An expert system approach to LCI database management

Pippa Notten & Bo Weidema
2.-0 LCA consultants
Why an expert system?

• Errors / inconsistencies not readily apparent
  – Many large tables each with hundreds of data entries
  – Data summed, normalised or otherwise aggregated
• Adequate supporting documentation often lacking
• Volume of LCI data steadily increasing
• Movement towards uniformity in format
Proposed approach

• Combine data retrieval and analysis features
  – a search for a product from a particular region returns
    the LCI for that region
  – also informs user of gaps or significant differences
    between that LCI and other relevant LCIs

• Take advantage of “multivariate viewpoint”
  – Use mathematical techniques developed to gain insight
    into multivariate systems
Potential tools

- **Exploratory factor-based / cluster analysis techniques, e.g. PCA**
  - Use information content of data
  - Essentially pattern recognition
  - Highlight large or unexpected differences between data samples

- **“Supervised learning tools” e.g. SIMCA**
  - Alert users to possible errors / inconsistencies in data
  - Provide “clues” when meta-information is lacking
Principal Component Analysis

• Reduces dimensionality of data
  – variation present in many variables represented in a small number of principal components

• Constructs a new space in which to view the data
  – enables identification of structures in the data that were previously obscured
PCA example

- Analysis of four different chemicals using four different data sources
- Benzene from data source C shows anomolous characteristics
PCA example

- PC “loadings” indicate which variables are most influential in differentiating between the cases
- Interpreted according to the length and orientation of the lines
- BOD seen to be cause of anomalous value
Practice & Limitations

• Only makes apparent what is “hidden” in data
  – Value increases with number of independent, well documented LCIs entered into system

• System must be “trained” and validated
  – Success rests on availability of data in common format with adequate supporting information

• Ideally for use across different databases
  – requires consistent data format and search criteria
Conclusions

• Considerable potential for expert system using multivariate statistical techniques
  – Identify unexpected variations, gaps, clustering etc.
  – “Unlock” information obscured by aggregation
  – Assist user in making more informed data selections
  – Provide checking / analysis when entering data

• Success requires consistent data format and sufficient number of well specified entries