How to perform mixture analysis on PARC HBM data in MCRA

Jasper Engel, WR-BIOM, 26-12-2022

Aim of this note

To describe how mixture analysis on PARC HBM data can be carried out by multivariate methods in the MCRA web system.

Introduction

In PARC project Real-life mixtures, data from HBM studies are analyzed for co-exposures that might be relevant for mixture risk assessment. This is done by multivariate methods which have been implemented in the MCRA system for probabilistic risk assessment (https://mcra.rivm.nl, see also van der Voet et al., 2020). In the current version of MCRA two mixture analysis procedures that were previously used in the Euromix and HBM4EU projects are implemented, namely network analysis and the SNMU method.

Network analysis of chemical correlation patterns in HBM data using network analysis methods was explored in the HBM4EU project (Ottenbros et al. 2021). A node in the network represents a substance, and an edge reflects the partial correlation between a pair of substances. Presentation of biomonitoring data in networks may help to interpret exposure patterns. In addition, substances may be grouped in "communities" based on the network topology, i.e. to group highly co-exposed substances.

The sparse nonnegative matrix underapproximation (SNMU) method was applied to select mixtures of chemicals from dietary co-exposures (Traore et al. 2018, Crépet et al. 2019) and breast milk contamination (Crépet et al. 2022). The approach is a combination of SNMU to select mixtures and estimate the exposure of individuals to these mixtures. Subsequently, cluster analysis is applied to group individuals based on their mixture exposure profiles. The approach originates from the use of nonnegative matrix factorization (NMF) to identify food consumption patterns and main mixtures of chemicals to which the French population was exposed (Bechaux et al 2013, Traore et al. 2016). In Euromix, it was found that SNMU (Gillis and Plemmons 2013), a method closely related to NMF, allows for control of the number of substances that are selected in a mixture (the sparsity of a mixture). Experiments were made with SNMU analysis of standardized exposures, and risk-based mixture selection (Crépet et al. 2019). Risk-based mixture selection weights the exposure per substance by their relative potency factors (RPF) to obtain results that are directly relevant for risk assessment. With missing RPF information, an alternative approach is to focus on stndradized exposures. Standardization is applied to avoid that subsyances with high levels and variation overwhelm the analysis. The implementation of the SNMU method in MCRA allows both risk-based mixture selection and standardized exposure-based mixture selection. In the first year of PARC, the focus is on standardized exposures because for substance groups other than pesticides the relative potency information is often missing or incomplete.

Network analysis and SNMU have been implemented in MCRA similar to the procedures described in Ottenbros et al. 2021 and Crépet et al. 2022. The steps outlined in this guideline are subject to change.

Procedure to run multivariate mixture analysis models in MCRA

It is assumed that you have an MCRA account. If not, register on https://mcra.rivm.nl. Registrations must be approved, so this may take some time.

Annex to this document, three data files are distributed:

1. Example study population.xlsx
2. Example_data_BasicCodebook_v2.0.xlsx

It is assumed that these files are available to you.

Steps

1. The data files are prepared for the case study of the FLEHS study and contain several tables as suggested by VITO. For your own data, please prepare these in the same format, see "Guidance how to use PARC HBM data in MCRA.docx" for more details.
2. Upload the data to MCRA:
    a. Login to MCRA at https://mcra.rivm.nl
    b. Go to Data. Click on your personal folder (labelled with your MCRA user name).
    c. Press the green '+' button in the lower right corner and choose 'Upload new file(s)'. Select the two data files above and upload them to MCRA.
3. Create an MCRA Exposure mixtures action:
    a. Go to Workspaces (Click All workspaces button top right in the blue bar). Press the green '+' button in the lower right corner to create a Workspace, e.g. 'Mixture selection'.
    b. In the workspace, press the green '+' button in the lower right corner to create an action
    c. Click 'Show all action types', and then scroll down and select Exposure mixtures
    d. Choose an appropriate name for your action (e.g. 'PARC HBM Mixtures') and click Next
    e. Choose Risk type 'Chronic'. This way, MCRA will calculate averages per individual person over all available days per person in the HBM data. Set Target level to 'Internal'. This will trigger MCRA to target calculations at the internal level, e.g. in an organ or in body fluids like blood or urine; any information available at external level will be translated to internal level using kinetic models). Next set Internal concentration to 'Human monitoring concentrations'. Finally set substance weighting to 'Standardised'. Standardized exposures have variance 1, which is of relevance for SNMU analysis.
    f. Click Create. Now the action is created but still has to be linked to the input data.
4. Link data:
    a. Click on Human monitoring data, in the list shown in the left panel (green section). Click on the pencil icon to select a Human monitoring data source. Browse to the folder with your uploaded data and select 'Example_data_MCRA_2022-09-08.xlsx' (or your own adapted version). At the bottom of the pop-up, choose 'Toggle all'. This will load all available data tables in the file. Then click Select.
    b. In the Human monitoring selection settings, choose the survey/study (PARC HBM sample data in the example file), and a sampling method (e.g. Urine (spot) in the example file). Then click Save changes (red button).
    c. [Optional] Click on Substances (purple section). Click 'set filter to select the substances of interest. Click Save. Then click Save changes (red button).
    d. Click on Populations (purple section). Select 'Use data' and click on the pencil icon to select a Populations data source. Browse to the folder with your uploaded data and select 'Example study population' (or your own adapted version). Then click Select.
5. Choose action settings for imputation and mixture analysis
    a. In the Human monitoring analysis module
        i. Choose setting how to handle censored values (e.g., as 0.5 * LOR, where the LOR (limit of reporting) means LOQ if that is provided, or else LOD.

ii. Also choose how to impute any missing values. Impute from data means that for each missing value a random value from the available measurements per substance is selected.
iii. Then click Save changes (red button).

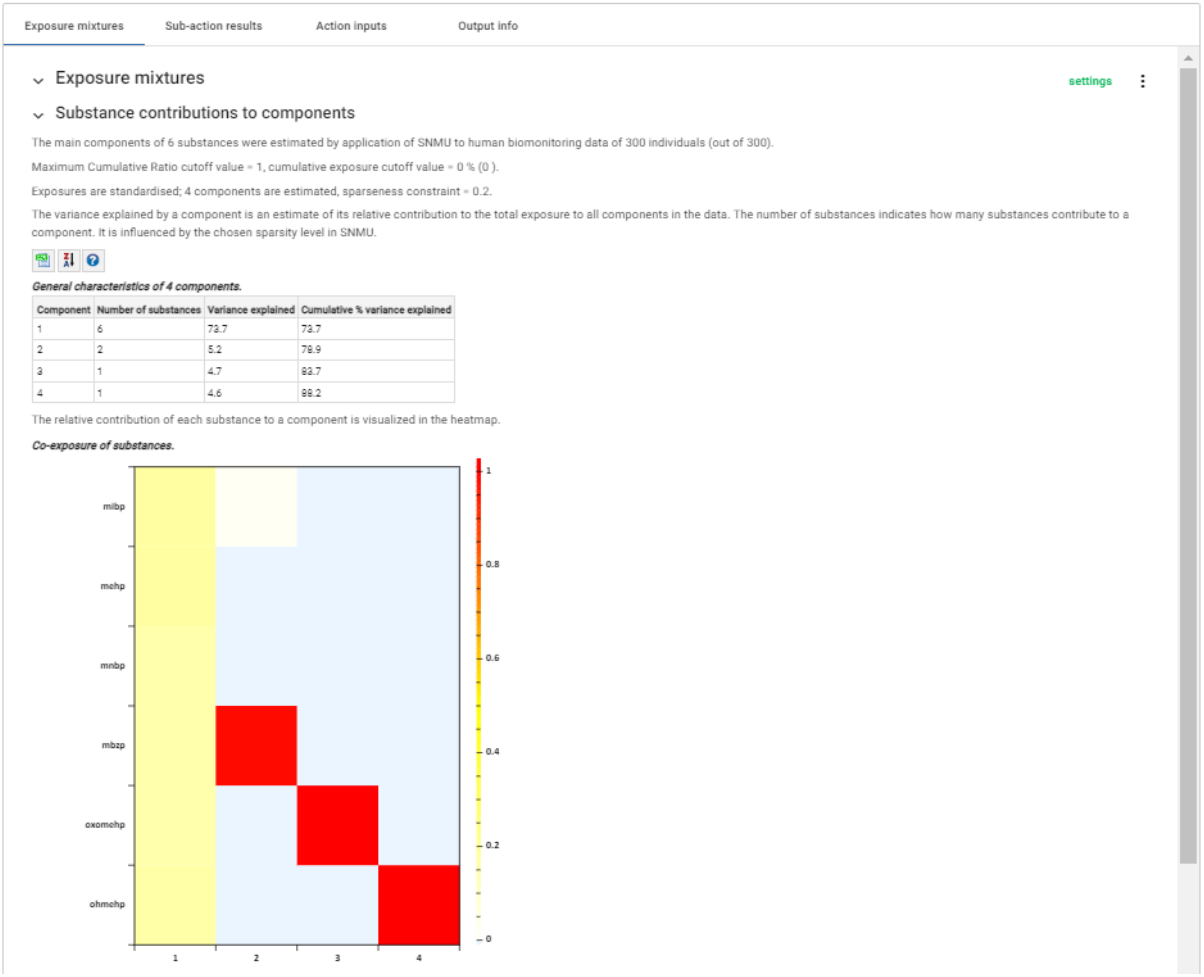
b. Click on Exposure mixtures. In mixture analysis settings choose how to carry out SNMU analysis.
i. Set Method to 'SNMU + hierarchical clustering' to first select mixtures by SNMU and subsequently identify population sub-groups with similar exposure to the mixtures.
ii. Optionally, adapt the SNMU settings for 'SNMU: number of mixtures', 'Sparseness parameter', and 'Iterations' (e.g. 4, 0.2 and 1000 in this example).
iii. Then click Save changes (red button). Note: you may have to scroll upwards to see the red button.
iv. Subsequently specify how many population sub-groups to identify by selecting 'Determine number of clusters automatically'. Then click Save changes (red button).

c. In mixture analysis setting select 'Apply network analysis' under network analysis type to carry out network analysis with community detection, in addition to SNMU analysis. Then click Save changes (red button). Note that the graphical lasso is implemented for network estimation and the walktrap algorithm is applied for community detection following the approach in Ottenbros et al. 2021.

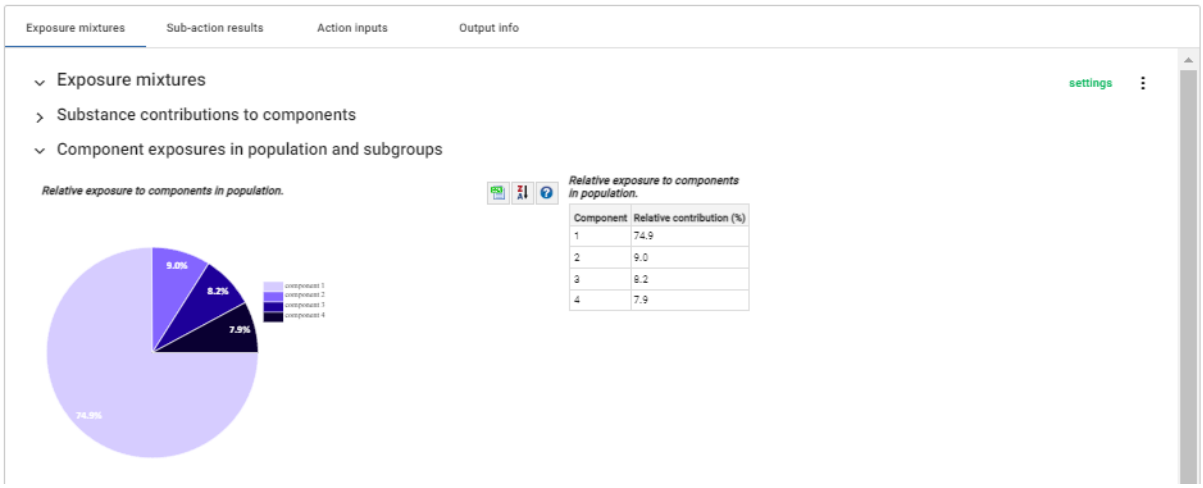
6. Run the action:
a. Click the Run button, triangle in the grey horizontal action bar at the top of the screen.
b. You are transferred to the Results screen. Wait for the run to finish. Meanwhile, you can change the name of the output if you want by clicking the pencil symbol.

7. Inspect the results:
a. When ran to completion, open the output by clicking the name of the output.
b. Click on 'Substance contributions to mixtures' and 'Additional details' for a figure that has been proposed to determine the optimal number of mixtures that should be selected (Béchaux et al. 2013, Mancini et al. 2021, Sy et al. 2013, Zetlaoui et al, 2011). The analysis can be adapted based on this result by changing the number of mixtures in step 5b. Additionally, inspect the characteristics of the components to check for convergence of the SNMU algorithm. The number of iterations listed in the table should be smaller than the value set in step 5b. If this is not the case, the maximum number of iterations should be increased in 5b.
c. Click on 'Substance contributions to components' to show the SNMU mixture selection results. A table with statistics for each mixture such as number of substances involved and explained variance is given. Additionally, the relative contribution of each substance to the mixture is presented in the heatmap. The pie-chart and table in the subsections per mixture offers more detailed information about the relative contributions of substances.
d. Click on 'Component exposures in population and subgroups' to study the (averaged) relative exposure of individuals in the population and sub-populations to the selected mixtures.
e. Click on 'Exposure by substances in (sub)population' to compare exposures in population sub-groups to the whole population at the level of the individual substances. The tables for the subgroups report the output of a t-test for a difference in average exposure (on a log-scale) in that cluster compared to the average in the other clusters. The significance reported by this test may be overoptimistic because the same data was used twice (for clustering and significance testing).
f. If a network analysis has been requested a carrot with 'Network analysis' will be visible. Here, the network is visualized. Communities of substances (here taken as mixtures) are indicated by the same color.

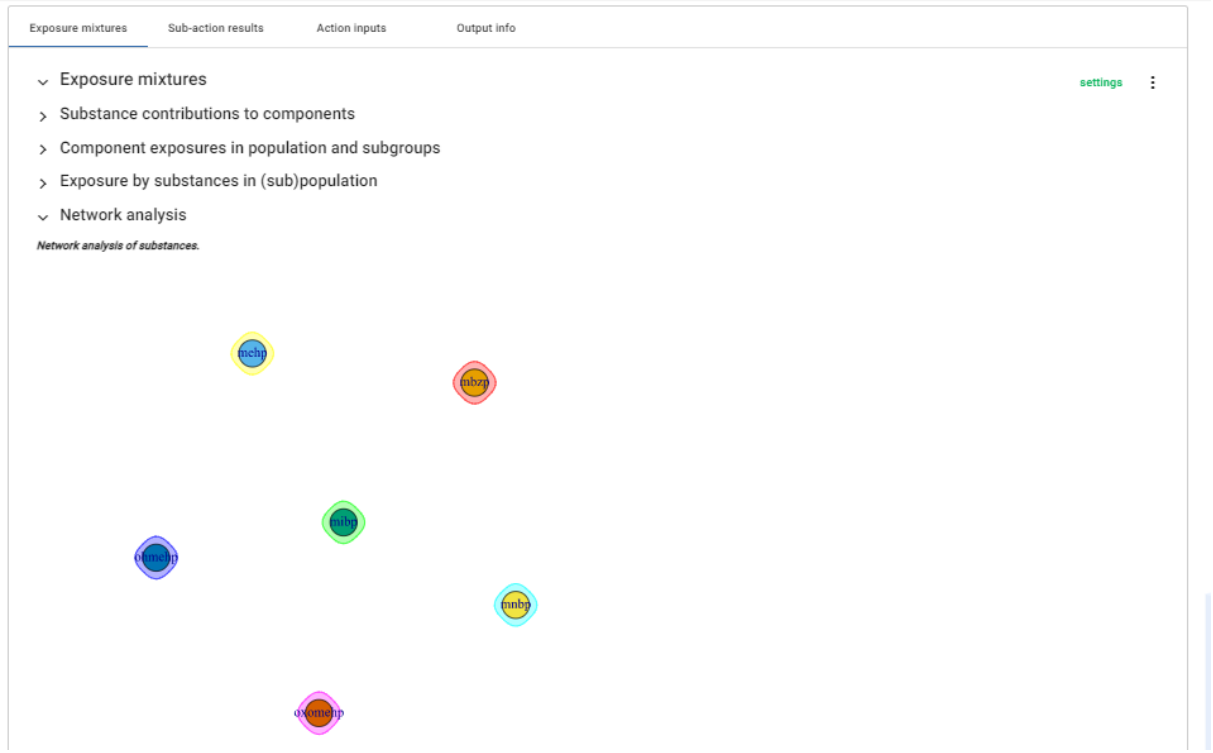
Example: SNMU analysis of urine concentrations of phthalates and some other substances in a population of Flemish teenagers (simulated data)



In this example, SNMU has been run on standardized data, with 4 mixtures and a sparseness constraint of 0.2. Censored values were imputed with 0.5 times the LOR and missing values (if any) were imputed from data. The first mixture consists of 20 substances and explains 71.6% of the variation in the exposure matrix. The heatmap shows that all substances contribute almost equally to this mixture. This is a consequence of the simulation procedure that was used to generate these data. In this case the first component roughly describes the average exposure for each substance. The data will be adapted later to include mixtures with contributions from a (more) limited number of substances.



SNMU returns for each individual the exposure to each mixture, from this a relative exposure can be calculated for that individual by dividing the exposure to a mixture by the total. The pie chart and table show these relative exposures, averaged across all individuals. The average relative exposure to mixture 1 is 43.3%, that to mixture 2 17.9%, etc.



The network analysis groups each substance in a separate community. This is expected because in this simulated data set exposures for substances were generated independently.

References

Béchaux, C., Zetlaoui, M., Tressou, J., Leblanc, J. C., Héraud, F., & Crépet, A. (2013). Identification of pesticide mixtures and connection between combined exposure and diet. *Food and chemical toxicology*, *59*, 191-198.

Crépet, A., Vasseur, P., Jean, J., Badot, P. M., Nesslany, F., Vernoux, J. P., ... & Mhaouty-Kodja, S. (2022). Integrating Selection and Risk Assessment of Chemical Mixtures: A Novel Approach Applied to a Breast Milk Survey. *Environmental health perspectives*, *130*(3), 035001.

Gillis, N. and Plemmons, R. J. Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis. Linear Algebra and its Applications, 438(10):3991–4007, 2013.

Mancini, Francesca Romana, Pauline Frenoy, Thibault Fiolet, Guy Fagherazzi, and Amélie Crépet. 2021. 'Identification of Chemical Mixtures to Which Women Are Exposed through the Diet: Results from the French E3N Cohort'. Environment International 152:106467. doi: 10.1016/j.envint.2021.106467.

Ottenbros, I., Govarts, E., Lebret, E., Vermeulen, R., Schoeters, G., & Vlaanderen, J. (2021). Network analysis to identify communities among multiple exposure biomarkers measured at birth in three Flemish general population samples. *Frontiers in public health*, *9*, 590038.

Sy, Mouhamadou Moustapha, Max Feinberg, Philippe Verger, Tangui Barré, Stéphan Clémençon, and Amélie Crépet. 2013. 'New Approach for the Assessment of Cluster Diets'. Food and Chemical Toxicology 52:180–87. doi: 10.1016/j.fct.2012.11.005.

Traoré, T., Béchaux, C., Sirot, V., & Crépet, A. (2016). To which chemical mixtures is the French population exposed? Mixture identification from the second French Total Diet Study. *Food and Chemical Toxicology*, *98*, 179-188.

Traoré, T., Forhan, A., Sirot, V., Kadawathagedara, M., Heude, B., Hulin, M., ... & Crépet, A. (2018). To which mixtures are French pregnant women mainly exposed? A combination of the second French total diet study with the EDEN and ELFE cohort studies. *Food and Chemical Toxicology*, *111*, 310-328.

Zetlaoui, Mélanie, Max Feinberg, Philippe Verger, and Stéphan Clémençon. 2011. 'Extraction of Food Consumption Systems by Non-Negative Matrix Factorization (NMF) for the Assessment of Food Choices'. Biometrics 67(4):1647–58. doi: doi: 10.1111/j.1541-0420.2011.01588.x.