS.

At least one author was present online to answer any questions during survey completion. It took the experts between 25 and 80 min each to complete the survey. The fixed interval method was used to elicit the parameters. For allocation fractions ϕ, the support [0, 1] for ϕ was divided into

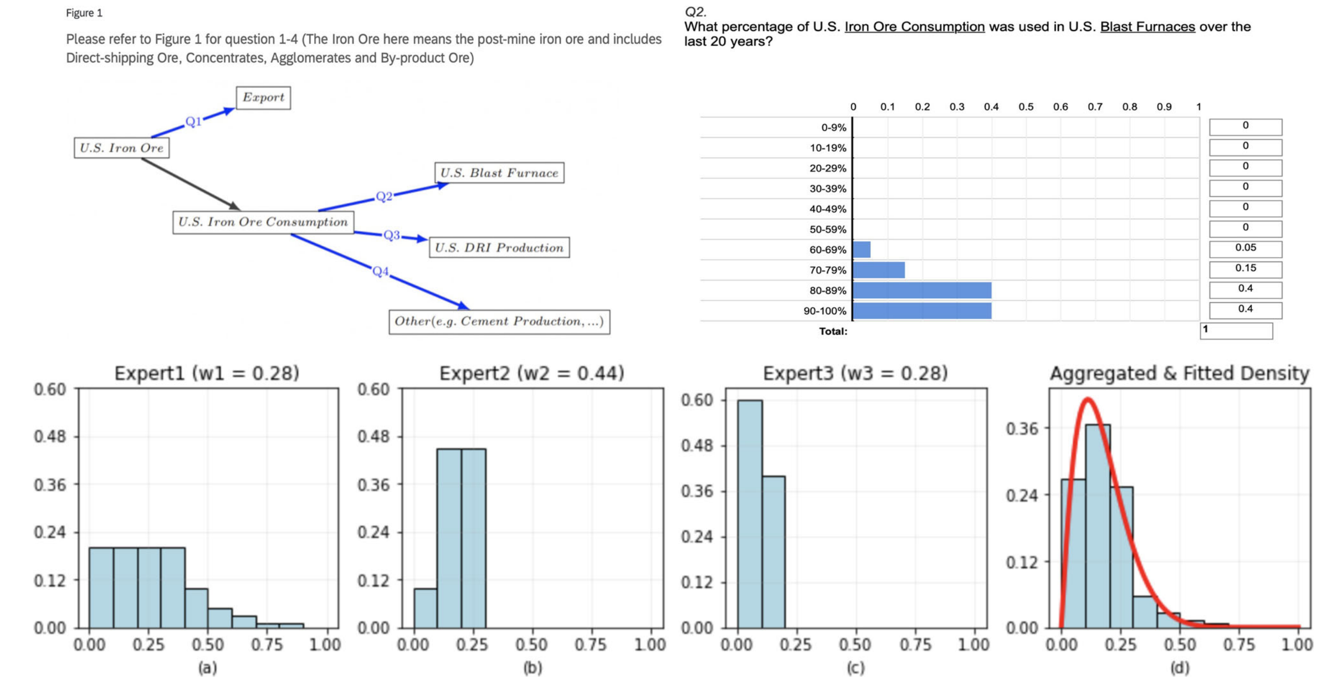


FIGURE 4 Prior elicitation required steel industry experts to complete an online Qualtrics Survey. The images above show an example question and expert response. The images below show an example calculation of an aggregated prior for the fraction of US iron ore that is exported. Only three expert responses are shown for the sake of concision. Their weights have been normalized accordingly to sum to unity.

10 equal-width intervals and Dirichlet hyperparameters were fit to the elicited experts' histograms as described in Section 2.4. For the external inputs, the expert first specifies the lower and upper bound, with the interval then divided into 10 equal-width intervals. Figure 4 shows an illustrative example of a survey question for eliciting an allocation fraction. An expert could enter the probability value for each interval in the box or else drag the bar across for the given interval until the summation of the bars is 1, otherwise the expert cannot continue to the next question.

Linear pooling was used to aggregate the responses from the multiple experts into a single proper prior PDF for each MFA variable. First, weights were assigned to each of the experts based on their responses to nine seeding questions using Cooke's method. A wide range of expert weighting values were obtained with the most informed expert having a weight of 0.299 and the least informed expert having a weight of less than 0.001. Figure 4 shows how the response of multiple steel experts are combined to form a single aggregated prior PDF for an MFA variable.

3.2 | Collecting MFA data records and constructing the prior for data noise standard deviation (σ)

After construction of the allocation fraction and external inflow priors, US steel flow MFA data were collected. We focus here on 2012 data although any year from the previous 20 years could also have been chosen. The steel flow data were collected from the United States Geological Survey (USGS) (United States Geological Survey, 2012a, 2012b, 2012c), World Steel Association (WSA) (World Steel, 2012) and Zhu et al. (2019). A complete record of all the collected MFA data is provided in Supporting Information Section 6.

The collected MFA data are published without accompanying uncertainty information while data error is inevitable. Subsequently, the noise for each piece of collected data is modeled as an independent relative error (see Equation 7) that follows a Gaussian distribution with zero mean and a standard deviation σ. Expert elicitation could be used to derive an informed prior for σ; for example, experts could be interviewed on the likely accuracy of USGS and WSA data. Further elicitation would significantly increase the surveying time. Therefore, in this study we avoid expert elicitation on the data noise parameters and instead the prior on σ is modeled as only "weakly informative" using a normal distribution truncated below zero and above 0.5 with hyperparameters set such that $\mathbb{P}(\sigma \leq 0.1)$ and $\mathbb{P}(\sigma \leq 0.3)$ are approximately 0.5 and 0.95, respectively. This imposes a reasonable probability that the data can be of high quality; for example, $\sigma \leq 0.1$.

3.3 | Case study results: US steel flow in 2012

The Bayesian inference is implemented using SMC in PyMC3 with the code adapted from Lupton and Allwood (2018). It takes approximately 17 h to generate 10,000 samples using an Intel(R) CoreTM i7-11800H CPU, 2.30 GHz. The prior and posterior results are shown in Figures 5 and 6,

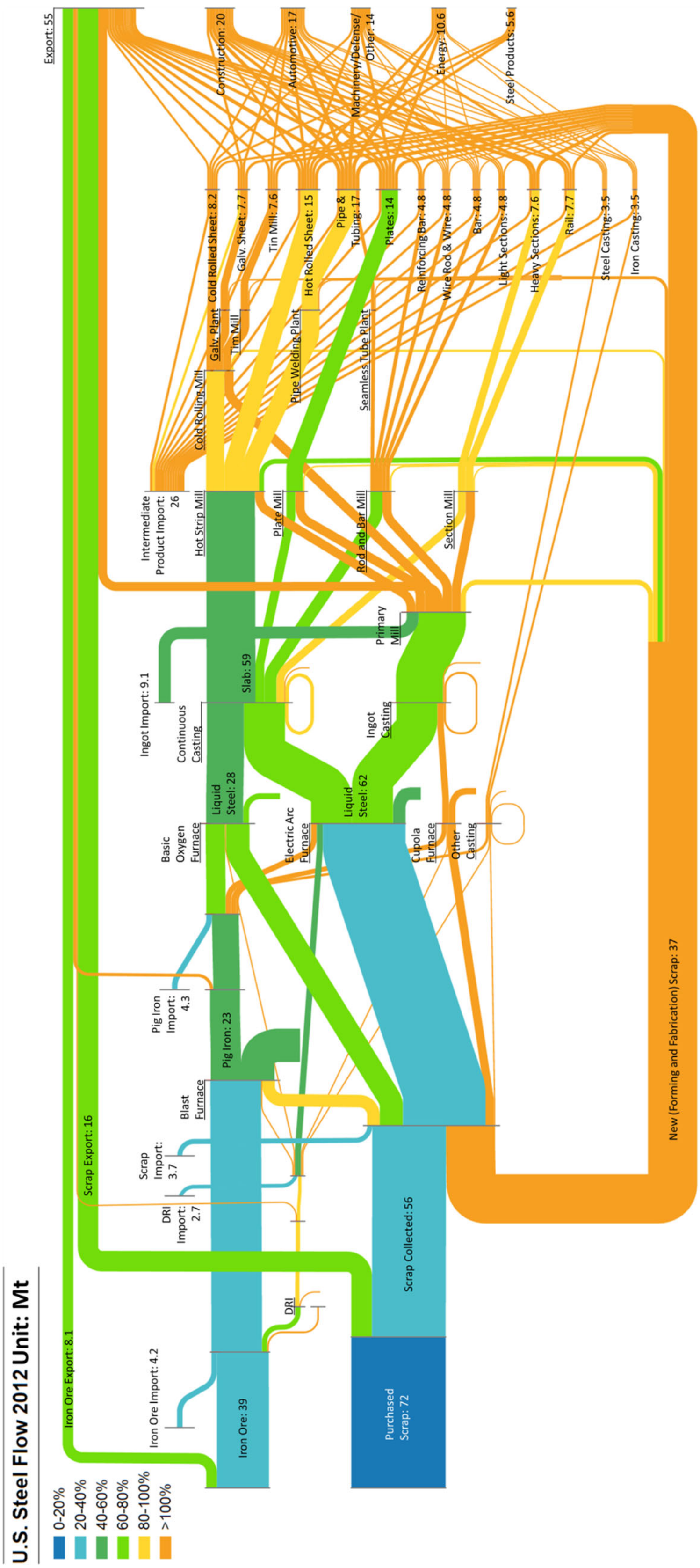


FIGURE 5 The prior for the US Steel flow in 2012. All numbers on the flows refer to the mean of the prior mass flow in units of million metric tons (Mt). The uncertainty percentages refer to the flow standard deviation as a percentage of the flow mean. All mass flows refer to steel except for the iron ore flows that include the non-iron mass (e.g., oxygen and gangue). Underlying data for Figure 5 are available in the Supporting Information.

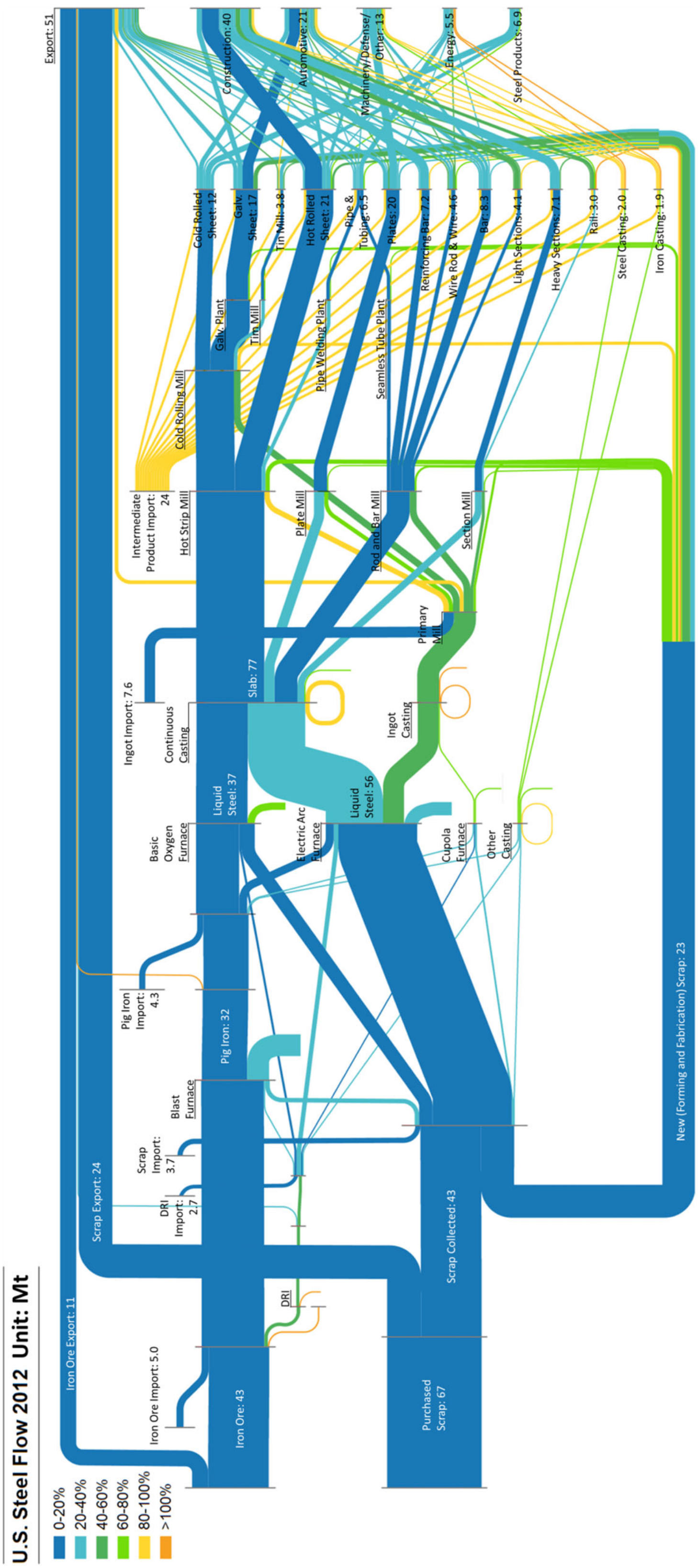


FIGURE 6 The posterior for the US Steel flow in 2012. All numbers on the flows refer to the mean of the posterior mass flow in units of million metric tons (Mt). The uncertainty percentages refer to the flow standard deviation as a percentage of the flow mean. All mass flows refer to steel except for the iron ore flows that include the non-iron mass (e.g., oxygen and gangue). Underlying data for Figure 6 are available in the Supporting Information.

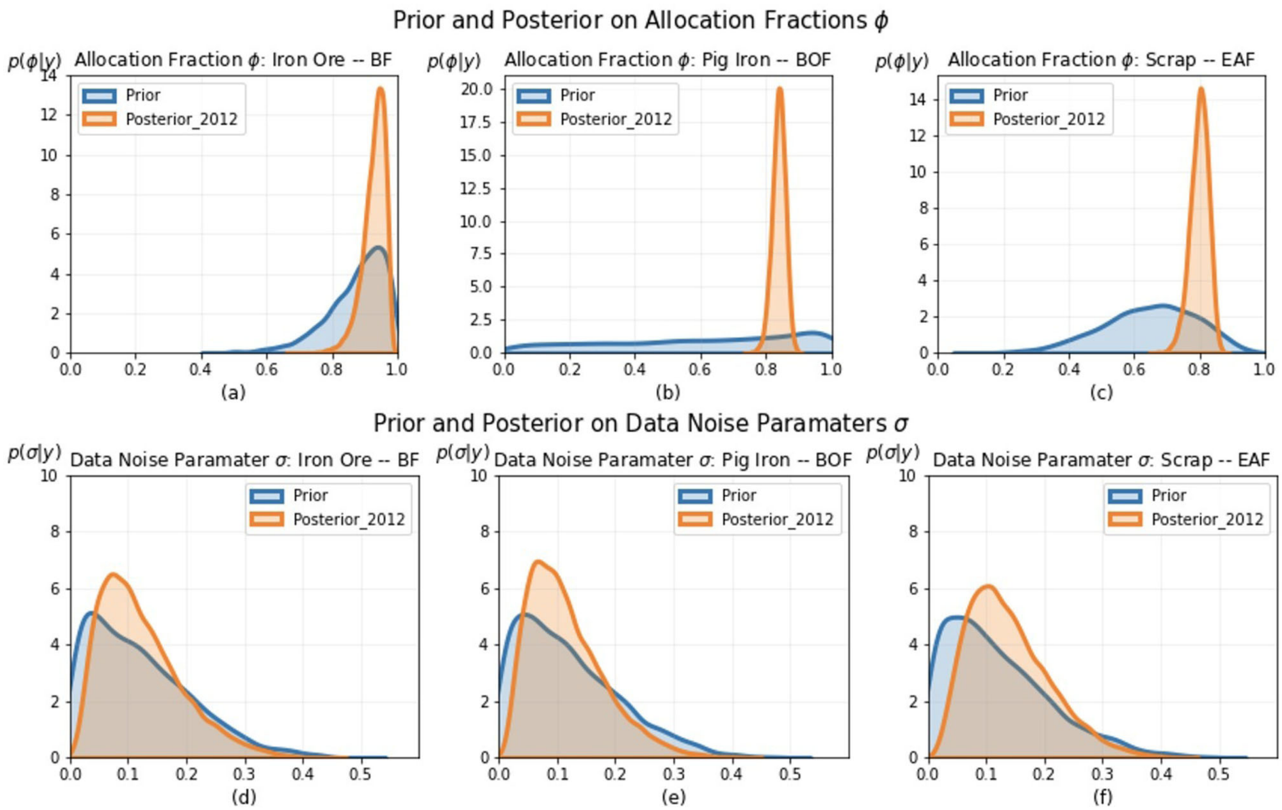


FIGURE 7 Examples of prior and posterior on allocation fractions (ϕ) and data noise parameters (σ) for annual US steel flow (2012).

respectively. The width of each line is proportional to the mean of the flow and the color indicates the uncertainty level, with a smaller relative uncertainty displayed in darker blue colors.

Figure 7 shows the prior and posterior distributions for ϕ and σ associated with three prominent upstream flows. There is a significant uncertainty reduction for all ϕ shown in Figure 7a–c. For example, despite a relatively flat prior for the fraction of (solid or liquid) pig iron consumed in the basic oxygen furnace (BOF), even with only 1 year of data the posterior is able to reflect that pig iron is mainly consumed in the BOF. Figure 7d–f presents the prior and posterior for σ. The posterior distribution for σ on "Iron Ore → BF" is more concentrated than the prior, indicating that the data noise can be learned from data. However, in comparison to ϕ, there is a lower uncertainty reduction in σ. In an effort to enhance data noise learning, we explored the use of multiple years of data to prompt a greater reduction in uncertainty.

3.4 | Enhanced learning using data collected from multiple years

One interpretation of the modest residual uncertainty for σ shown in Figure 7 is that there is still a low amount of information from data for inference if using data only from 2012. However, some of the ϕ are very likely to be similar across years. In addition, the measurement noise ϵ in different years is also possibly a realization from a similar distribution. If the allocation fractions ϕ and data noise parameter σ on regularly reported MFA data (e.g., the USGS data record on "Iron ore → BF") can be verified to be similar across a multi-year time period, then there is the potential to leverage multiple years' worth of MFA data to enhance the learning of the data noise. Therefore, we used Bayes factor analysis (see Supporting Information Section 4) to check and justify modeling the allocation fractions and data noise parameters as constant across five years worth of USGS and WSA data (2012–2016), allowing the inference to be rerun using these additional years' data.

Figure 8 shows the posteriors on ϕ and σ when utilizing one year (2012) versus five years worth (2012–2016) of data. Figure 8 shows a reduction in σ uncertainty when leveraging these extra data. For "Iron ore → BF,", utilizing 2012–2016 data reduces both the uncertainty on σ and its mean value. On the other hand, the posterior on "Pig Iron → BOF" shows a reduced uncertainty for σ but an increase in its mean. This could be because the data noise is indeed low in the "Iron ore → BF" data and high in the "Pig Iron → BOF" data. However, readers should note that the data noise calculated here reflects not only the collected MFA data quality but also any inadequacy in the modeling; for example, from the MFA network structure used to the assumption of constant data noise errors across the five years.

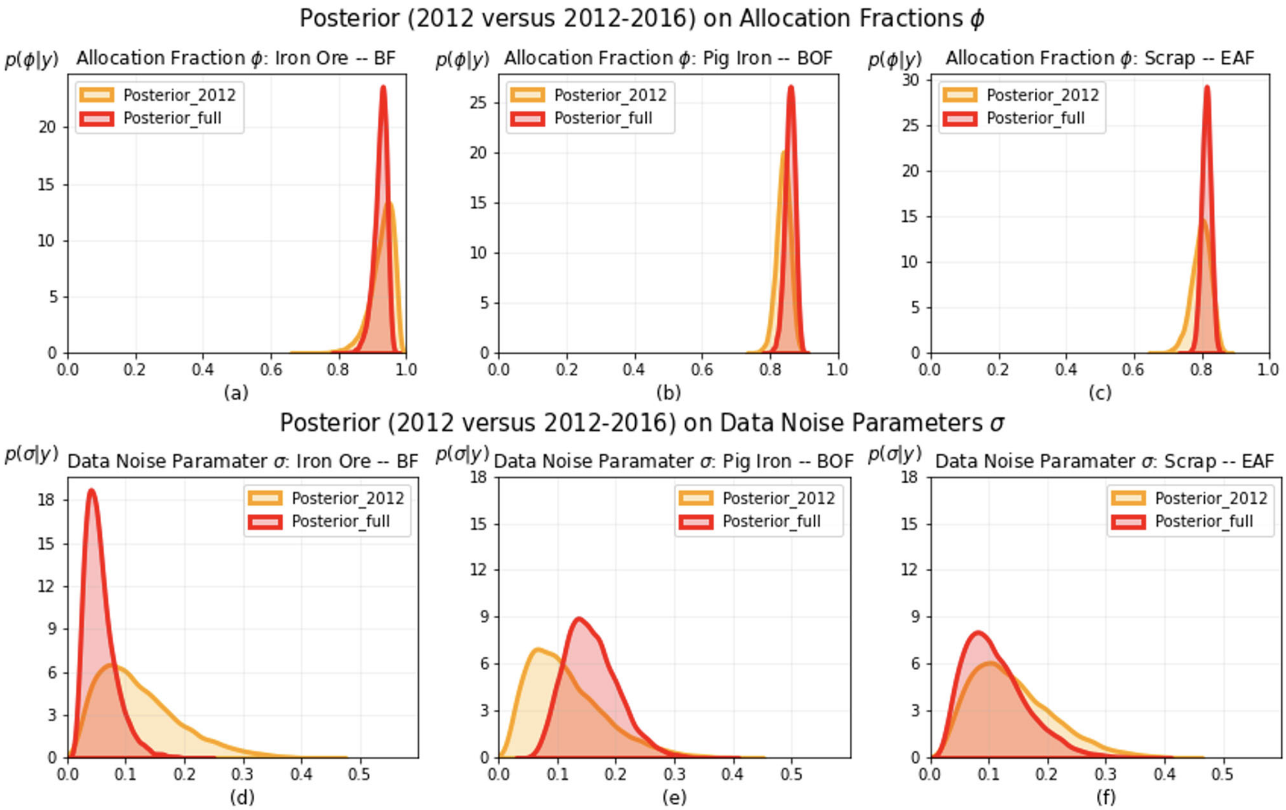


FIGURE 8 Examples of posteriors on allocation fractions (ϕ) and data noise parameters (σ) using data only from 2012 versus full data from 2012-2016.

4 | DISCUSSION AND CONCLUSIONS

Bayesian inference is a general probabilistic method for uncertainty quantification and updating as more data become available. Its potential uses in industrial ecology extend beyond MFA and include UQ for inventory and environmental impact data in life cycle assessments, as previously demonstrated in modeling of end-of-life waste management systems (Lo et al., 2005) and production using emerging manufacturing technology (Liao et al., 2023). This paper has introduced, adapted, and demonstrated the use of expert elicitation techniques to define proper prior distributions for Bayesian inference in MFA, and demonstrated how the noise level in collected MFA data may itself be learned from data. Below, we discuss the lessons learned from the case study on US steel flows, limitations of these approaches, and future work to advance the Bayesian approach in MFA.

4.1 | Using expert elicitation in MFA

Expert elicitation allows the use of informed priors in MFA and will result in a quick reduction in parametric uncertainties when combined with collected data. In cases where no or negligible MFA data can be collected, expert elicitation provides a statistically robust method of estimating material flows. Expert elicitation may also reveal model-form mistakes in the MFA structure that were not apparent beforehand.

A potential drawback is the time needed to find experts, develop well-posed elicitation questions, and process the responses. For the case study, the authors were concerned that each interview might take more than 1 h and that there would be significant confusion regarding our request to answer questions with histograms and PDFs rather than point values. However, most interviews were completed within 30 min and the authors only received approximately one survey clarification question per expert. Even experts who were unaccustomed to PDFs were comfortable completing the questions after reviewing the example question and solution we posted at the start of the survey. There was one case where the sum of upper bound elicited allocation fractions from a single source node was less than unity. While this does not prevent hyperparameter fitting to a Dirichlet PDF it does suggest that either the expert made a mistake or that the expert thought that the model structure was incorrect, believing that there should be another destination node from that source node. A more sophisticated elicitation procedure could ask experts to evaluate the probability of candidate MFA model structures in addition to MFA parametric values. Overall, the fixed interval method used in the case study was found to be an effective and quick method of eliciting allocation fraction and external inflow priors.

In the case study, a single weight was assigned to each expert. A potential problem is that the expert might not be uniformly informed across the domain of interest; for example, an expert might be an authority on steel recycling but relatively uninformed on primary steelmaking. In the case study, it was possibly a case of non-uniform expertise that affected the derivation of the aggregated prior for the allocation fraction from (liquid and solid) pig iron to the BOF. The expert who received the highest weighting gave a very different answer from the other experts in response to the elicitation question on "Pig Iron → BOF," indicating that less than 50% of the iron flows into the BOF. The high weighting given to this expert meant that the aggregated prior for pig iron to BOF was relatively uninformed (flat) despite the other seven experts responding that the majority of pig iron flows into the BOF (see Figure 7). One option to avoid such problems would be to assign multiple weights to each expert corresponding to different areas of expertise; however, that would require the development and asking of more seeding questions tailored to those different areas of expertise.

A source of confusion for one of the experts was whether the survey was asking for uncertainty on a given variable (e.g., $\theta = \frac{\text{Amount of iron ore exported}}{\text{Amount of iron ore produced}}$) or its variability over space and time; for example, the histogram of the $\frac{\text{Amount of iron ore exported}}{\text{Amount of iron ore produced}}$ for each month in a year or across regions. The survey was eliciting the former (uncertainty instead of variability) and we found it useful to clarify that we were seeking uncertainty by demonstrating a "toy" example and also using equations to define the variable of interest.

4.2 | Learning the MFA variables and data noise parameters

The uncertainty reduction in the US steel mass flows between the prior and posterior is shown by the increase in dark blue colors between Figures 5 and 6. The ability to quantify and reduce these uncertainties in a principled manner using Bayesian inference is appealing as it can lead to more informed decision- and policy-making. For example, the potential effects of import tariffs are clearer in upstream production (where imports are shaded blue in Figure 6) than in downstream fabrication where planning for the consequences is not as easy. Figure 6 also confirms Zhu et al.'s suggestion that US production of reinforcing bar (which typically acts as a sink for steel recycling impurities) is small at less than 25% of construction demand; therefore, improved scrap separation and refining technologies will be needed if more US end-of-life scrap is to be diverted away from export and recycled domestically (Zhu et al., 2019). Elsewhere, the benefits of deploying a new, more energy-efficient mill technology can be calculated with greater confidence for the cold rolling mill (calculated to have processed an expected 32.3 Mt in 2012 with a standard deviation of 3.27 Mt) compared to the primary mill (calculated to have processed a smaller expected value of 23.1 Mt in 2012 but with a larger standard deviation of 8.69 Mt).

A new MFA approach explored in this article has been to incorporate data noise parameters as random variables. The allocation fractions and the data noise parameters are then learned simultaneously with collected MFA data. Intuitively it is harder to achieve great uncertainty reductions using this method because the same amount of data is used to provide information on more parameters, even though it is a more honest representation if we do not know the true data noise. To combat this, we used Bayes factor analysis to verify the existence of time-invariant data noise parameters, allowing us to incorporate multiple years worth of data for greater uncertainty reductions.

The speed advantages of the Bayesian approach are from the easy updateability of the results. When newly acquired data become available, Bayesian updating requires only the uncertainty results (posteriors) of the previous analysis, and then performing of the posterior sampling step in Section 2.6. In contrast, methods such as least squares reconciliation require re-running an analysis from scratch that includes the original plus newly acquired data sets. As observed elsewhere (Lupton and Allwood, 2018), a potential drawback of the Bayesian approach to MFA is the computational cost of the Monte Carlo-based algorithms. Modeling the data noise parameters as random variables significantly increased the stochastic dimension of the problem and in turn computational cost, from 3 h per run of the python script if the data noise parameters are prescribed as constants, to 17 h per run of the python script when the data noise parameters are modeled as random variables. Therefore, there is a trade-off between the amount of bias and the computational cost. However, the computational speed can be increased by, for example, using multi-core processors to run algorithms that can be parallelized or applying approximate Bayesian techniques such as variational inference (Blei et al., 2017).

4.3 | Conclusions and future work

The Bayesian framework provides a mathematically rigorous method of quantifying and then updating the uncertainty in MFA, supporting informed decision- and policy-making. This paper has introduced, adapted, and demonstrated the use of expert elicitation techniques to define proper prior distributions for MFA and illustrated how the noise level in collected MFA data may itself be learned from data. Files have been made available (via the Supporting Information) to help readers apply these methods, including templates for conducting expert elicitation surveys, calculating expert weights and aggregating priors, and performing Bayesian inference using MFA priors and collected data.

The Bayesian approach to MFA provides a mathematically principled procedure to incorporate expert knowledge alongside sparse, noisy, and often incomplete data records: a data-informed model learning approach. We plan to investigate how this model-and-data relationship can be leveraged to create intelligent data acquisition strategies for seeking out the most informative data that can tell us what the MFA structure should look like, and what its parameters values are. These remain important challenges in MFA and can be approached by combining the use of Bayesian inference for uncertainty quantification early in the MFA exercise with the principles of Bayesian experimental design (Chaloner & Verdinelli, 1995; Müller, 2005).

ACKNOWLEDGMENTS

The authors would like to thank all the steel experts interviewed for this study. This material is based upon work supported by the National Science Foundation under Grant No. #2040013.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supporting information of this article.

ORCID

Daniel Cooper https://orcid.org/0000-0003-2903-0468

REFERENCES

Abbas, A., Budescu, D., Yu, H.-T., & Haggerty, R. (2008). A comparison of two probability encoding methods: Fixed probability vs. fixed variable values. *Decision Analysis*, *5*, 190–202. https://doi.org/10.1287/deca.1080.0126

Andrieu, C., de Freitas, N., Doucet, A., & Jordan, M. I. (2003). An introduction to MCMC for machine learning. *Machine Learning*, *50*, 5–43. https://doi.org/10.1023/A:1020281327116

Aspinall, W. P., & Cooke, R. M. (2011). Quantifying scientific uncertainty from expert judgement elicitation. *Risk and Uncertainty Assessment for Natural Hazards* (pp. 64–99). Cambridge University Press. https://doi.org/10.1017/CBO9781139047562.005

Azzolina, D., Berchialla, P., Gregori, D., & Baldi, I. (2021). Prior Elicitation for Use in Clinical Trial Design and Analysis: A Literature Review. *International Journal of Environmental Research and Public Health*, *18*(4), 1833. https://doi.org/10.3390/ijerph18041833

Bader, H.-P., Scheidegger, R., Wittmer, D., & Lichtensteiger, T. (2011). Copper flows in buildings, infrastructure and mobiles: A dynamic model and its application to switzerland. *Clean Technologies and Environmental Policy*, *13*, 87–101. https://doi.org/10.1007/s10098-010-0278-4

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer Series in Statistics. Springer. https://doi.org/10.1007/978-1-4757-4286-2

Bertsekas, D. P., & Tsitsiklis, J. N. (2008). *Introduction to Probability*. Athena Scientific Optimization and Computation Series. https://ocw.mit.edu/resources/res-6-012-introduction-to-probability-spring-2018/part-i-the-fundamentals/MITRES_6_012S18_Textbook.pdf

Betancourt, M. (2018). A Conceptual Introduction to Hamiltonian Monte Carlo. arXiv:1701.02434. https://doi.org/10.48550/arXiv.1701.02434

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877. https://doi.org/10.1080/01621459.2017.1285773

Bonnin, M., Azzaro-Pantel, C., Pibouleau, L., Domenech, S., & Villeneuve, J. (2013). Development and validation of a dynamic material flow analysis model for French copper cycle. *Chemical Engineering Research and Design*, *91*(8), 1390–1402.

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (Eds.). (2011). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC. https://doi.org/10.1201/b10905

Cencic, O. (2016). Nonlinear data reconciliation in material flow analysis with software stan. *Sustainable Environment Research*, *26*, 291–298. https://doi.org/10.1016/j.serj.2016.06.002

Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, *10*(3), 273–304. https://doi.org/10.1214/ss/1177009939

Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, *19*(2), 187–203. https://doi.org/10.1111/j.1539-6924.1999.tb00399.x

Colson, A. R., & Cooke, R. M. (2018). Expert elicitation: Using the classical model to validate experts' judgments. *Review of Environmental Economics and Policy*, *12*(1), 113–132.

Cooke, R. M. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. Oxford University Press.

Cooke, R. M., ElSaadany, S., & Huang, X. (2008). On the performance of social network and likelihood-based expert weighting schemes. *Reliability Engineering & System Safety*, *93*(5), 745–756. https://doi.org/10.1016/j.ress.2007.03.017

Cullen, J., & Cooper, D. (2022). Material flows and uncertainty. *Annual Review of Materials Research*, *52*(1), 525–559. https://doi.org/10.1146/annurev-matsci-070218-125903

Cullen, J. M., & Allwood, J. M. (2013). Mapping the global flow of aluminum: From liquid aluminum to end-use goods. *Environmental Science & Technology*, *47*(7), 3057–3064. https://doi.org/10.1021/es304256s

Cullen, J. M., Allwood, J. M., & Bambach, M. D. (2012). Mapping the global flow of steel: From steelmaking to end-use goods. *Environmental Science & Technology*, *46*(24), 13048–13055. https://doi.org/10.1021/es302433p

Daneshkhah, A., & Oakley, J. E. (2010). Eliciting multivariate probability distributions. In K. Böcker (Ed.), *Rethinking risk measurement and reporting* (pp. 179–204). Risk Books. https://doi.org/10.4135/9788132114062.n2

Degroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121. https://doi.org/10.1080/01621459.1974.10480137

Delbecq, A., Ven, A., & Gustafson, D. (1986). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Green Briar Press.

Dessai, S., Bhave, A., Birch, C., Conway, D., Garcia-Carreras, L., Gosling, J. P., Mittal, N., & Stainforth, D. (2018). Building narratives to characterise uncertainty in regional climate change through expert elicitation. *Environmental Research Letters*, *13*(7), 074005. https://doi.org/10.1088/1748-9326/aabcdd

Doucet, A., de Freitas, N., & Gordon, N. (2001). An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice* (pp. 3–14). Springer.

Eggstaff, J., Mazzuchi, T., & Sarkani, S. (2014). The effect of the number of seed variables on the performance of Cooke's classical model. *Reliability Engineering & System Safety*, *121*, 72–82. https://doi.org/10.1016/j.ress.2013.07.015

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, *100*(470), 680–701. https://doi.org/10.1198/016214505000000105

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. https://doi.org/10.1214/06-BA117A

Gelman, A., Bois, F., & Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, *91*(436), 1400–1412. https://doi.org/10.1080/01621459.1996.10476708

Genest, C., & Zidek, J. V. (1986). Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, *1*(1), 114–135.

Gigone, D., & Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology*, *65*, 959–974.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in practice*. Chapman & Hall.

Gottschalk, F., Scholz, R. W., & Nowack, B. (2010). Probabilistic material flow modeling for assessing the environmental exposure to compounds: Methodology and an application to engineered nano-TiO_2 particles. *Environmental Modelling & Software*, *25*(3), 320–332. https://doi.org/10.1016/j.envsoft.2009.08.011

Gottschalk, F., Sonderer, T., Scholz, R. W., & Nowack, B. (2010). Possibilities and limitations of modeling environmental exposure to engineered nanomaterials by probabilistic material flow analysis. *Environmental Toxicology and Chemistry*, *29*(5), 1036–1048. https://doi.org/10.1002/etc.135

Graedel, T. E. (2019). Material flow analysis from origin to evolution. *Environmental Science & Technology*, *53*(21), 12188–12196. https://doi.org/10.1021/acs.est.9b03413

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790423

Johnson, S. R., Tomlinson, G., Hawker, G. A., Granton, J. T., & Feldman, B. M. (2010). Methods to elicit beliefs for Bayesian priors: A systematic review. *Journal of Clinical Epidemiology*, *63*(4), 355–69.

Kopec, G. M., Allwood, J. M., Cullen, J. M., & Ralph, D. (2016). A general nonlinear least squares data reconciliation and estimation method for material flow analysis. *Journal of Industrial Ecology*, *20*(5), 1038–1049. https://doi.org/10.1111/jiec.12344

Kullback, S. (1997). *Information theory and statistics*. Courier Corporation.

Liao, J., Huan, X., Haapala, K., & Cooper, D. (2023). A Bayesian approach to modeling unit manufacturing process environmental impacts using limited data with case studies on laser powder bed fusion cumulative energy demand. Accepted for publication in *Procedia CIRP*.

Linstone, H., & Turoff, M. (1975). The Delphi method: Techniques and applications. *Technometrics*, *18*, https://doi.org/10.2307/3150755

Livingstone, S., Betancourt, M., Byrne, S., & Girolami, M. (2016). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli*, *25*, https://doi.org/10.3150/18-BEJ1083

Lloyd, S. M., & Ries, R. (2007). Characterizing, propagating, and analyzing uncertainty in life-cycle assessment: A survey of quantitative approaches. *Journal of Industrial Ecology*, *11*(1), 161–179. https://doi.org/10.1162/jiec.2007.1136

Lo, S.-C., Ma, H.-W., & Lo, S.-L. (2005). Quantifying and reducing uncertainty in life cycle assessment using the Bayesian monte carlo method. *Science of The Total Environment*, *340*(1-3), 23–33.

Lupton, R. C., & Allwood, J. M. (2018). Incremental material flow analysis with Bayesian inference. *Journal of Industrial Ecology*, *22*(6), 1352–1364. https://doi.org/10.1111/jiec.12698

Mathieux, F., & Brissaud, D. (2010). End-of-life product-specific material flow analysis. Application to aluminum coming from end-of-life commercial vehicles in Europe. *Resources, Conservation and Recycling*, *55*(2), 92–105. https://doi.org/10.1016/j.resconrec.2010.07.006

McBride, M., Garnett, S., Szabo, J., Burbidge, A., Butchart, S., Christidis, L., Dutson, G., Ford, H., Loyn, R., Watson, D., & Burgman, M. (2012). Structured elicitation of expert judgments for threatened species assessment: A case study on a continental scale using email. *Methods in Ecology and Evolution*, *3*, 906–920. https://doi.org/10.1111/j.2041-210X.2012.00221.x

Meylan, G., Reck, B., Rechberger, H., Graedel, T., & Schwab, O. (2017). Assessing the reliability of MFA results: The cases of rhenium, gallium, and germanium in the US economy. *Environmental Science & Technology*, *51*, 11839–11847. https://doi.org/10.1021/acs.est.7b03086

Montangero, A., & Belevi, H. (2007). Assessing nutrient flows in septic tanks by eliciting expert judgement: A promising method in the context of developing countries. *Water Research*, *41*(5), 1052–1064. https://doi.org/10.1016/j.watres.2006.10.036

Moyé, L. A. (2008). Bayesians in clinical trials: Asleep at the switch. *Statistics in Medicine*, *27*(4), 469–482. https://doi.org/10.1002/sim.2928

Müller, P. (2005). Simulation based optimal design. *Handbook of Statistics*, *25*, 509–518. https://doi.org/10.1016/S0169-7161(05)25017-4

Murphy, A. H., & Winkler, R. L. (1974). Credible interval temperature forecasting: Some experimental results. *Monthly Weather Review*, *102*(11), 784–794. https://doi.org/10.1175/1520-0493(1974)102⟨0784:CITFSE⟩2.0.CO;2

Nga, D., Anh Duc, T., & Kei, N. (2014). Modification of uncertainty analysis in adapted material flow analysis: Case study of nitrogen flows in the Day-Nhue River Basin, Vietnam. *Resources, Conservation and Recycling*, *88*, 67–75. https://doi.org/10.1016/j.resconrec.2014.04.006

Oakley, J. E. (2010). Eliciting univariate probability distributions. In K. Böcker (Ed.), *Rethinking risk measurement and reporting* (pp. 155–178). Risk Books. https://doi.org/10.1590/1806-9126-RBEF-2021-0424

O'Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *The Statistician*, *47*, 21–35.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain judgements: Eliciting experts' probabilities*. John Wiley & Sons, Ltd. https://doi.org/10.1002/0470033312

Plous, S. (1993). *The psychology of judgment and decision making*. Mcgraw-Hill Book Company.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Robert, C. P., & Casella, G. (2004). *Monte Carlo statistical Methods*. Springer. https://doi.org/10.1007/978-1-4757-4145-2

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting*, *15*(4), 353–375. https://doi.org/10.1016/S0169-2070(99)00018-7

Schwab, O., & Rechberger, H. (2018). Information content, complexity, and uncertainty in material flow analysis. *Journal of Industrial Ecology*, *22*(2), 263–274. https://doi.org/10.1111/jiec.12572

Sniezek, J. A. (1992). Groups under uncertainty: An examination of confidence in group decision making. *Organizational Behavior and Human Decision Processes*, *52*(1), 124–155. https://doi.org/10.1016/0749-5978(92)90048-C

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*, 1467–1478.

United Nations Comtrade Database (2012). International trade statistics yearbook. https://comtradeplus.un.org/Publication/ITSY

United States Geological Survey (2012a). Iron and steel. https://www.usgs.gov/centers/national-minerals-information-center/iron-and-steel-statistics-and-information

United States Geological Survey (2012b). Iron and steel scrap. https://www.usgs.gov/centers/national-minerals-information-center/iron-and-steel-scrap-statistics-and-information

United States Geological Survey (2012c). Iron ore. https://www.usgs.gov/centers/national-minerals-information-center/iron-ore-statistics-and-information

Von Toussaint, U. (2011). Bayesian inference in physics. *Reviews of Modern Physics*, *83*, 943–999. https://doi.org/10.1103/RevModPhys.83.943

Wang, D., & Romagnoli, J. A. (2003). A framework for robust data reconciliation based on a generalized objective function. *Industrial & Engineering Chemistry Research*, *42*(13), 3075–3084. https://doi.org/10.1021/ie0206655

Winkler, R. L. (1967). The assessment of prior distributions in Bayesian analysis. *Journal of the American Statistical Association*, *62*(319), 776–800. https://doi.org/10.1080/01621459.1967.10500894

Wiser, R., Jenni, K., Seel, J., Baker, E., Hand, M., Lantz, E., & Smith, A. (2016). Expert elicitation survey on future wind energy costs. *Nature Energy*, *1*, 16135. https://doi.org/10.1038/nenergy.2016.135

World Steel (2012). Steel statistical yearbook. https://worldsteel.org/wp-content/uploads/Steel-Statistical-Yearbook-2012.pdf

Zapata-Vázquez, R., O'Hagan, A., & Bastos, L. (2014). Eliciting expert judgements about a set of proportions. *Journal of Applied Statistics*, *41*, 1919–1933. https://doi.org/10.1080/02664763.2014.898131

Zhu, Y., Syndergaard, K., & Cooper, D. R. (2019). Mapping the annual flow of steel in the United States. *Environmental Science & Technology*, *53*(19), 11260–11268. https://doi.org/10.1021/acs.est.9b01016

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dong, J., Liao, J., Huan, X., & Cooper, D. (2023). Expert elicitation and data noise learning for material flow analysis using Bayesian inference. *Journal of Industrial Ecology*, *27*, 1105–1122. https://doi.org/10.1111/jiec.13399