Structuring Complex Results using Network Maps and Hierarchical Charts

Lanters, Chloe Annelies; Fantke, Peter

Published in:
Procedia CIRP

Link to article, DOI:
10.1016/j.procir.2017.11.147

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
25th CIRP Life Cycle Engineering (LCE) Conference, 30 April – 2 May 2018, Copenhagen, Denmark

Structuring complex results using network maps and hierarchical charts

Chloe A. Lanters\textsuperscript{a}, Peter Fantke\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}Department of Management Engineering, Technical University of Denmark, Bygningstorvet 116, 2800 Kgs. Lyngby, Denmark

*Corresponding author: Tel.: +45 45254452, fax: +45 45933435. E-mail: pefan@dtu.dk

Abstract

Results from quantitative exposure and life cycle assessments are often complex, rendering their interpretation and communication to non-experts difficult. However, such data can be disaggregated and structured using visualization techniques to increase their interpretability. We present a simple, interactive tool that allows disaggregating data according to user preferences and flexibly visualizing these data in quantitative network maps and hierarchical column charts. We show in a case study on a chemical in flooring that our tool can help users to rapidly identify exposure hot-spots and trace back related pathways. Our tool can be applied to various types of results from chemical substitution, life cycle impact assessment, and high-throughput risk screening to improve decision support by better results interpretation and communication.

Keywords: quantitative exposure results; data visualization; chemical alternatives assessment; life cycle impact assessment; high throughput risk screening

1. Introduction

1.1. Understanding complex results for better decisions

Humans and ecosystems are exposed to various chemical substances contained in and used along life cycles of our many consumer products. Some substances may cause negative effects on human health or the environment, such as flame retardants, paraben preservatives, or pesticides [1-3]. The need to address the societal challenge of identifying and minimizing negative effects from releases of and exposures to potentially hazardous chemicals is therefore integral part of the United Nations Sustainable Development Goals (SDGs).

Aiming to address this challenge and providing related decision support for e.g. helping product designers to identify sustainable materials has resulted in developing different methodological frameworks. This includes methods to assess human and ecological exposures in the context of chemical alternatives assessment (CAA), life cycle impact assessment (LCIA), and risk-based high-throughput screening (HTS), as well as relevant for sustainability-focused strategies, such as striving toward a circular economy [4] and a non-toxic environment [5]. In addition, exposure-related decision support is required in frameworks targeting sustainable management of chemicals and international phase-out programs, such as the Stockholm Convention on Persistent Organic Pollutants [6], have led to replacing harmful chemicals with safer alternatives in many product applications [7,8]. Nevertheless, several substitution efforts resulted in only incremental rather than fundamental improvement, leading to shifting the burden from one type of impact (e.g. human toxicity) to another (e.g. groundwater contamination) or replacing chemicals with similarly harmful alternatives [1,8], which is due to the lack of understanding all relevant aspects and their contribution to exposure and related effects. In response, metrics and methods have emerged that quantitatively couple exposures across populations and pathways in a full life cycle perspective [9-12] with the aim to uncover related trade-offs, avoid burden shifting and target relevant exposure hot-spots along product life cycles. Such hot-spot results are ideally integrated with other impact categories, e.g. in a life cycle assessment (LCA) or product design context to ultimately provide comprehensive and science-based decision support for improved technologies.

However, interpreting and communicating rather complex...
results that consider multiple environmental media, exposure pathways and populations is challenging for practitioners, demanding at times the interpretation of mathematical models involving matrix or other complex structures [11,13-16].

As an illustrative example, we take a framework recently proposed to couple near-field (i.e. the direct vicinity of exposed workers or consumers; usually indoors) and far-field (i.e. exposures mediated via outdoor emissions) environments for different population groups on a consistent mass balance basis [11]. In this framework, product, near-field and far-field compartments, and humans are arranged in columns and rows of a matrix whose elements describe direct transfers of chemicals between compartments or to human receptors. Matrix inversion yields cumulative transfer and exposure estimates passed on to practitioners for interpretation (see [11] for details). While elegant and transparent in deriving such exposure results, practitioners might find it difficult to trace back pathways and transfers contributing to overall exposure. Hence, there is a practical need to visualize and communicate such results in a structured, simple and flexible way to accommodate different user perspectives and interests without requiring extensive expertise in quantitative exposure science.

1.2. Techniques to visualize complex data

Impact, exposure and similar data are commonly structured in tables, column, line, or pie charts, spider diagrams [17-19], or network diagrams for presenting model results; Sankey diagrams for visualizing flows; and tree maps for illustrating hierarchies [17,20]. Matrix, property, and cluster heat maps have been discussed to inform, respectively, decision makers and stakeholders in CAA [21], HTS [22], and LCA [17], where especially the latter help in contrasting large amounts of heterogenic data using hierarchical trees and color scales [23]. Multi-dimensional graphs are proposed to support LCA-based decisions using influence diagrams [24]. Finally, California’s Department of Toxic Substances Control (DTSC) proposes a qualitative (receptor) network map to visualize relationships between chemicals in consumer products and human and ecological receptors, including information on sources, exposure mechanisms, media and routes, and receptor groups [25]. Each map is specific to a particular scenario, relying on expert knowledge to manually construct relevant pathways.

All mentioned techniques are designed to support decision making involving quantitative data, yet, familiar and easy to interpret column charts are preferred by practitioners for visualizing quantitative results [26]. DTSC’s conceptual map could be valuable to complement column charts or other visualizations, but needs to be adapted to process quantitative exposure (or other relevant) data. Combining both easy to use charts and network maps is necessary to provide practitioners and decision makers with tools to more transparently interpret and more easily communicate complex assessment results.

1.3. Study objectives

To address this gap, our objectives are to (a) develop a tool of network maps and hierarchically nested column charts to structure and visualize complex assessment results at different disaggregation levels, and (b) test our tool in an illustrative example on a set of quantitative receptor and pathway specific exposure data to support decisions by e.g. product designers.

2. Methods

We construct a data set for an illustrative case study along which we develop and test our methods for structuring and visualizing complex information. As a starting point, we use results representing a set of cumulative exposure estimates of different population groups (adult workers, adult and children (<5 years old) consumers, and adult and children general population) via several exposure pathways (inhalation of air, gaseous and aqueous dermal uptake, and ingestion of drinking water, fish, and other food) and environments (near-field and far-field) to phenol (CAS 122-99-6) used as adhesive in wood flooring. We define that workers are exposed during 50 days of installing the floor, adult and children consumers are exposed over 10 years of living in a household with the installed floor, and the (general) population is exposed over 10 years to related emissions reaching outdoor environments. Quantitative exposure estimates are calculated by combining models for chemicals in building materials [27] and far-field emissions [28] structured in a matrix of Product Intake Fractions (PiF) linking chemical mass taken in by the different population groups to the mass originally contained in flooring [11]. We disaggregate our exposure estimates according to different categories, i.e. population and age groups, exposure environments and pathways. Overall cumulative exposure as sum over all contributing estimates is calculated to relate the contribution of each category (e.g. ‘exposure pathways’) or individual category items (e.g. ‘inhalation’ as a particular exposure pathway). Overall exposure is finally disaggregated at different levels of detail to contrast e.g. ‘inhalation’ and ‘ingestion’ pathways for particular population groups.

For structuring and visualizing exposure results in this case study, we first develop quantitative network maps to provide an overview of links between all categories and items. In a summary network map, we summarize all relations between categories by showing numerical values in connected boxes representing the items for each category (in columns), where the sum over each column equals the overall aggregated result. Box shadings indicate relative contributions to the overall result, while line thickness of connections indicates relative distribution of the overall result to items of a given category. Weights can be exaggerated by a scroll bar to further highlight differences between connections of similar magnitude. The user is able to select the order of the categories to explore the relationships between adjacent items.

To trace specific connections, e.g. focusing on exposure of children in the near-field environment, we develop a partial network map, where the user can progressively disaggregate the overall aggregated result following a single item per category column with flexible order of categories. The item with the highest value in its category is always disaggregated by default, although we allow for the manual selection of any given item. This network map yields the highest possible level of detail for any given dataset, while limiting its information to a set of user-selected items. Connection weights and box shading apply as described for the summary network map.

To provide an easy to understand overview of results at all given levels of detail and in flexible order, we finally develop column charts following the category order and item selection defined by the user for the partial network map. Each chart’s
3. Results

Exposure assessment results presented in our case study for phenoxyethanol in wood flooring are meant for illustration purposes only and do not indicate any actual risk for humans.

3.1. Summary network map

The summary network map for our case study is arranged according to user-defined order of categories, i.e. for example environment, population group, age group, and exposure pathway (Fig. 1). Values of all items per category sum up to an overall exposure of $6.78 \times 10^{-3}$ mg intake across population groups per mg phenoxyethanol in flooring. Items per category are arranged in decreasing order with highest values on top. In our case, adult consumers are exposed highest, namely via inhalation in the near-field (i.e. in households). Ranks and values of items per category, and connections between items change with chemical and scenario as a function of chemical properties and exposure duration per population group.

For the defined scenario, our summary network map communicates distributions of overall exposure within the different categories at the first disaggregation level, and quantitative exposure connections between items of distinct categories. Arranging items by decreasing magnitude and using color shading and scientific number formats allow users to quickly identify items with highest exposure per category (e.g. ‘adults’), and to compare exposure magnitudes across items and relative to overall exposure. Missing connections (e.g. ‘workers’ to ‘child’) indicate non-quantifiable relations. Sometimes, there is no model available to quantify a relation, which should be indicated by e.g. a dotted connection, to be clearly distinguishable from non-quantifiable connections. Existing connections show the contributions of individual items to items of another category, with variable line weights indicating relative differences in exposure magnitude, e.g. in our case adults show highest exposure as ‘consumers’ and ‘workers’, and to a lesser extent as general ‘population’.

3.2. Partial network map

We use partial network maps to complement the overall exposure distribution given in the summary map with detailed information on the distribution of specific items within any category, e.g. ‘far-field’, among items of other categories, e.g. ‘inhalation’, ‘food ingestion’ and ‘water ingestion’ (Fig. 2).

Partial network maps thereby visualize only part of the overall network of items according to users’ preferences (e.g. disaggregating ‘far-field’ exposure will only show inhalation and ingestion pathways, while ignoring dermal uptake in our example only relevant in the near-field (see Fig. 2a). This would also be seen when rearranging Fig. 1 in a way such that environments are linked to exposure pathways, where ‘near-field’ but not ‘far-field’ would have connections to ‘dermal’ pathways. While only showing part of the picture, partial network maps allow setting focus on specific points of interest or taking a particular perspective. In our example, the order of selected categories (in Fig. 2a starting from ‘age groups’ and then disaggregating to ‘population groups’ etc. compared to starting in Fig. 2b from ‘population groups’ and further going to ‘environments’ etc.) and manual selection of an item for further disaggregation (in Fig. 2a choosing the general ‘population’ even though it is not the population group with highest exposure, and in Fig. 2b choosing the ‘child’ age
group to see how children are exposed via different pathways, despite again not being the age group showing the highest exposure) can be flexibly selected to accommodate any user preference. By default, items with highest values within each category are automatically disaggregated further. Changes to the selection of category order have the effect of clearing all downstream items and categories, and returning the immediately upstream items to their default order. Clearing downstream categories is necessary, since categories cannot be disaggregated at more than one level; hence, such a selection requires a new category order from this point of disaggregation onwards.

While partial network maps can display more details than summary maps, the progressively more narrow definition of the population groups and other categories does not allow for cross-comparisons between all categories, e.g. only ‘far-field’ is shown in Fig. 2a as a result of disaggregating ‘general population’ exposure, while only ‘near-field’ is shown in Fig. 2b as result of ‘consumers’ exposure. Eventually, with selections made for every category, the resulting exposure is e.g. for one completely disaggregated population and age group that cannot be broken down further from the given input data. While connections in the summary network map only directly link two items of distinct categories, connections in partial network maps also relate to upstream linked items due to the progressive disaggregation. Items with zero contributions are not visualized and are also unable to be selected for further disaggregation. The default order of items (i.e. largest contributing item per category always topmost) is intended to organize the partial network map logically for easy interpretation by users. Arranging items according to their category contribution yet allowing users to further detail different items offers flexibility and a comprehensive level of disaggregation detail. However, caution should be taken when interpreting resulting partial maps as user-chosen items may not necessarily have the highest contribution to overall results within their respective categories.

### 3.3. Hierarchical column charts

In addition to network maps and to provide easy to use and familiar charts, we use nested 2D vertical unstacked column charts to represent quantitative results for given categories, ordered in a hierarchical way according to any user-defined category order (Fig. 3). Such a series of hierarchical column charts is attractive for immediate analytical use and as input for decision making [26,29,30]. We use column grouping rather than stacking to allow for a representation of quantities in logarithmic scale on all y-axes, which accommodates the structuring of results that typically span several orders of magnitude, such as exposure or toxicity results [31-33].

A distribution of overall result (in our case again overall cumulative exposure) at the first user-selected category is provided in the first chart (Fig. 3a). Here, we can see for our example how exposure is distributed within ‘pathways’ (with inhalation being highest) but aggregated over environments, population and age groups, corresponding to the left-most column of above network maps if the same category is chosen. The same category is used for disaggregation in the second chart (Fig. 3b), while using another user-defined category (in our example ‘population’ groups) to split the first category (‘pathway’) into separate columns (with ‘consumers’ as largest contributor to inhalation and ‘dermal’ pathways, where the sum of each column cluster equals the same item’s value in Fig. 3a). Columns in both charts sum up to the overall result, while providing different levels of detail.

![Hierarchical column charts](image-url)
but provides an additional disaggregation step (in our example the 'environment'), while maintaining the same column color code in both charts (e.g. 'consumers' in blue in both Fig. 3b and 3c). In our case, Fig. 3c provides insight into the exposure environment of each population group, highlighting the relevance of near-field exposure in this particular scenario. Finally, Fig. 3d further details the third-level category for a user-specified item, in our example distributing 'far-field' population exposure across age groups. Additional details in Fig. 3c and 3d also come with a loss of information in these charts for the items not selected from the higher levels, highlighting that the last two charts do not necessarily show details for the largest contributing items to overall results.

4. Discussion

4.1. Applicability and limitations

Structuring and visualizing complex data according to user-specified categories helps to build trust in the data at hand by providing at the same time both an easy-to-understand overview of the data and a disaggregation of the data beyond which they cannot (easily) be further detailed. This allows the identification of important contributors or focus on areas of specific interest for the user.

While some studies have advocated to go beyond column charts for visualizing complex results from exposure, risk or sustainability assessments [19,34], hierarchical column charts allow to meaningfully conveying quantitative information in a structured way when focusing on individual chemicals, life cycle stages, impact categories or scenarios, if they are systematically nested in a way to represent different levels of information detail (hierarchical column charts) or if they are organized to provide an overview of important connections (summary network map) or pathways (partial network maps). For example, contrasting exposure levels in our example for different population groups for each exposure pathway separately (Fig. 3b) allows identifying from a single chart the predominantly exposed group per pathway and at the same time shows the relative importance of all pathways across population groups. If this level of detail is not required, population groups can be summed per pathway (Fig. 3a), while additional details for a particular pathway (Fig. 3c) or population group can be presented in a subsequent chart. Hence, our set of hierarchical column charts (or only a subset of these) can be applied in various decision-making contexts and assessment frameworks, where quantitative results need to be interpreted or communicated to other stakeholders. However, this representation of quantitative data also comes with an important limitation. The representation of the data through aggregating exposure (or any other) results in column charts does not provide any information on uncertainty. Relative differences in column heights should be interpreted with care, regardless whether results are presented in log scale or not and regardless whether results represent relative or absolute values.

Using comparative metrics for expressing quantitative data in CAA, LCIA and HTS studies, including exposure matrix data used in our illustrative case study, allows contrasting different aspects at distinct levels of detail, such as in our example contrasting population groups, exposure pathways and environments. Despite the intrinsic complexity, the visualization of the raw information as interactive pathways and hierarchical charts is intended to make the results easy to interpret and communicate to others. Access to an easily used and interpreted data structuring and visualization tool can support practitioners and decision makers contrasting scenarios and better understanding the data they are working with. The first step, however, is always that the user identifies a suitable set of categories and items that should be applied to structure and visualize the data at hand. This can be challenging if the data are reported using loose or inconsistent terminology. Hence, while the application of our tool has been demonstrated in our case study by linking it to the results of an exposure matrix, any data reported in disaggregated units can be used as input as long as different aggregation levels or connections between categories are possible. Thereby, our tool allows the use of any absolute or relative metric and unit as long as data remain comparable across levels of detail.

Our network maps and column charts can complement higher level screening frameworks that compare for example different chemicals or products by focusing on specific pathways or parameters, such as heat maps as demonstrated by e.g. [22]. While in such cases other frameworks show the differences between substances, our visualization tool provides a high level of detail across pathways and processes for one chemical, meaning that it can be implemented at a more detailed stage of assessment and allows the comparison of exposures within a particular product application scenario. Although only one chemical or product can be visualized at a time, the insight into key drivers and pathways is valuable.

Another advantage of our tool is that four disaggregation levels with a maximum of four items per category were shown in the example application; however, a larger number of both items and categories can be incorporated into the structure with manual changes to the data input structure defined by the user. However, in some cases it is necessary to adapt the visualization structure to specific input data structures, which requires manually adapting formatting and formulae. This is a limitation of our current tool and further development could focus on implementing an automatized check of the data to be used, determining the number of items and categories required to be shown and adapting the data structure accordingly. Removing the need for manual changes can vastly improve the applicability of the tool and reduce the complexity of implementing new scenarios.

4.2. Future research needs

Our network maps and hierarchical column charts should be tested and adapted to the requirements of different LCIA, CAA, and HTS application contexts. To then further improve transparency and increase confidence in quantitative results used in decision support, it is important to also visualize the uncertainty of complex data in network maps and column charts at hand. This requires reporting uncertainty ranges, which can usually be obtained via different techniques in the different assessment frameworks, along with nominal results. In column charts, e.g. error bars can represent uncertainty intervals, disaggregated according to the shown level of detail.
if available. In network maps, uncertainty can be represented by separate boxes adjacent to the nominal results, perhaps with a shading scheme to indicate the relative magnitude.

5. Conclusions

Our data structuring and visualization tool can provide users in CAA, LCIA, and HTS as well as other fields using quantitative data with an automated and visually appealing way of structuring complex results from e.g. quantitative exposure assessment, requiring only minimal user input while offering flexibility on the level of detail and the desired user focus. Furthermore, if results are clearly structured, less time is required to compare scenarios and identify most problematic or most desired solutions as input to improved decision support for expert and non-expert users.

Acknowledgements

This work was supported by the Marie Curie project QuanTox (grant agreement no. 631910) funded by the European Commission under the Seventh Framework Programme.

References


Appendix A. Visualization tool in spreadsheet format

Our visualization tool is available free of charge upon request from the authors (corresponding contact: pefan@dtu.dk).