Data-Analysis of Environmental Air Pollutant Monitoring Systems in Europe

Kristina Voigt, Gerhard Welzl, ²Rainer Brüggemann
GSF – National Research Center for Environment and Health
Working Group Biostatistics,
²Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin
Content

- Air Pollutant Monitoring Systems
- Objects, Attributes
- Data-analysis Methods
- Evaluation of Data-Matrix 15 x 5
  - Data-analysis by Hasse Diagram Technique
  - Data-analysis by Data Reduction Methods
    - PCA
    - POSAC
- Discussion of Results
15 Member States of EEC
Air Pollution Monitoring / EEC

Concept of public’s right to information

Proposal for a Directive of the European Parliament and of the Council on public access to environmental information exists

- Preserving, protecting and improving the quality of the environment
- Protecting human health
- Prudent and rational utilization of natural resources
- Promoting measures at the international level to deal with regional or world-wide environmental problems
Necessity to Find and Analyze Air Pollutant Monitoring IS

Where on the Internet?

http://www.stadtklima.de/stuttgart/s-luft/links.htm

Quality of the Systems
Search Strategy for Air Pollutant Monitoring Information Systems

- Key Air Quality Links http://unr.edu/homepage/daved/airqual.html
- Search in GOOGLE http://www.google.com
  - "air monitoring"
  - "name of European capital"
Data-Matrix 15 x 5

**15 Objects:** Air Pollutant MS in European capitals

**5 Variables:** Evaluation Criteria

**Scoring System:**
- 0: bad
- 1: medium
- 2: good
Evaluation Criteria (1)

- NU: Number of chemicals
- ME: Type and duration of measurements
- ST: Measurement stations
- PR: Data presentation
- BM: Background material
Evaluation Criteria (2)

- NU: quantitative, type of pollutant
- ME: temporal
- ST: spatial
- PR: descriptive
- BM: descriptive
Evaluation Criterion: NU
Number of chemicals monitored

<table>
<thead>
<tr>
<th>≤ 4</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-7</td>
<td>1</td>
</tr>
<tr>
<td>&gt; 7</td>
<td>2</td>
</tr>
</tbody>
</table>
## Evaluation Criterion: ST
Measurement stations in capital

<table>
<thead>
<tr>
<th>Number of Stations</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2 stations</td>
<td>0</td>
</tr>
<tr>
<td>3-5 stations</td>
<td>1</td>
</tr>
<tr>
<td>&gt; 6 stations</td>
<td>2</td>
</tr>
</tbody>
</table>
# Data-Matrix (abbreviated)

<table>
<thead>
<tr>
<th>Abb.</th>
<th>NU</th>
<th>ME</th>
<th>ST</th>
<th>PR</th>
<th>BM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUS</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>BEL</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>DEN</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>FIN</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FRA</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Analysis of Data-Matrix with Different Methods

- Hasse Diagram Technique
- Dimension Reduction Methods
- Similarities and Difference in Methods
Chemical or Environmental Data-Matrix

Analysis of a Data-Matrix

- Variance
- Distance
- Order Relation

- Scatterplots
  - Parallel coordinates

- Principle Component Analysis
  - Correspondence Analysis

- Cluster Analysis
  - Bertin Strategies

- Hasse Diagrams
  - Formal Concept Analysis

- Multidimensional Scaling

- Facet Theory
  - POSAC
  - Discrete Mathematics

Order Relation

Distance

Variance

Multidimensional Scaling

Cluster Analysis

Bertin Strategies

Parallel coordinates

Scatterplots

Principle Component Analysis

Correspondence Analysis

Hasse Diagrams

Formal Concept Analysis

Facet Theory

POSAC

Discrete Mathematics
HDT: Four Point Program (1)

- Set of Elements: Ground set $O$
- Information Base IB
- Find a common orientation for all properties
- Analysing $x, y \in O$ whether one of the following relations is valid
HDT: Four Point Program (2)

Analysing $x, y \in O$ whether one of the following relations is valid

- $x \sim y$ (equivalence)
- $x \leq y$ or $x \geq y$ (comparability)
- $x \parallel y$ (incomparability, there is a “contradiction in the data of $x$ and $y$”).
Hasse Diagram: Comparabilities / Characterizing Numbers

- **V**: Comparabilities
- **U**: Incomparabilities
- Complex with equivalent objects

- **NECA**: Number of equivalence classes
- **NL**: Number of levels
- **NEL**: Number of elements in the level, which contains the most elements
Basis of the HDT is the assumption that we can perform a ranking while avoiding the use of an ordering index.

HDT is very appropriate for a comparative evaluation of objects when a multicriterial assessment is envisaged.

Hasse diagrams visualize so-called partially ordered sets (posets).
Hasse Diagram Technique (HDT) (2)

- **Objects** (elements)
- **Criteria** (attributes, variables)
- **Equivalent objects**: Different objects that have the same data with respect to a given set of attributes
- **Maximal objects** (greatest element)
- **Minimal objects** (least element)
- **Isolated objects**
D-Matrix: Analysis of Successors

- Geometrical analysis of a Hasse diagram to investigate substructures
- Preferably a maximal object is chosen
- Key element \( (k) \)
- All elements located lower than that of the key element: successors

\[ G(k) = O(k) \setminus \{k\} \]
W-Matrix: Dissimilarity-Matrix

- Describes the influence of the attributes on the Hasse diagram
- **W-matrix** contains the mutual comparisons of the Hasse diagrams
- Heart of the analysis
- Dissimilarity-matrix: The larger the matrix-entries are, the greater is the difference between the successor sets for the element $k$ and hence between the Hasse diagrams.
PCA: Principal Component Analysis

- MVSP Statistics Package Kovach Computing Services
  http://www.kovcomp.co.uk/mvsp/
- Aim: Reduction of the Dimensionality of the Data
- The new variables, the principal components, are defined as linear functions of the original variables
POSAC

- Partially Ordered Scalogram Analysis with Coordinates
- Systat 10 program
- Iteratively computes a configuration of points in a two-dimensional space according to the partial order model
- Order relations are considered as the essential aspect of the data to be preserved in the data-analysis
POSAC

- Data reduction on attribute side
- 2 Dimensions
- LOV (Latent Order Variables)
- Percentage of order relations are kept / lost
Application of Analyses Methods

15X5 DM

D-Matrix → HDT → W-Matrix

Dimension Reduction

PCA

POSAC

Correlation of Attributes with LOV1/LOV2
HDT with LOV1/LOV2
Hasse Diagram Results (1)

- **Maximal Objects**: \{DEN\}, \{FRA\}, \{GER\}
- **Minimal Objects**: \{LUX\}, \{GRE\}
- **Comparabilities** \(V(N)\): 57
- **Incomparabilities** \(U(N)\): 48
- \(K = 0\)
Hasse Diagram: Levels

6 Levels

4 Objects in largest level
D-Matrix: Analysis of Successors

- Maximal objects are taken as key objects
- Search of all objects which are located lower than the key object
- GER 10 successors
- FRA 11 successors
- DEN 4 successors
W-Matrix: Dissimilarity Matrix

Case 3 (leaving out the attribute ST) leads to 26 changes in the original diagram

Measurement stations in capital is the most important attribute

Data-matrix: 16 X 4 (NU, ME, PR, BM)

ST left out
Data Reduction Methods

PCA case scores

Axis 2

Axis 1

Vector scaling: 2.27

PCA

POSAC
PCA Biplot

Vector scaling: 2.27
Component 1 is influenced by NU, BM, ME, PR
Component 2 is influenced by ST
Component 3 has a positive loading 0.754 for BM
POSAC Plot for Data-Matrix 15 x 5

- Dimension reduction
- 2 latent order variables
- 93.4 % correct
- Minimal (00000) and maximal (22222) are added.
Interpretation of LOVs using ANOVA

- **LOV (1)** is highly correlated with **ST**
  - F-Statistik: 26,556
- **LOV (2)** is highly correlated with **PR** (ME)
  - F-Statistik: 32,7
- **LOV (1)** can be described by **ST**
- **LOV (2)** can be described by **PR**
- **POLAR Items**
Interpretation of LOVs using ANOVA

Attribute: Measurement stations in capital is described by LOV(1)

Attributes: Way of presentation on the Internet is described by LOV(2)
Scatter Plots of Variables ST and PR

\[ .=0, \times=1, + = 2 \]
Comparison of Hasse Diagrams

15x5

15x2
Hasse Diagram of the reduced Data-matrix 15 X 2

ST = 2

ST = 0,1
Interpretation POSAC/HASSE

- Maximal objects: same, Minimal objects: plus {POR}
- More V, less U
- Generation of HD from POSAC plot leads to clearer structures
- Better interpretation
- Two parts:
  - Right: ST = 2
  - Left: ST = 0,1
- Still further research necessary
POSAC - MPOSAC

- Maybe 2 dimensions insufficient
- Higher dimensions proposed
- MPOSAC = Multidimensional POSAC
- Not available (yet) in Systat
- Available in HUDAP
- Instead: Subgroups according to the influences of PCA (Variable BM) are looked upon
Discussion of Results: Objects

- Germany, Denmark, France offer valuable data on air pollution monitoring (HDT, POSAC)
- Luxembourg, Greece show rather low information (HDT)
Discussion of Results: Objects

- POSAC analysis supports the results of the initial Hasse diagram technique approach
- Initial data-matrix 15 x 5 can be reduced to 15 x 2 (latent order variables)
- Very few changes but improvement for interpretations (ST = 2, ST = 0,1)
Discussion of Results: Attributes

- ST (measurement stations): highly correlated with LOV1: POSAC
- PR (way of presentation): highly correlated with LOV2: POSAC
- ST: highly correlated with Component 1: PCA
- BM: highly correlated with Component 2: PCA
- ST: W-Matrix (HDT)
- Importance of spatial aspect of air monitoring information systems
Outlook

- More and different Evaluation Criteria
- Combination of HDT with other MVS methods
- POSAC – MPOSAC
- 2 Dimensions – more Dimensions
Further Research

- Multivariate explorative statistical methods offer simple and effective tools for graphical analysis of data-matrices
- Ranking using Hasse Diagram Technique
- Choice and Preference of Method(s) is problem-driven
- Combination of Methods is the Aim of future research
Collaboration between MVS und HDT

Rainer Brüggemann, Stefan Pudenz
IGB, Berlin, Criterion Berlin

Kristina Voigt, Gerhard Welzl
München, GSF-IBB
Data-Analysis of Environmental Air Pollutant Monitoring Systems in Europe

Dr. Kristina Voigt
GSF - National Research Center for Environment and Health
Institute for Biomathematics and Biometry
Ingolstaedter Landstrasse 1
85764 Neuherberg
Germany
Tel: ++49 (0)89/3187-4029
Fax: ++49(0)89/3187-3127
e-mail: kvoigt@gsf.de
URL: http://www.gsf.de/institute/ibb/voigt
http://www.wiz.uni-kassel.de/dain