Data Implications for Environmental Safety and Health Assessments (ES&H) in the Semiconductor Industry

– *A Proposal for Data Standards*

Gregory J. McRae
Yue Chen

MIT - Chemical Engineering

Overview of Presentation

Key Point: There is a critical need and many benefits from establishing industry standard ES&H data bases and models

Data base Attributes

- Incorporate information about source and uncertainties
- Linked to other related information (e.g. CAS,…)
- Can represent information of different quality
- Separation between corporate confidential and public information but with the same data structures
- Framework for expansion and further research

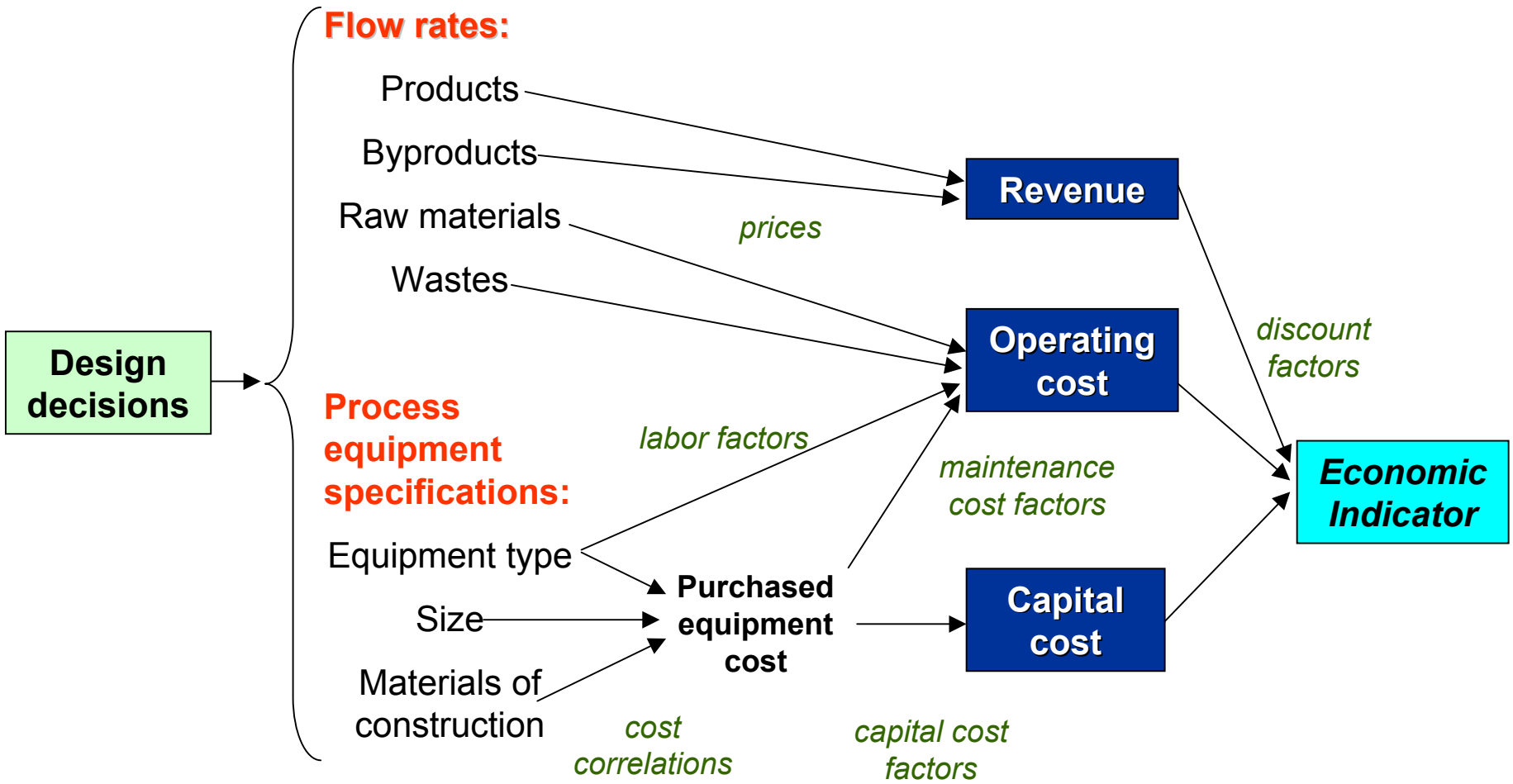
Lessons From Other Areas

1. NIST evaluated kinetics data bases and Webbook
2. Computer Aided Process Engineering CAPE-OPEN, IK-CAPE
3. Gene sequences and genomics
4. Machine Tools and CAD/CAM
5. XML Based Data exchange

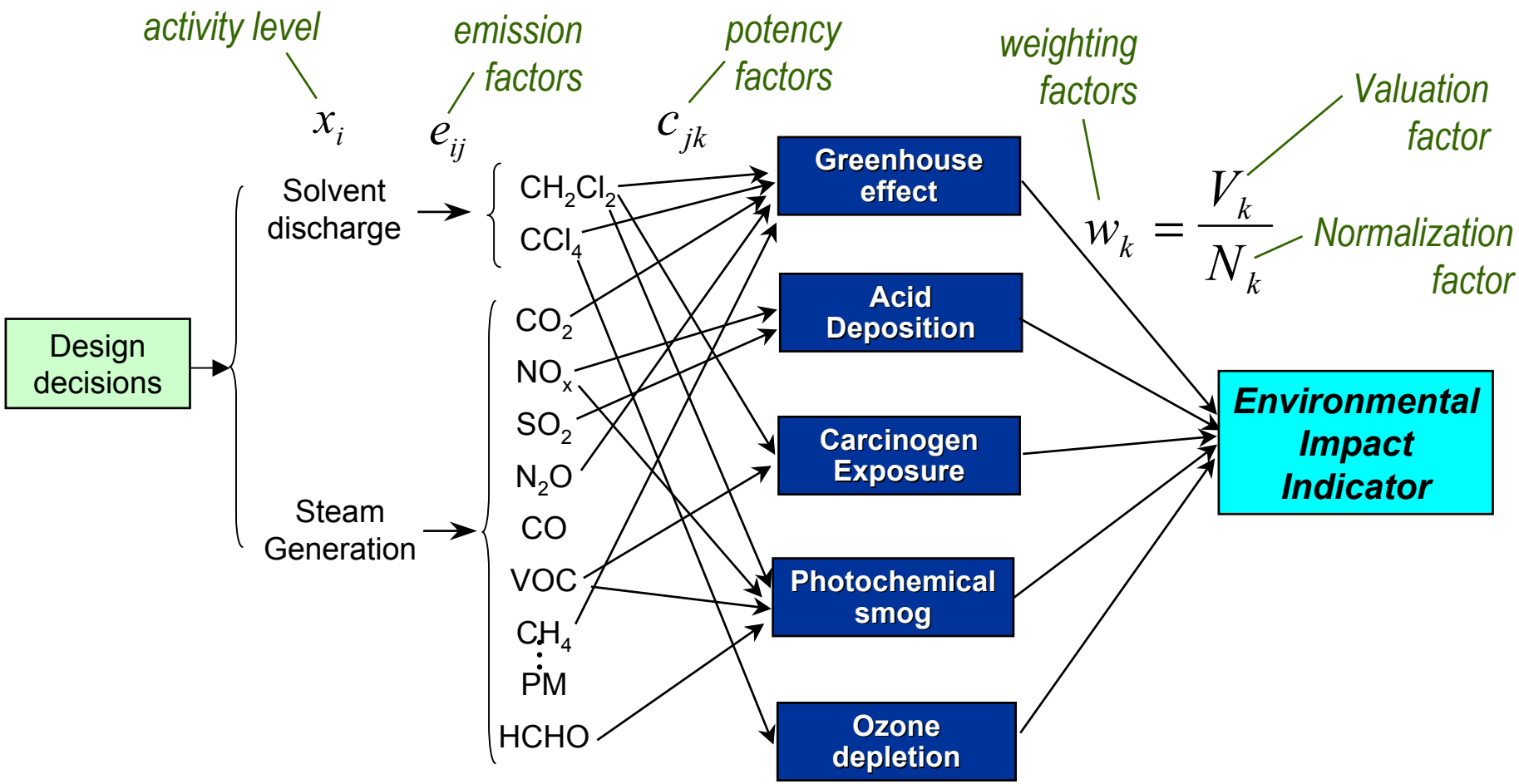
Questions that Drive our Current Research

1. How to improve both commercial and environmental performance of new and existing technologies?
2. How to measure economic, environmental and life cycle impacts?
3. How to cope with the large uncertainties in environmental evaluations?
4. How to make decisions when there are multiple objectives – environment/economics?
5. How to avoid the problems in the first place?

EVALUATION: Economic Cost of Ownership Model (COO)



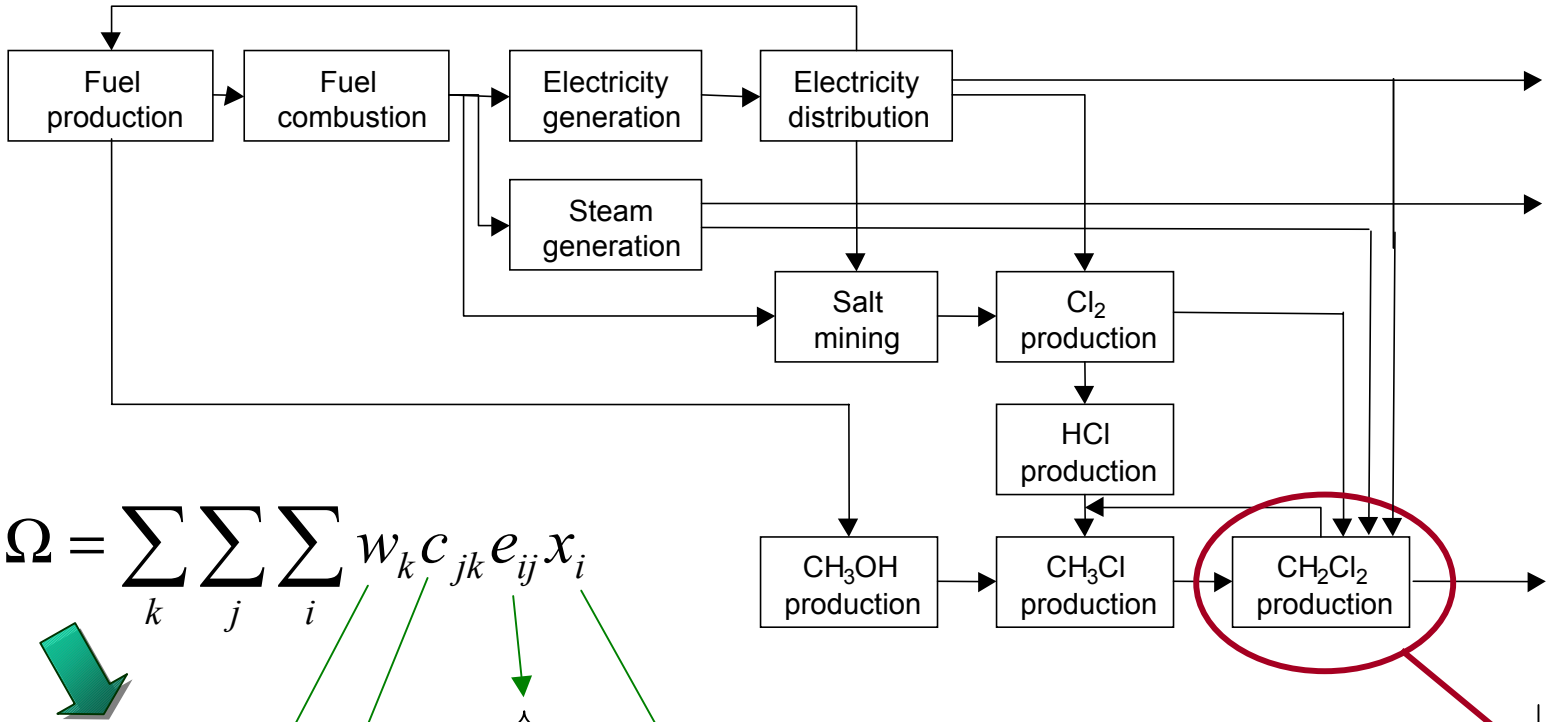
Evaluation: Environmental Impact Objective Functions



$$\Omega = \sum_k \sum_j \sum_i w_k c_{jk} e_{ij} x_i$$

Input Output Matrices for Life Cycle Assessment

Life-cycle emission inventories are correlated



$$\Omega = \sum_k \sum_j \sum_i w_k c_{jk} e_{ij} x_i$$

$$\Omega = \mathbf{w}^T \mathbf{C}^T \mathbf{E} \left(\mathbf{I} - \mathbf{D} \right)^{-1} \mathbf{x}$$

- $\mathbf{x}$ — *process utilization vector*
- $\left(\mathbf{I} - \mathbf{D} \right)^{-1}$ — *direct requirements matrix*
- $\mathbf{E}$ — *direct emissions matrix*
- $\mathbf{C}^T$ — *characterization factor matrix*
- $\mathbf{w}^T$ — *weighting factor vector*

TOXIC RELEASE INVENTORY: introduction

- Since 1987, manufacturing firms required to report releases of toxic chemicals
- data available to the public
- 370 chemicals in original list;
  current list has more than 600 chemicals
- corporations under pressure to decrease their releases
- most efforts focus on the reduction of the total mass of TRI chemicals released
- is risk being reduced?

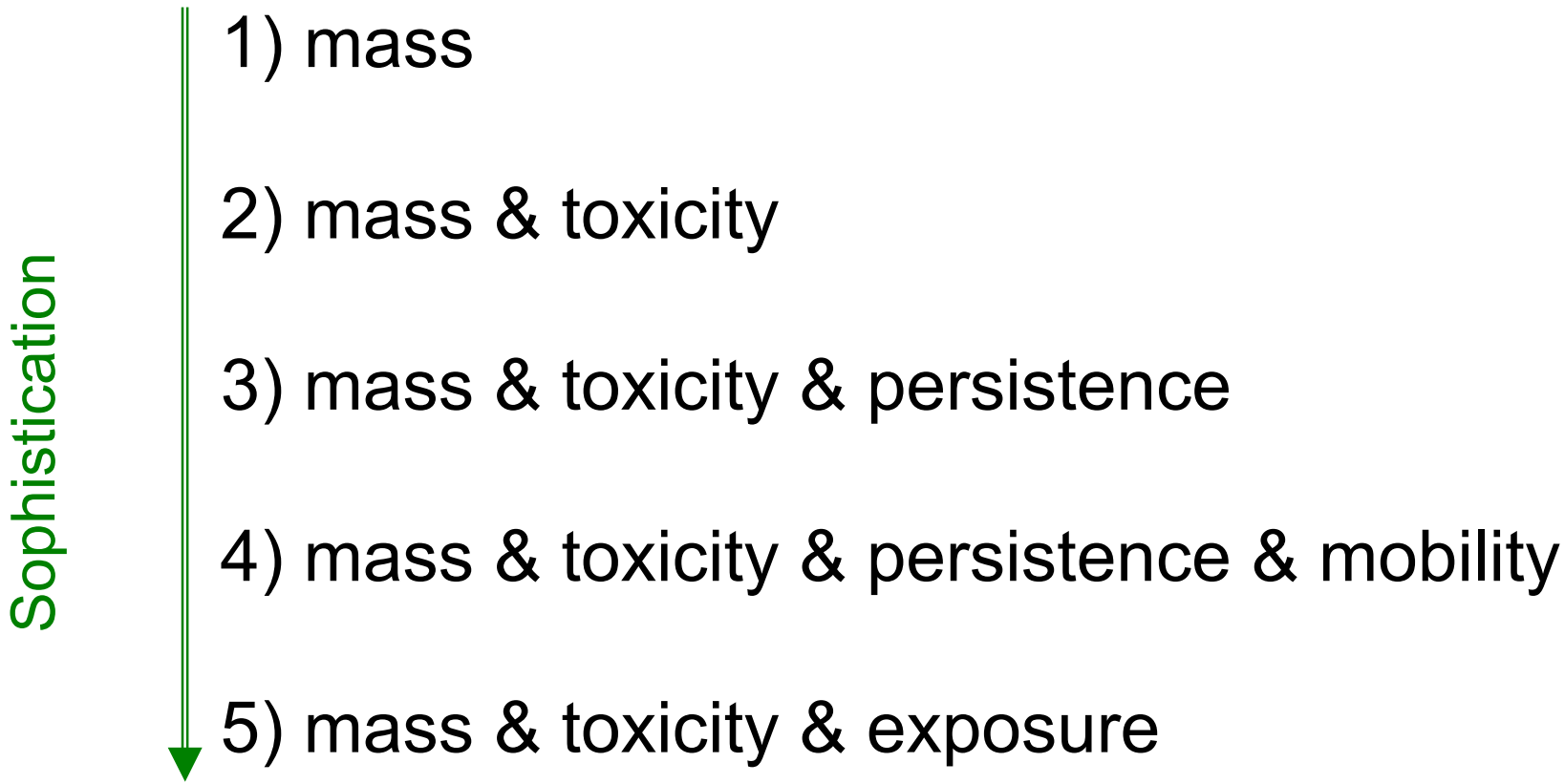
TOXIC RELEASE INVENTORY: analysis

- Releases considered:
  - total air emissions (stack + fugitive)
  - surface water discharges
  - transfers to wastewater treatment plants (POTWs)
- Toxicity indicators:
  - human exposure model (inhalation + ingestion)
  - cancer and chronic non-cancer endpoints
- Uncertainties:
  - physical properties
  - toxic potency factors
  - exposure model parameters

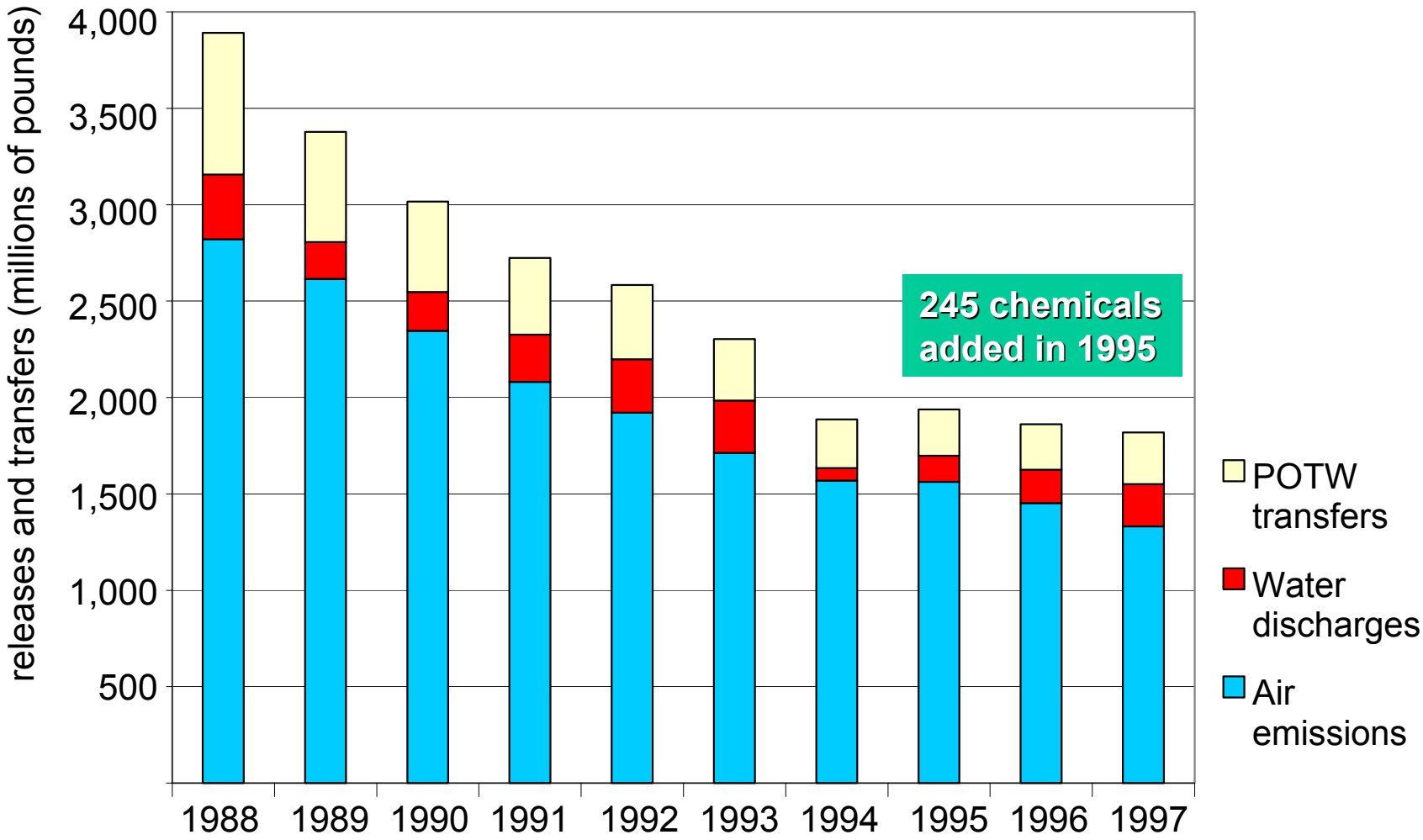
425 chemicals with releases in 1988-97

> 10,000 independent uncertainty distributions

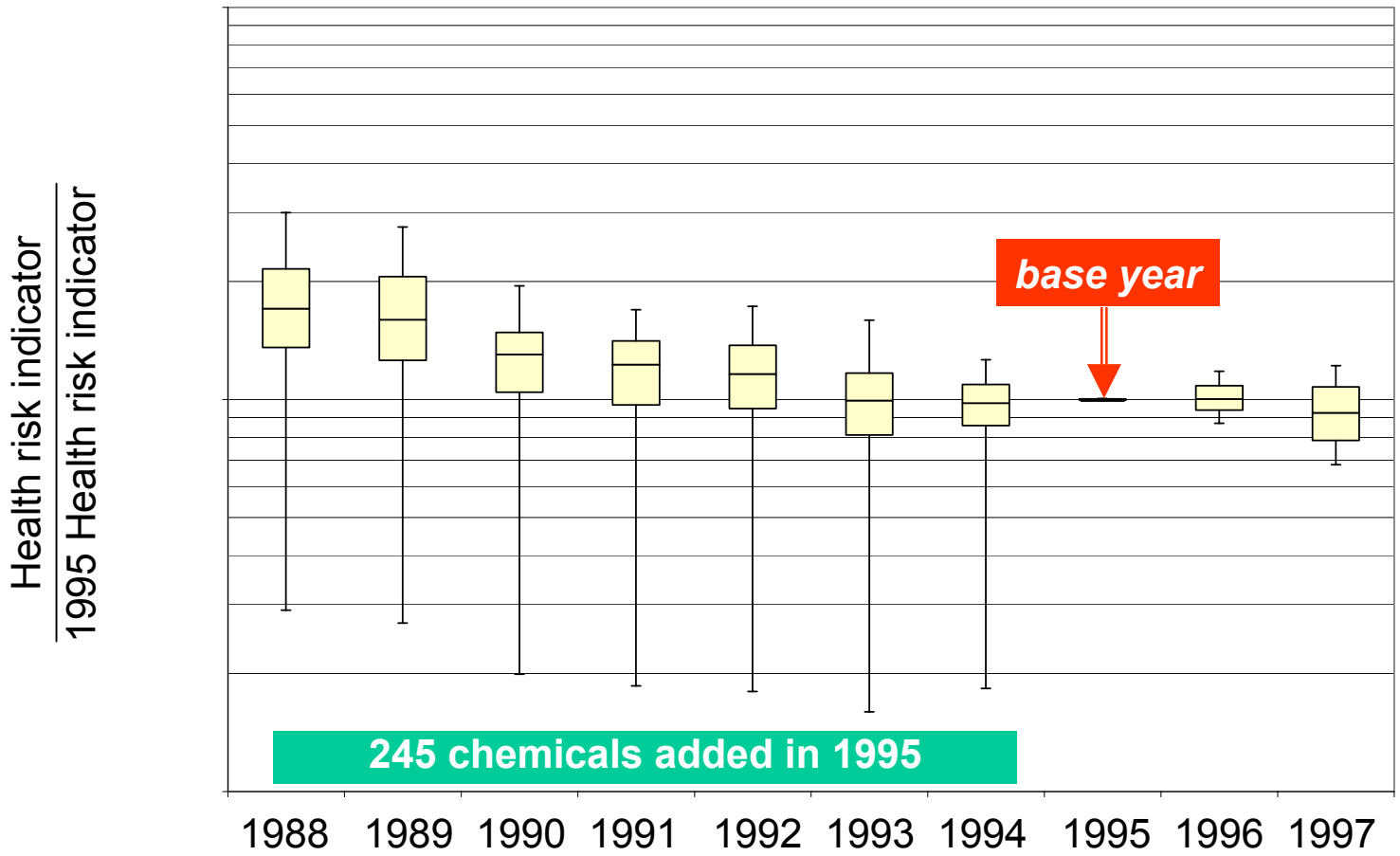
Toxicity Indicators: Hierarchical Structure



Toxic Release Inventory: Mass Contributions

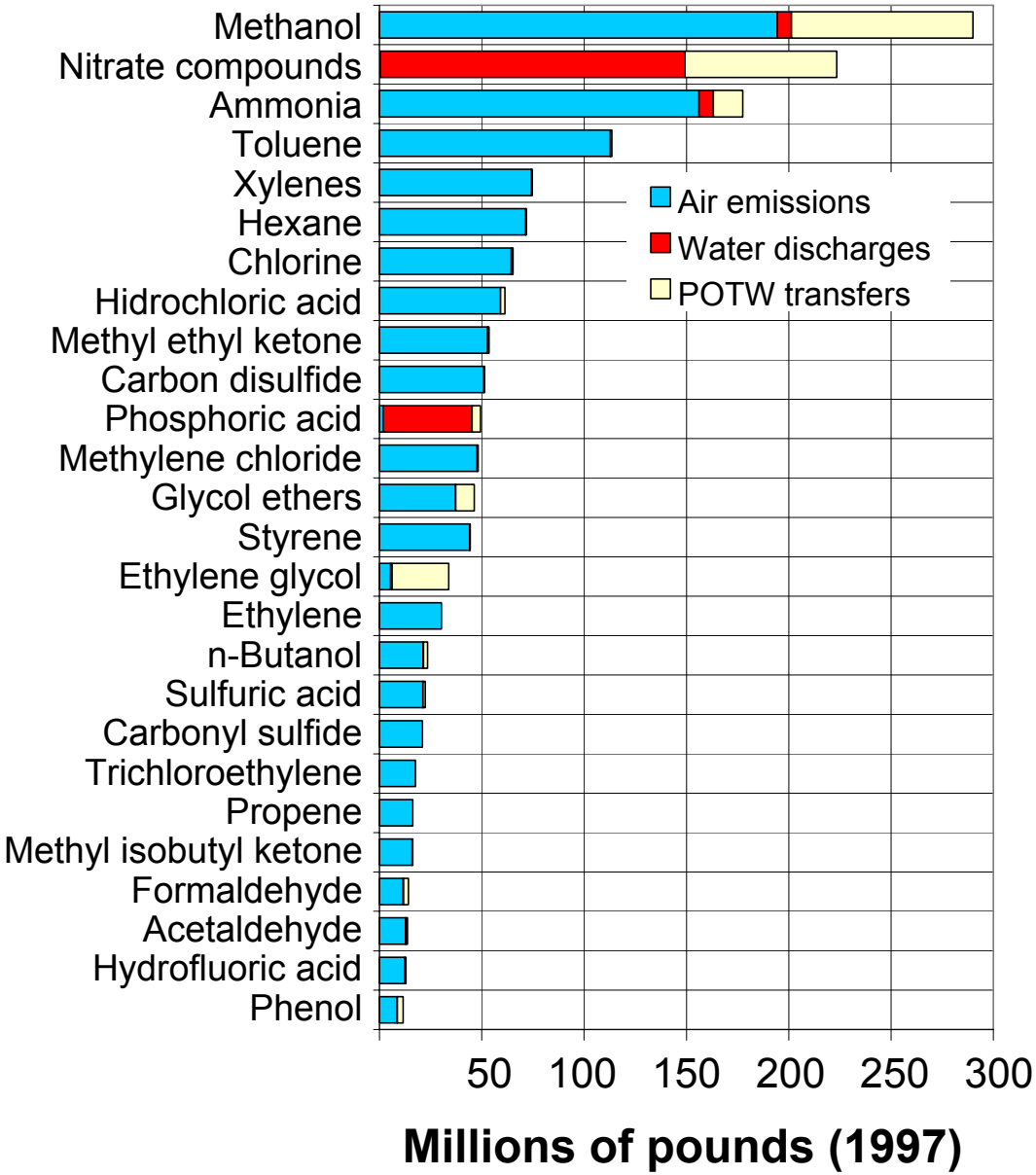


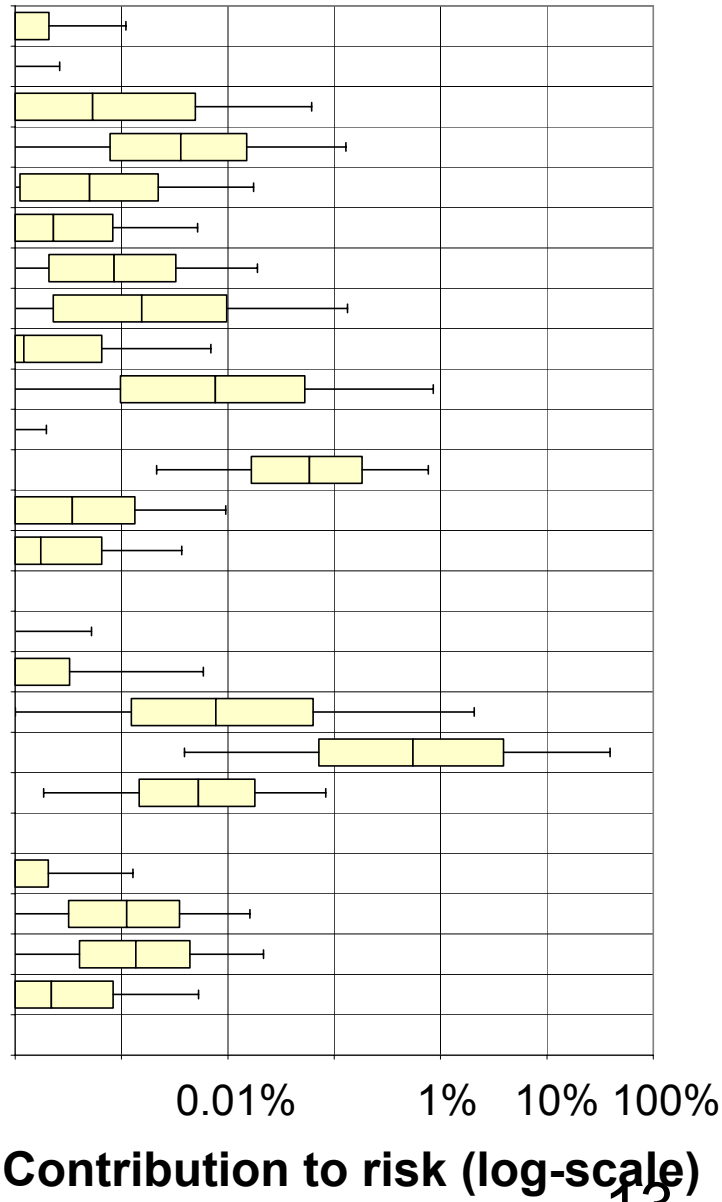
Toxic Release Inventory: Risk Relative to 1995



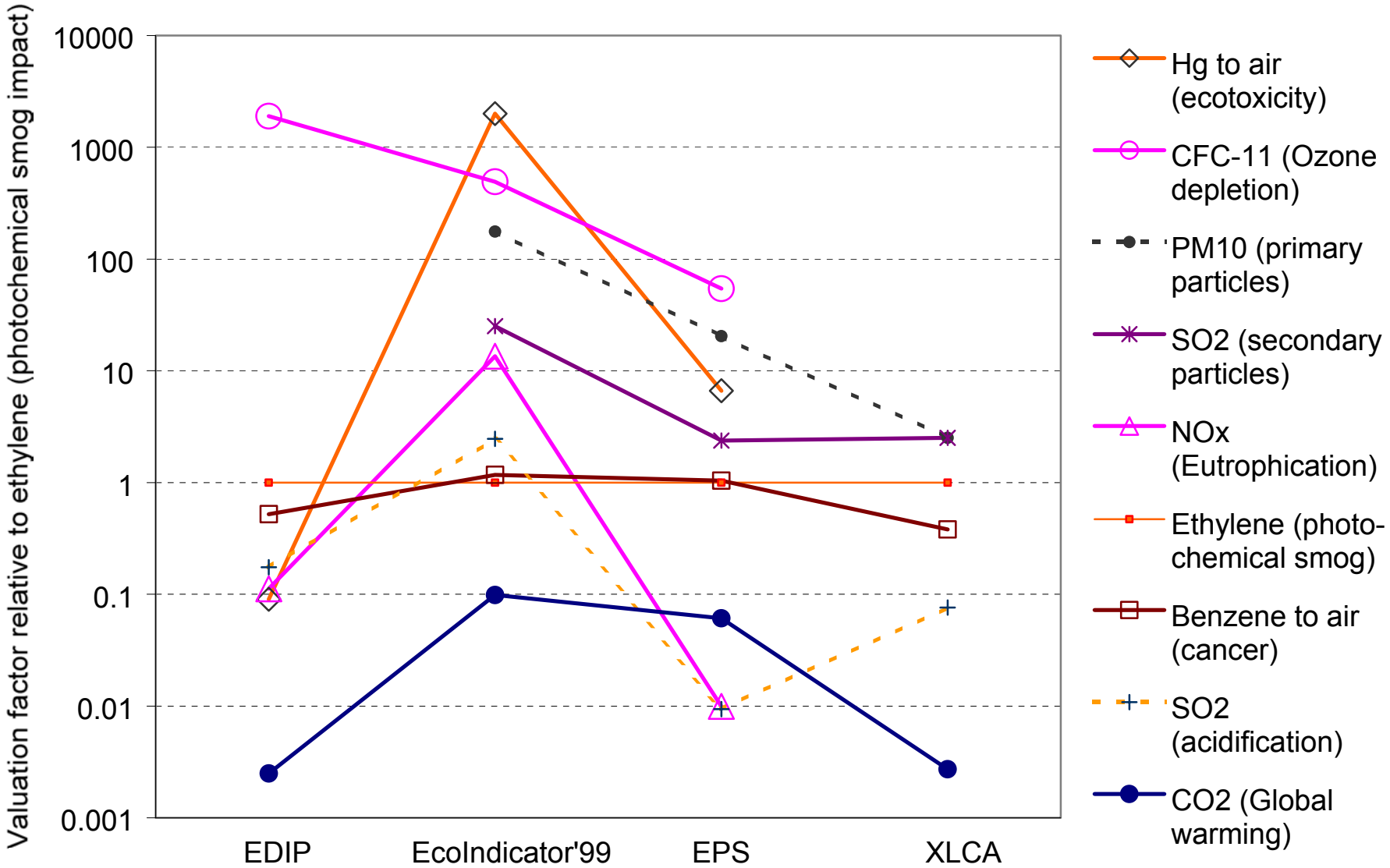
Not much progress since 1995...

Toxic Release Inventory: Chemical Contributions

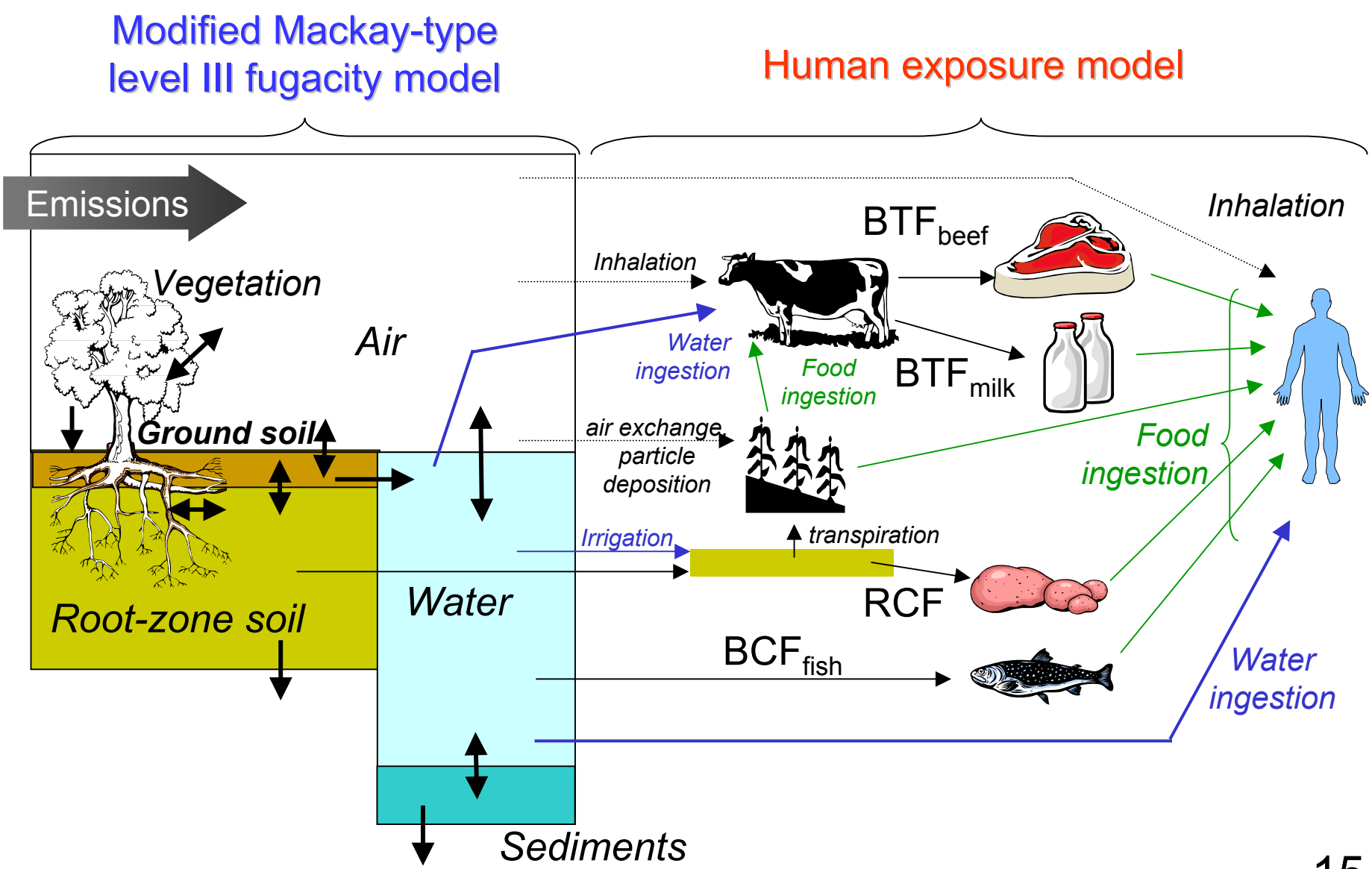




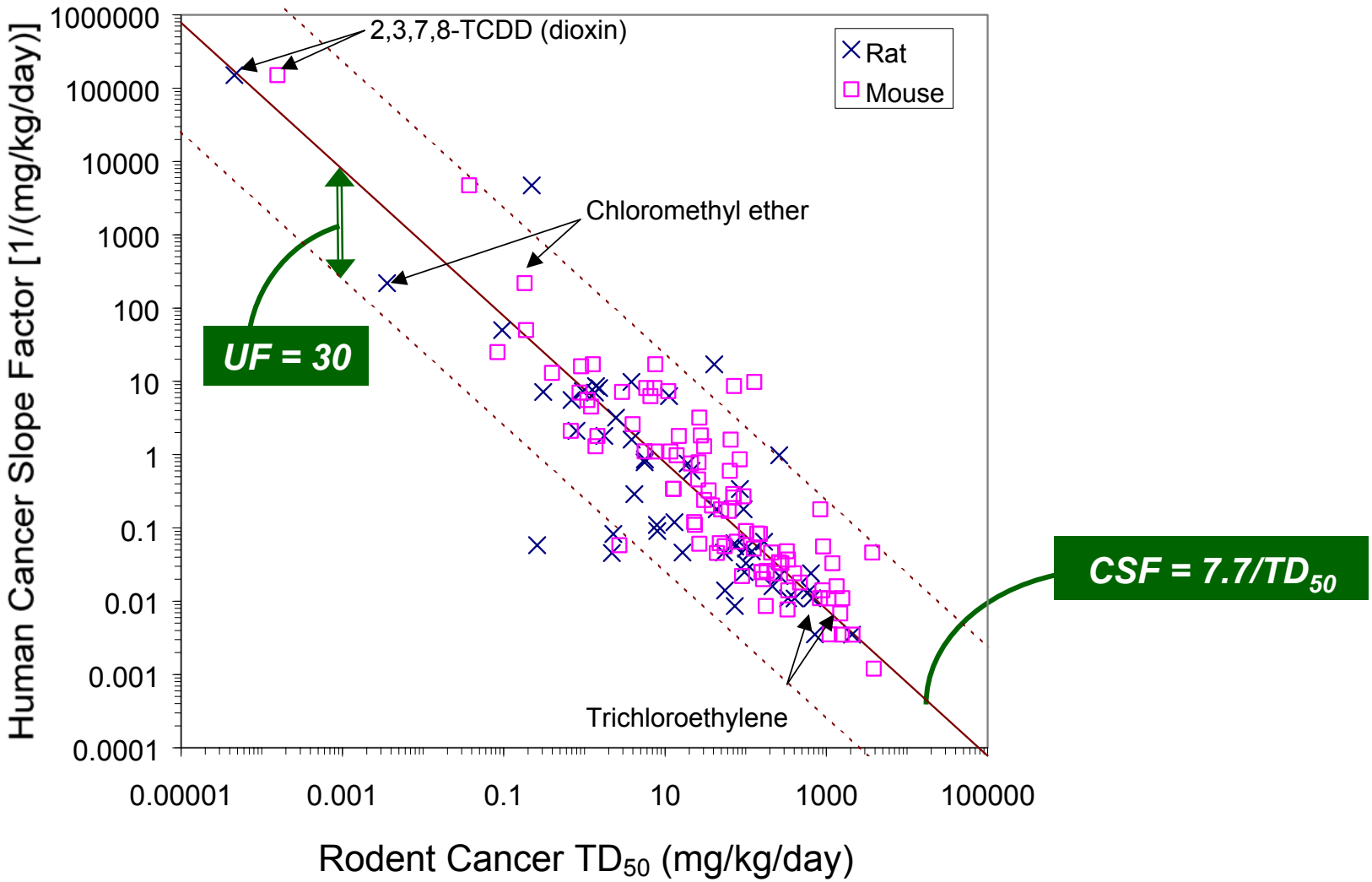
Environmental Indicators: Large Uncertainties



Human Exposure Modeling: Complex Interactions



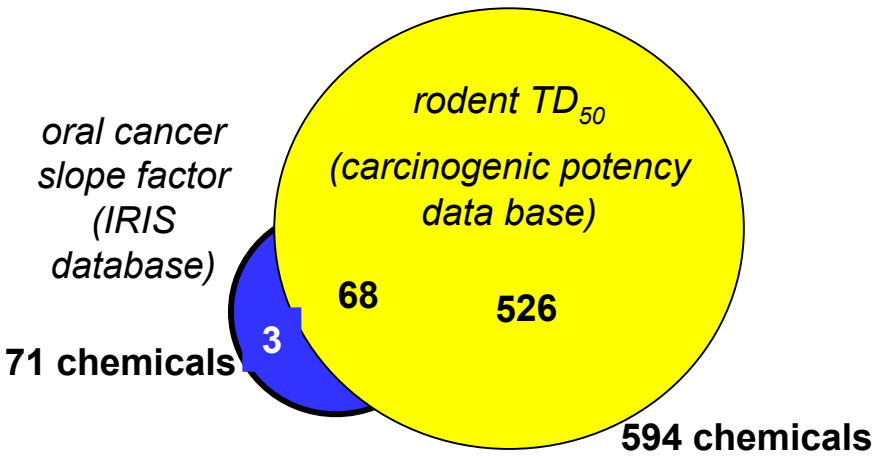
COMBINING DATA OF DIFFERENT QUALITY: example

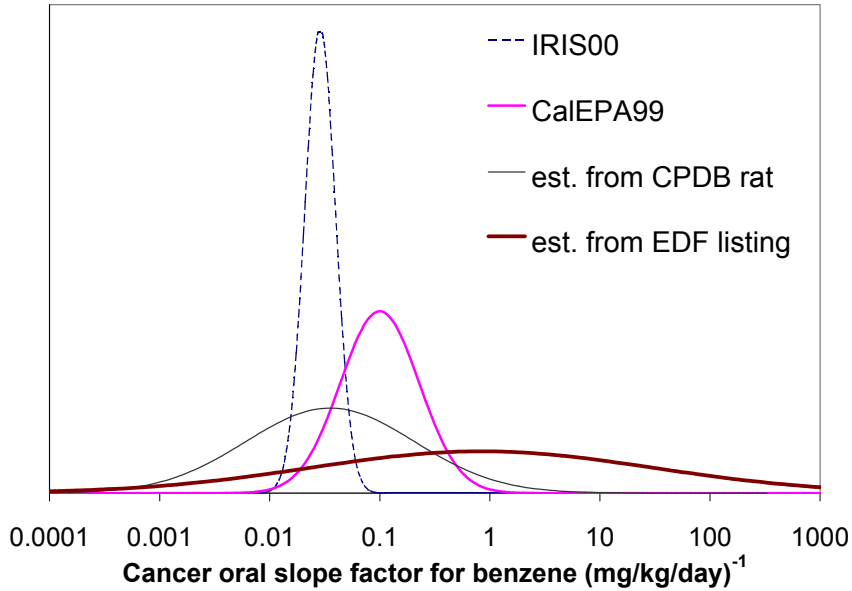


$CSF \sim LogN(7.7 / TD_{50}, 30)$

Combining Data of Different Quality: Distributions

Availability of toxicity data (cancer)





$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)} \;\alpha\; p(y \mid \theta)p(\theta)$$

Posterior Distribution

Prior Distribution

Likelihood Function

$$p(y) = \int p(\theta)p(y \mid \theta)d\theta$$

Methological Issues: Practical Issues

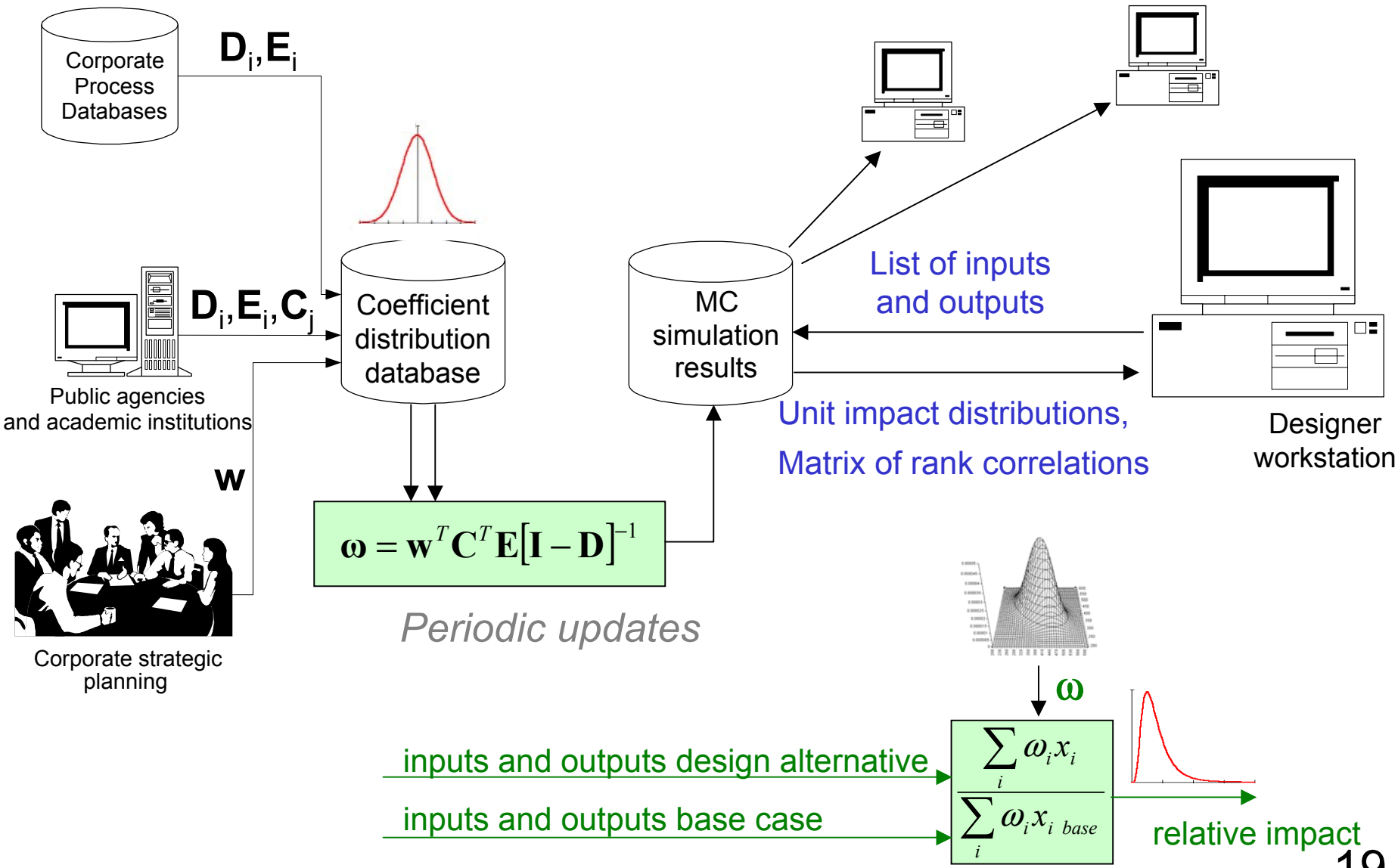
ISSUES

- How to organize the data?
- How to reduce computational cost of evaluation?
- How to implement procedures consistently within an organization?
- How to balance conflicting objectives?
- How to maximize learning from analysis?

TOOLS

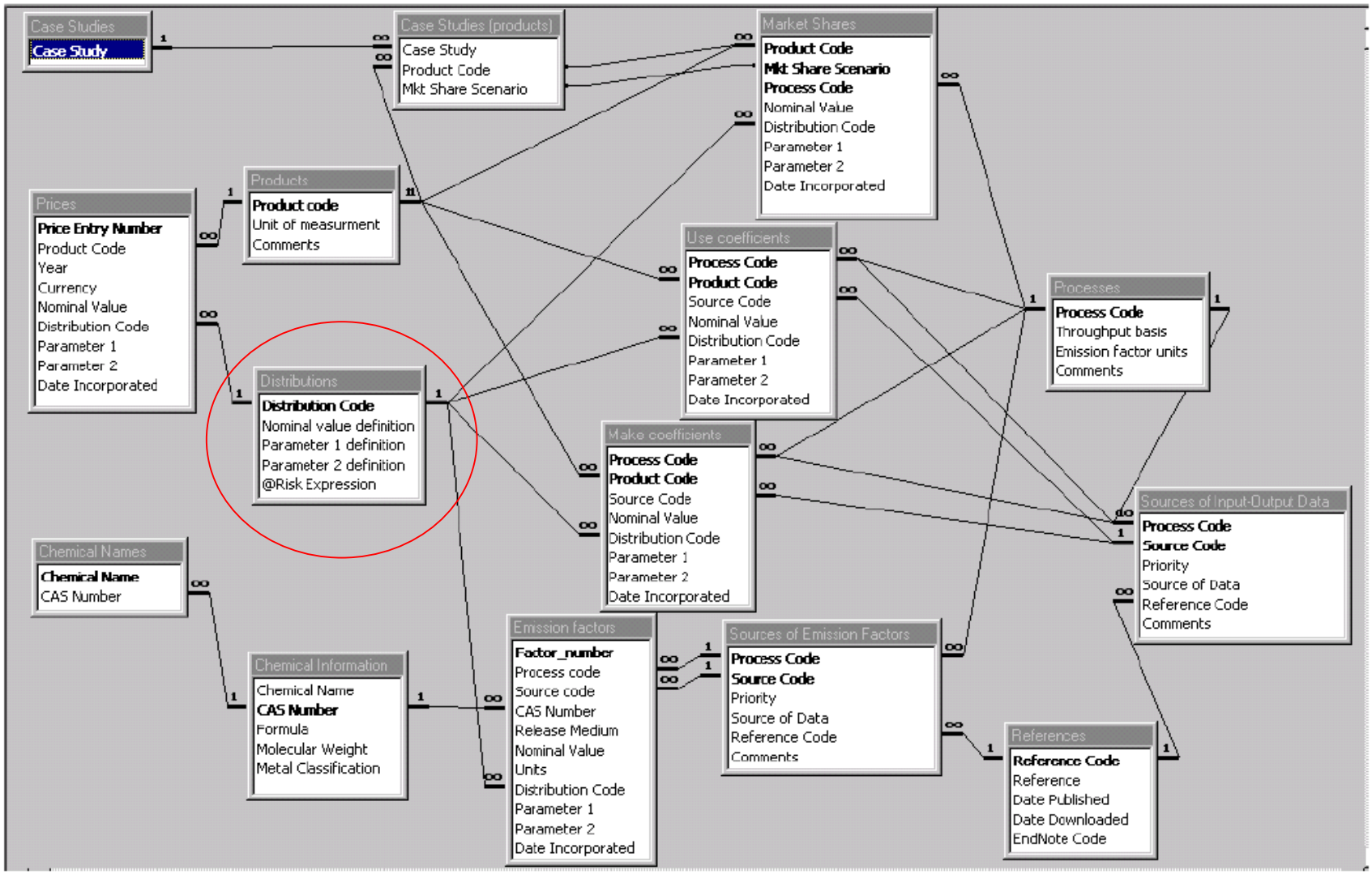
- Input/Output matrices
- Use of multivariate distribution models
- Centralized management of databases and models
- Multi-attribute utility functions
- Sensitivity/Uncertainty analysis

Implementation: Data Base Management



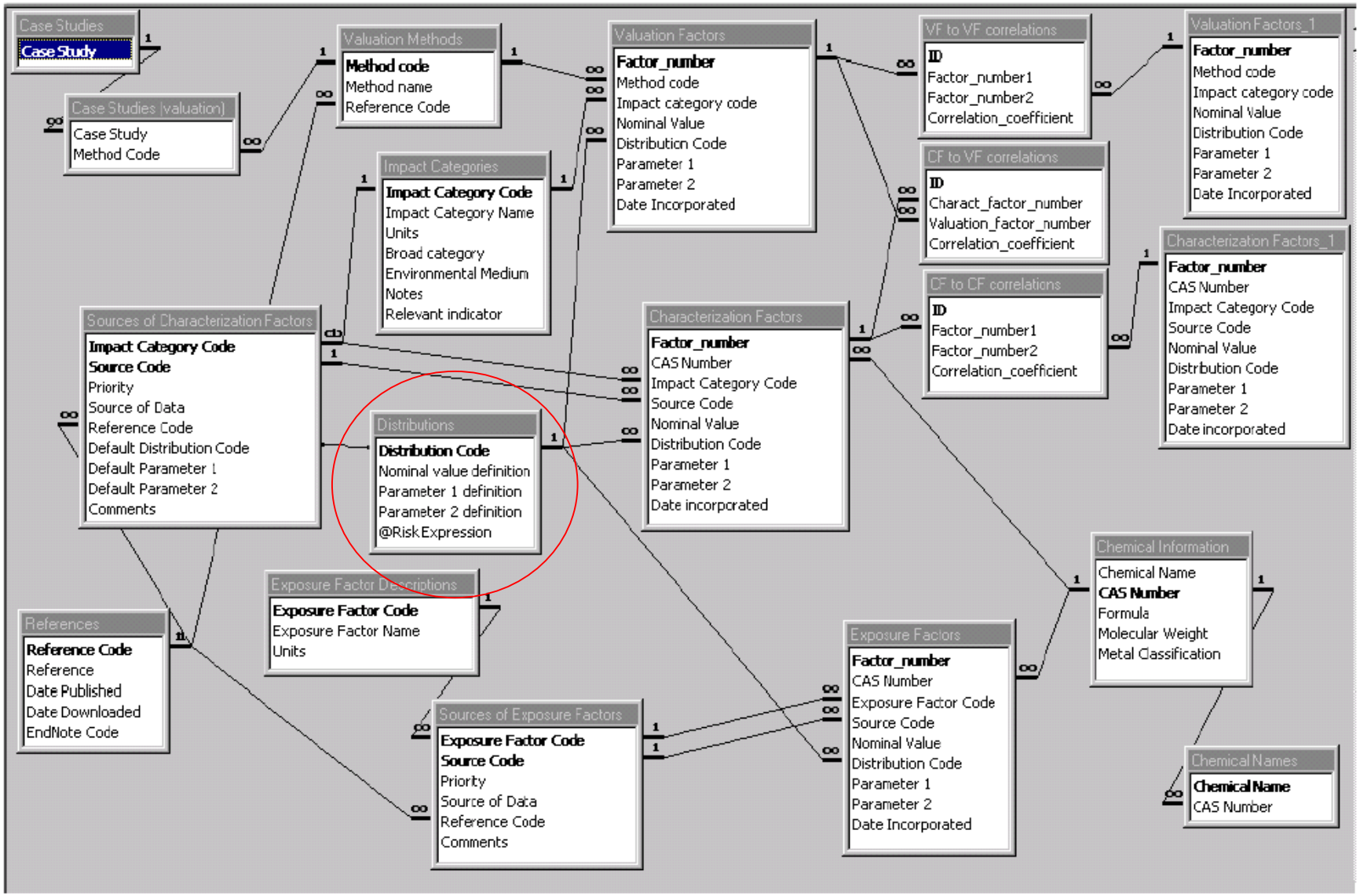
Data Storage: Example of Database Structure*

Input-output data and economic information

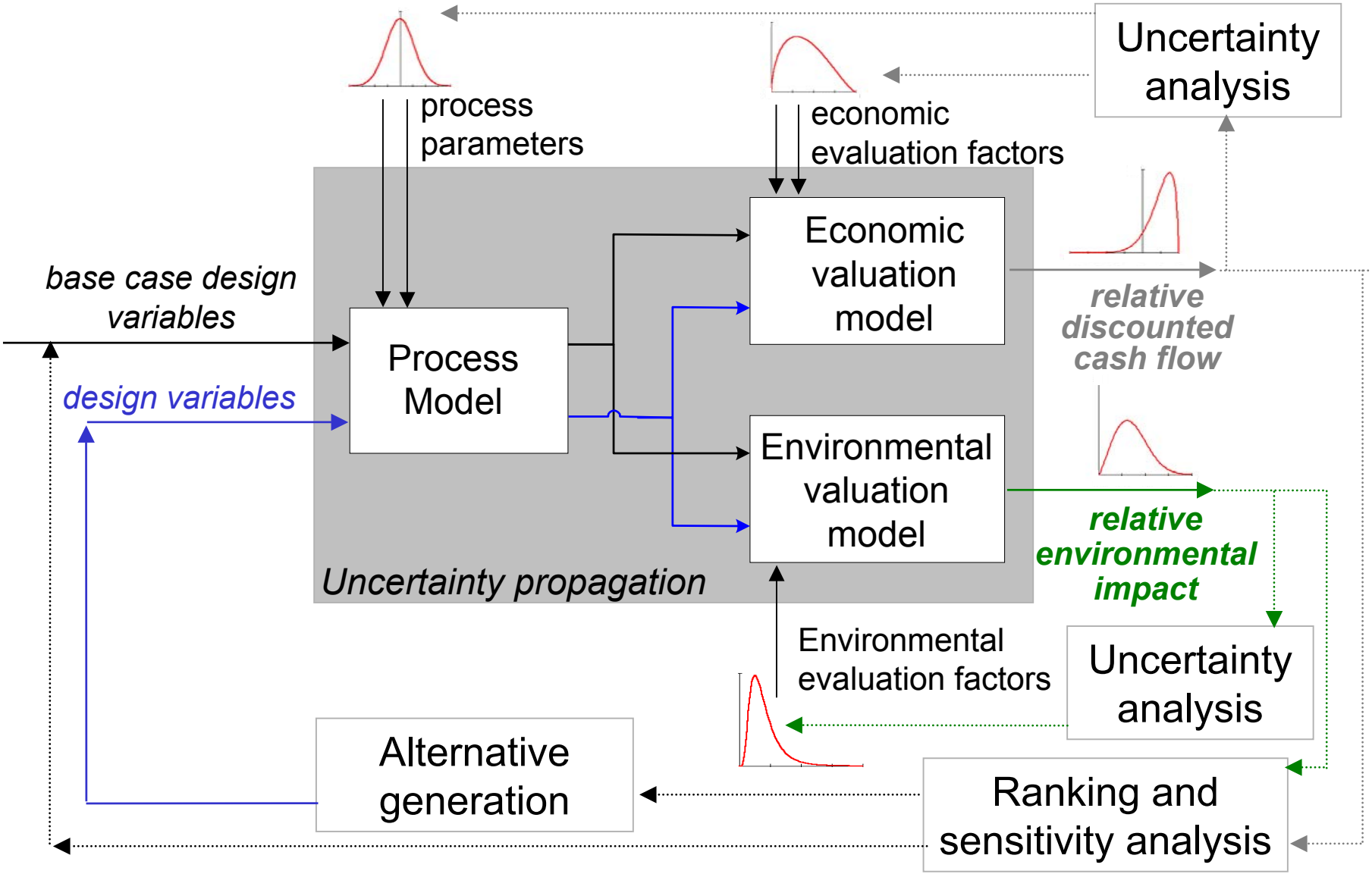


* Cano-Ruiz, 2000

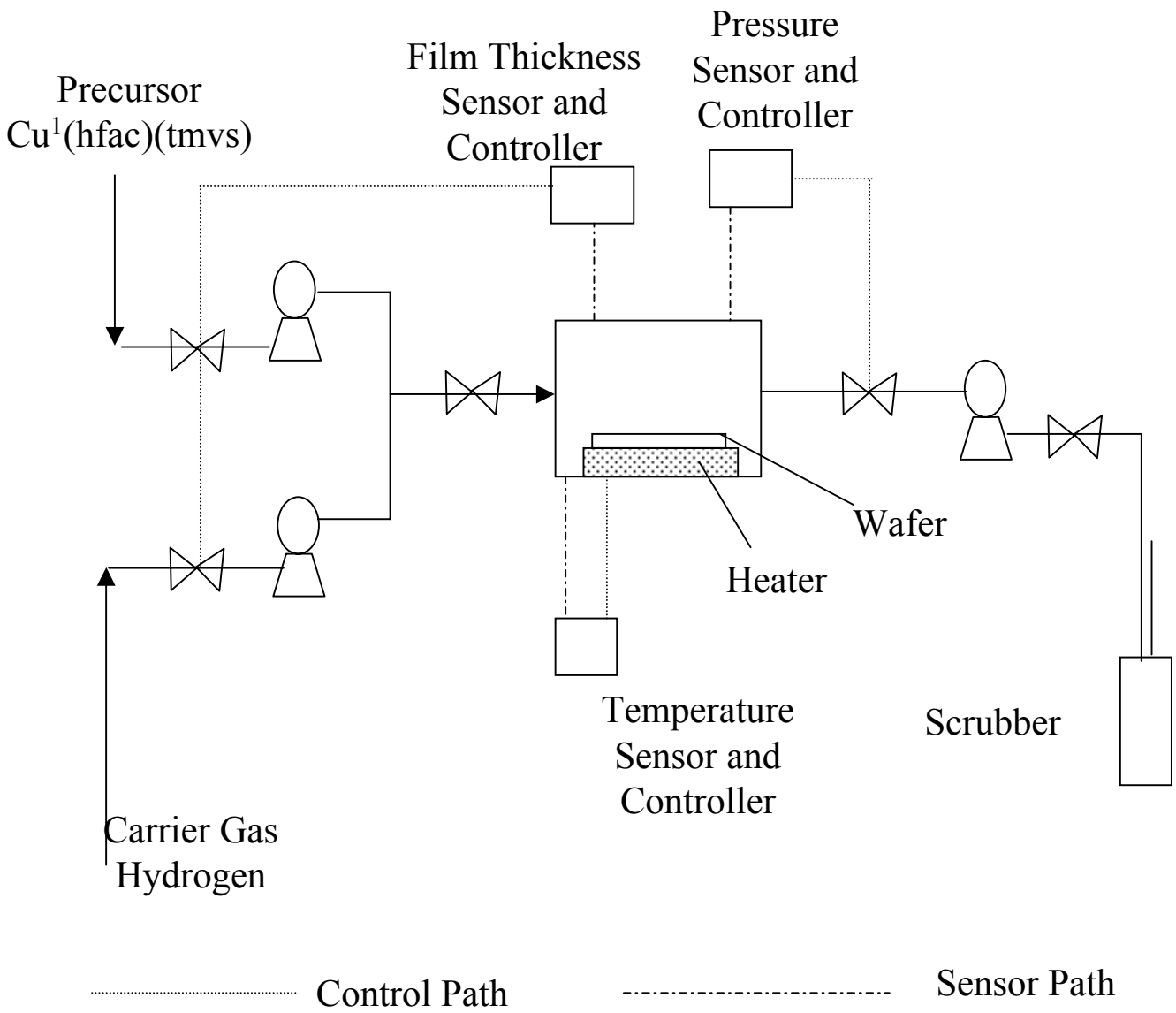
Chemical Properties and Valuation Factors



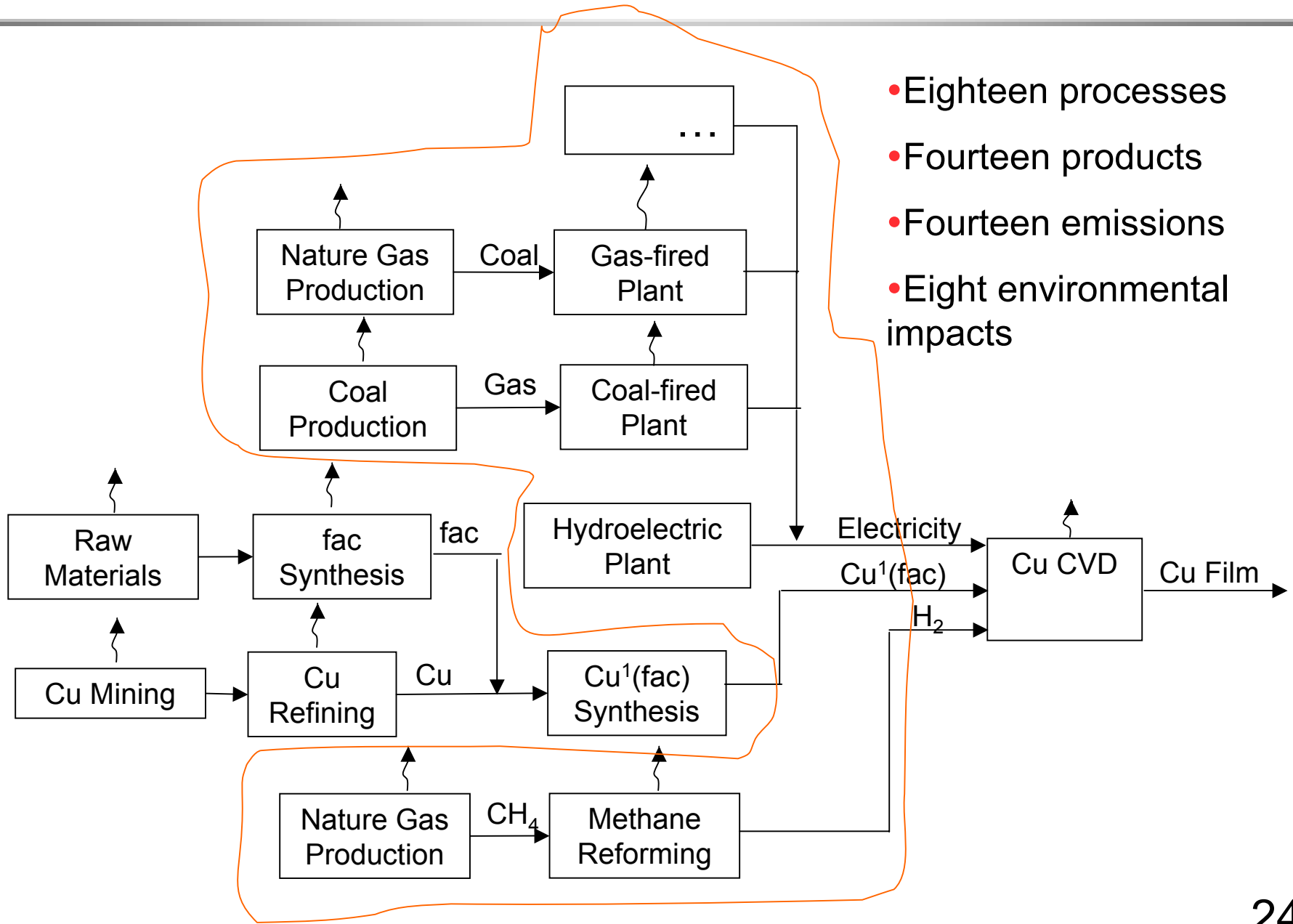
Process Evaluation: decision-making procedure



Case Study II—Cu CVD

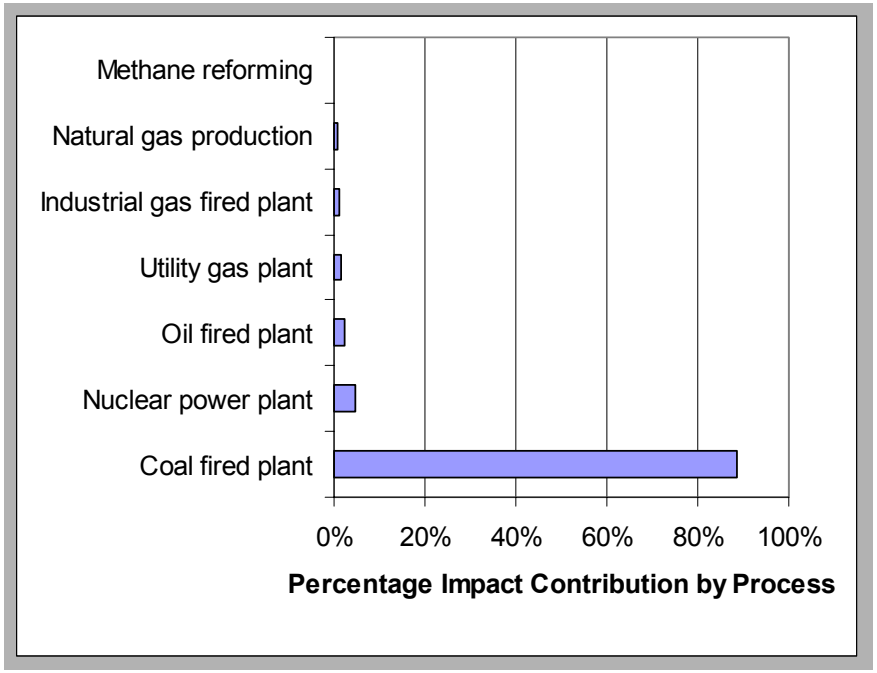


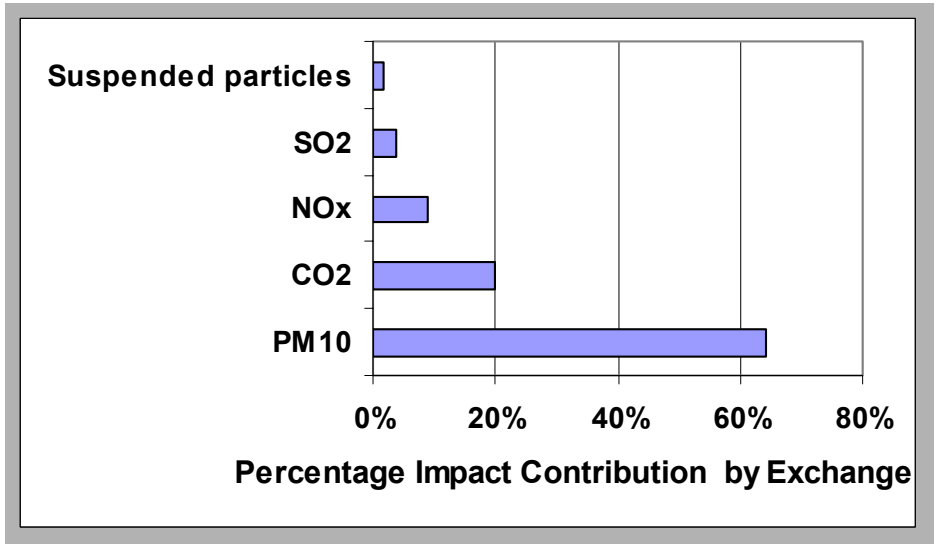
A Smaller Case



- Eighteen processes
- Fourteen products
- Fourteen emissions
- Eight environmental impacts

Evaluation Results





- Largest impact is from energy usage if coal fired power plant is used
- PM-10 is major component of impact

Conclusions

- We need to establish a working group to look at ES&H data base architecture and population
- Desirable attributes of ES&H Decision Support tools
  - Establish confidence in decisions
  - Focus alternative development efforts
  - Guide allocation of information gathering resources