Accounting for complex environmental exposure situations: a classification approach
Benoit Lalloue, Jean-Marie Monnez, Cindy Padilla, Wahida Kihal, Denis Zmirou-Navier, Séverine Deguen

To cite this version:

HAL Id: hal-01432884
https://hal.archives-ouvertes.fr/hal-01432884
Submitted on 12 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Accounting for complex environmental exposure situations: a classification approach

Benoît Lalloué1,2,3,4, Jean-Marie Monnez3,4, Cindy Padilla1,2, Wahida Kihal1, Denis Zmirou-Navier1,2,5, Séverine Deguen1,2

1Inserm, UMR 1045-IRSET (Institut de recherche sur la santé et le travail), France
2EHESP Rennes, Sorbonne Paris Cité, France
3Lorraine University, CNRS UMR 7502, Institut Elie Cartan, France
4University of Lorraine, ENSEA, CNRS UMR7502, BIGS (INRIA Nancy - Grand Est / IECN), France
5Lorraine University Medical School, France

Background

✓ Everyone is constantly exposed to several environmental exposures with positive or negative health effect.
✓ Studies which consider this complex environmental setting are rare.
✓ There is a scientific and political call for a realistic and “holistic” approach of cumulative exposure.
✓ There is a need for methods able to handle cumulative exposures.

Study design and data

✓ Lyon metropolitan area (1.2 million inhabitants, 527 km²), France
✓ French census blocks (2000 inhabitants on average)
✓ Environmental exposures groups:
  • NO₂ annual concentration (2 variables)
  • Noise levels (3 variables)
  • Traffic exposition (2 variables)
  • Industrial proximity (4 variables)
  • Green spaces (2 variables)

Results (MFA)

✓ The four first components explain respectively 30%, 15%, 13% and 11% of the total variance
✓ Major components interpretation:
  • 1st: air pollution and traffic proximity
  • 2nd: industrial proximity
  • 3rd: noise and green spaces

<table>
<thead>
<tr>
<th>Components</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Pollution</td>
<td>31.98</td>
<td>5.06</td>
<td>1.04</td>
</tr>
<tr>
<td>Noise</td>
<td>15.97</td>
<td>13.23</td>
<td>68.00</td>
</tr>
<tr>
<td>Industrial Proximity</td>
<td>0.13</td>
<td>78.34</td>
<td>5.13</td>
</tr>
<tr>
<td>Traffic Proximity</td>
<td>32.22</td>
<td>0.38</td>
<td>1.09</td>
</tr>
<tr>
<td>Green Spaces</td>
<td>19.71</td>
<td>2.99</td>
<td>24.74</td>
</tr>
</tbody>
</table>

Table 1: Contribution of each groups to the three first components (in %)

Results (HC)

✓ HC applied on the 10 first components of the MFA
✓ 5 cumulative exposure categories have been created using HC

Objectives

✓ Use data mining techniques to create a composite exposure index.
✓ Assess the environmental burden experienced by populations in their living environment.
✓ Illustrate this approach on a large French metropolitan area.

Methods

✓ Data mining technics, highlight underlying structures in data
✓ Multiple Factor Analysis (MFA); analyze variables by groups, give the same weight for each group and can include both quantitative and qualitative variables.
✓ Hierarchical Clustering (HC): minimizing the within-category inertia and maximizing the between-categories inertia.

Results (HC)

✓ HC applied on the 10 first components of the MFA
✓ 5 cumulative exposure categories have been created using HC

<table>
<thead>
<tr>
<th>Category:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO₂</td>
<td>≈</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>++</td>
</tr>
<tr>
<td>Noise</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>≈</td>
<td>+</td>
</tr>
<tr>
<td>Traffic</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>Industries</td>
<td>-</td>
<td>--</td>
<td>--</td>
<td>+++</td>
<td>--</td>
</tr>
<tr>
<td>Green spaces</td>
<td>++</td>
<td>+++</td>
<td>-</td>
<td>--</td>
<td>---</td>
</tr>
</tbody>
</table>

Table 2: Categories’ characteristics
(from +++ extremely higher than average to --: extremely lower than average, ≈: near the average value)

Conclusion and perspectives

✓ Data analysis technics can help to obtain insight about the different exposure profiles in an area with easily performed and interpreted tools
✓ This approach can help stakeholders to identify areas of higher “environmental burden”
✓ As a perspective, extend to other areas and indicators of living environment (public transport accessibility, health professionals density, primary good store availability …)