Uncertainty Handling in Data Pertaining to Material Flow Analysis

D. Dubois, H. Fargier, D. Guyonnet

IRIT - Universite Paul Sabatier, Toulouse, France
BRGM, Orleans
dubois@irit.fr, fargier@irit.fr, d.guyonnet@brgm.fr

projet ANR ECOTECH ANR-11-ECOT-0002 ASTER

Introduction

Material flow analysis

Material flow analysis consists in calculating the quantities of a certain product transiting a network of local entities referred to as processes, considering input and output flows and including the presence of storage.

Collected data often conflicts with the mathematical model because data is more or less erroneous

Data reconciliation

In MFA it consists in minimally modifying, measured or estimated quantities so as to respect the rules of mass balance

Material flow analysis

A network of processes with input and output flows + storage.

Balance equation for a single process :

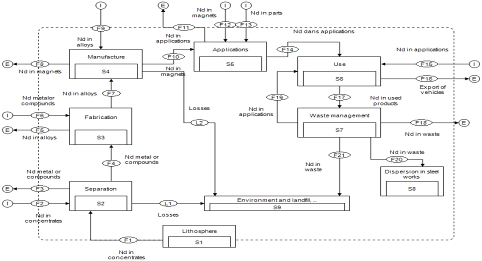
$$\sum_{i=1}^{n} IN_i = \sum_{j=1}^{m} OUT_j + \Delta s$$

where Δs is the amount of stock variation.

In order to balance the mass flows in a given network equations that model the flows, data $IN_i, OUT_j, \Delta s$ must be modified

Example of Network of Rare Earth flows

Neodymium in permanent magnets



Data Reconciliation

Adjusting data to comply with the equations of a model (mass balance in the case of MFA).

- A problem existing already in the early times of statistical science (XVIIIth century).
- The usual approach is : **least squares** (Legendre, 1805 ; Gauss...)

Outline of the talk

- analyse the pros and the cons of this approach
- present an alternative approach based on possibility theory, that takes into account uncertainty of data in a more explicit way
- Show differences and connections between the two approaches.

Usual Formulation of the Problem

Many data reconciliation packages assume

- Estimates of known quantities are in the form of mean values and st. deviations of normal random variables (some quantities are unknown).
- But the measured estimates (mean values) are generally not solutions to the balance flow problem.
- How to modify the measured estimates so as to satisfy the flow equations ?

Weighted Least Squares

Find x minimizing $\sum_{i=1}^{k} \frac{(x_i - \hat{x}_i)^2}{(\sigma_i)^2}$ under flow balance constraints

(a weighted Euclidean distance to the estimated data)

Reconciliation of Variances

The balance equations being linear, the reconciled vector of flow values is known to be a linear transformation of the measured values,

- The distribution of the reconciled values is obtained by transforming the joint pdf of the data via the linear transformation.
- If the data have a multivariate normal distribution, so have the reconciled data.
- The covariance matrix of the reconciliated data is then easily derived from the one of the original data.

Some comments on the least squares approach

Often justified by assuming the measured flows are random variables

Probabilistic model assumed with finite variance
Central Limit Theorem → Measurement errors follow a Gaussian law (with diagonal covariance matrix)
maximum likelihood principle + Normal distributions → Least squares are ideal

- This is not in agreement with the history of statistics :
  - Least squares are older than the CLT (Legendre 1805, Gauss, before).
  - Original justifications : Euclidean distance, ease of analytical computation, intuitive appeal of the average.
- Least squares make sense for numerous homogeneous noisy measurement results of the same quantity.

Uncertainty and probability

Uncertainty is pervading human activities, and especially decision.

- **What is uncertainty** : not knowing if an event of interest will (has) occur(red), if a state of facts is true.
- **Sources of uncertainty**
  - *Variability* : repeatable unpredictably changing observations (measurements)
  - *Lack of information* : imprecision inherent to human knowledge (non-probabilist)
  - *Too much information* : conflicting testimonies (contradictions)
- **Formal representations**
  - *Variability* : probability distributions (precise random observations)
  - *Imprecision* : intervals with confidence levels (imprecise coherent testimonies)
  - Too much information : NEED TO RECONCILE

Data in MFA of rare earths

Specificity of the MFA data

data, pertaining to the material flow of rare earths

- are seldom precise, but seldom random :
- scarce data about many material flows

- data have heterogeneous origins
  - Subjective estimations from experts
  - Data coming from documents of several nature, sometimes old.
  - Quantities for mixes of several products instead of single ones
- The estimated values clearly do not stem from a Gaussian random process.
- Even if they do, not enough information to isolate a distribution

Principles of a reconciliation methodology

Minimizing the distance between a solution to the MFA and the available data

- For defining reconciled values it sounds more convincing to try and justify the choice of a distance for error minimization than to appeal to abstract probability principles.
- Minimizing a quadratic deviation may incur a significant deviation of some x_i with respect to $\hat{x}_i$ (outliers),
- In the case of expert information one may wish to stay as close as possible to all relevant data

the choice depends on whether experts may be viewed as outliers or not.

Interval Reconciliation

An alternative non-probabilistic data model

- Each informed flow is represented by an interval containing its true value, all the wider as information is poor.
- Missing data are accommodated : the domain of possible values
- Does not make an assumption as to whether the data come from a random process or not.

Conjunctive Fusion Approach

Perform the intersection between

- the domain defined by balance flow equations
- and the domain constraints defined by intervals specified for the measured values,

and update ranges of variables accordingly.

Modeling information with fuzzy intervals

Interval reconciliation may have no solution (interval data too narrow) or may fail to alter the original intervals (if too wide).

Fuzzy intervals

Well suited to expert information. An expert can usually provide :

- an interval of values which he believes contains the "true" value
- a"preferred" (most plausible) value(s) within that interval

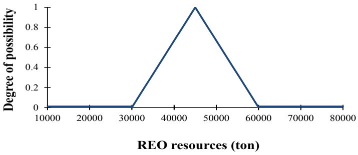
A possibility distribution with triangular shape.

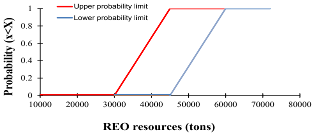
Example

expert information regarding the resources of a rare earth deposit.

More generally, a fuzzy interval represents a flexible range of a constant value or a family of probability distributions expressing imprecision.

Fuzzy intervals





Imprecise data for Nd-magnet

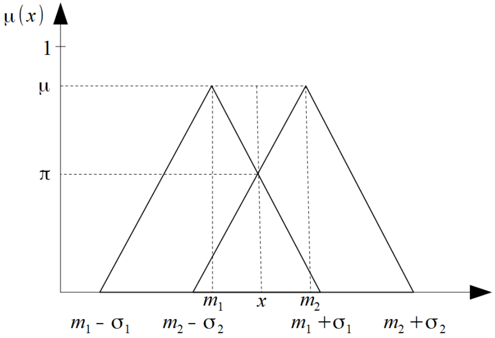
Flow/Stock	Min value	Max value	Preferred value	Flow/Stock	Min value	Max value	Preferred value
F1	180.0	250.0	215.0	F15	320.0	390.0	355.0
F2	2.0	5.0	3.5	F16	290.0	440.0	365.0
F3	160.0	230.0	195.0	F17	520.0	630.0	575.0
F4	9.0	25.0	17.0	F18	14.0	18.0	16.0
F5	45.0	65.0	55.0	F19	3.0	5.0	4.0
F6	190.0	240.0	215.0	F20	350.0	430.0	390.0
F7	150.0	200.0	175.0	F21	150.0	180.0	165.0
F8	3.0	7.0	5.0	S1	180.0	250.0	215.0
F9	160.0	180.0	170.0	S2	80.0	95.0	87.0
F10	180.0	230.0	205.0	S3	200.0	350.0	275.0
F11	250.0	310.0	280.0	S4	350.0	430.0	390.0
F12	610.0	630.0	620.0	S5	180.0	230.0	205.0
F13	240.0	400.0	320.0	P1	5.0	8.0	6.5
F14	780.0	960.0	870.0	P2	32.0	38.0	35.0

A $\max - \min$ fusion approach to reconciliation

The interval constraint satisfaction problem becomes a consistency degree optimisation problem

- The possibility distribution attached to each value of a flow evaluates the local compatibility of this choice with respect to the original estimate.
- Perform the intersection between the domain defined by balance flow equations and the flexible domain constraints on the measured values,
- the degree of consistency with the data of a set of flow values respecting the MFA model is supposed to be the worst local compatibility degree
- The optimal compatibility obtained by maximizing the worst local compatibility degree is a global measure of consistency of the data with the MFA model
- Plausible reconciled values can be derived

Intersecting fuzzy intervals



- On the left : expert estimates
- On the right : result of propagating other estimates via the MFA constraints
- The height of the intersection is the degree of consistency between the two information items.

Form of the results

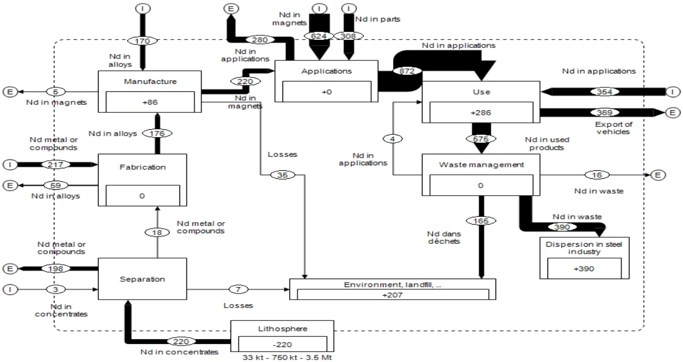
- degree of consistency of the MFA model with the data : if too low,
  - data collection to be reconsidered
  - MFA model to be questioned
- Fuzzy intervals of more or less plausible reconciled values resulting from the merging of imprecise data and constraints of the MFA models
- Precise estimates of reconciled flows consisting of most plausible values inside their reconciled fuzzy intervals

Informally : reconciled values are as close as possible to the central estimates, penalizing the largest deviation by maximizing the minimum among the compatibility degrees of all flow values, while respecting the MFA model.

Reconciled flows for Nd-magnet

Flow/Stock	Support		Core		Flow/Stock	Support		Core	
	min	max	min	max		min	max	min	max
F1	180.0	250.0	214.7	218.0	F15	320.0	390.0	348.7	354.8
F2	2.0	5.0	3.2	3.5	F16	290.0	440.0	366.1	379.1
F3	160.0	230.0	195.5	198.8	F17	520.0	630.0	570.2	579.5
F4	9.0	25.0	16.3	17.7	F18	14.0	18.0	16.0	16.4
F5	45.0	65.0	55.1	56.9	F19	3.0	5.0	3.9	4.1
F6	190.0	240.0	210.5	213.7	F20	350.0	430.0	386.5	393.5
F7	150.0	200.0	169.9	173.1	F21	150.0	180.0	163.7	165.6
F8	3.0	7.0	5.3	5.7	S1	180.0	250.0	214.7	218.0
F9	160.0	180.0	166.7	168.5	S2	80.0	95.0	87.7	89.0
F10	180.0	230.0	206.9	211.2	S3	200.0	350.0	268.5	281.5
F11	250.0	310.0	280.4	285.6	S4	350.0	430.0	386.5	393.5
F12	610.0	630.0	618.2	619.9	S5	187.0	226.0	205.3	207.2
F13	240.0	400.0	311.5	319.5	P1	5.0	8.0	6.4	6.6
F14	780.0	960.0	862.2	870.2	P2	32.0	38.0	35.2	35.7

Sankey Diagram for Nd-magnet



Estimation vs.fusion paradigms for reconciliation

Statistical approach v.s. flexible constraint-based approach : radically different problems addressed

- **Usual Least square method = Estimation + uncertainty propagation problem** : find the ideal unbiased estimate (the least square solution) and compute its distribution induced by the distributions of the data
  The resulting distributions may be wider than the original ones...
- **Information fusion problem** : find the tightest ranges for reconciled values by projecting the result of merging model constraints and data constraints. The reconciled values are computed from plausibility values inside these ranges.
  The resulting distributions, if any, are always inside the original ones...

Conclusion

Advantages of our approach in the rare earth case :

- *Its clear conceptual setting*, avoids artificially appealing to frequentist probability. It is a constraint-based fusion view, not an estimation process.
- *Its flexibility*, for representing uncertainty of data (statistical or subjective).
- The possibility of solving max – min problems by *standard methods and software*.

A software for data reconciliation and an application of the max – min approach to stocks and flow analysis for rare earths in the anthroposphere at the European scale is available at BRGM, as one of the results of the ASTER project.

More details on the method in

Didier Dubois, Hélène Fargier, Meissa Ababou, Dominique Guyonnet.
A fuzzy constraint-based approach to data reconciliation in material flow analysis.
International Journal of General Systems, Vol. 43 N. 8, p. 787-809, 2014