Source Analysis of Dioxins in River Water using Non-negative Matrix Factorization

Ohtsuka N[1], Minomo K[1], Hashimoto S[2]

[1] Center for Environmental Science in Saitama, 914 Kamitanadare, Kazo, Saitama 347-0115, Japan
[2] National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan

Introduction

Non-negative matrix factorization (NMF) is the approximation of $n \times m$ non-negative matrix V by a product of $n \times r$ non-negative factor matrix W and $r \times m$ non-negative load matrix H [1]. Applying NMF to concentration data of multiple compounds in environmental samples is expected to obtain the profiles and contributions from pollution sources. However, the studies demonstrated NMF analysis to environmental pollutions are very limited. To verify the advantage of NMF analysis for environmental pollutions, in this study, we applied NMF to the dioxin concentration data for river water in Japan.

Materials and methods

The data set used in this study consists of concentrations of 29 congeners for dioxins (PCDDs, PCDFs and DL-PCBs) in 12 water samples from Ayase River, which flows in Saitama Prefecture, Japan, collected once a month from April 2003 to March 2004 [2]. Concentrations of dioxins are usually evaluated in terms of toxic equivalent (TEQ). Therefore, the data set was converted into the TEQ compositions on the basis of WHO-2006 TEF as a matrix V for NMF calculation. The matrices W and H obtained by NMF calculation correspond to TEQ compositions of pollution sources and to TEQ-based source contribution rates in the samples, respectively. The Kullback-Leibler divergence was used as the objective function of NMF calculation [1, 3]. An NMF calculation was performed with the algorithm proposed by Hsieh [3]. Furthermore, the results were confirmed its convergence using the algorithm proposed by Lee and Seung [1]. Since the initial values of the matrices W and H are random numbers, it is rare to obtain the most appropriate result with a series of NMF computations. For this reason, NMF calculations were executed 50 times for each case where the factor number r was 1 to 11, and the result with the smallest objective function value was adopted for each number of factors.

Results and discussion

Estimation of the number of pollution sources

Figure 1 shows the relationship between the number of factors, r and the objective function value obtained by NMF. The objective function value drastically decreased between giving one-factor and two-factors and then monotonously decreased with increasing the r. This profile indicates that mainly two kinds of pollution sources influenced on the concentrations of dioxins in the Ayase River.

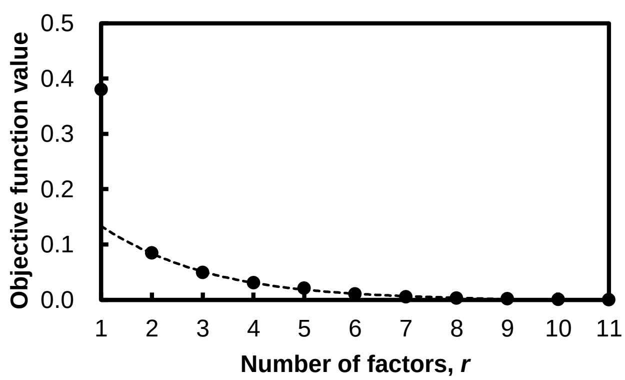

Fig. 1 Plots of the objective function values vs. number of factors, r

Profiles and contributions of pollution sources

The matrix *W* obtained by NMF corresponds to the TEQ compositions of pollution sources. The congener profiles in TEQ resulting from two pollution sources are shown in Fig. 2. The obtained matrix *H* corresponds to the TEQ-based contribution rates for the pollution sources in each water sample. The contributions of the two pollution sources to the total TEQ are shown in Fig. 3. Annual contribution of the pollution source #2 (Factor #2) was particularly high in the irrigation period from May to August, and then gradually decreased during from September to March which corresponds to the non-irrigation period. The predominant congeners of the source #2 were 1,2,3,7,8-PeCDD and 1,2,3,4,6,7,8-HpCDD (Fig. 2), which contained as impurities in CNP and PCP formulations, respectively [4]. PCP and CNP formulations had been widely used as herbicides in paddy fields in Japan. From these facts, the pollution source #2 is considered to be dioxins discharged from paddy field via irrigation water. The TEQ of the pollution source #1 (Factor #1) was relatively small through the year.

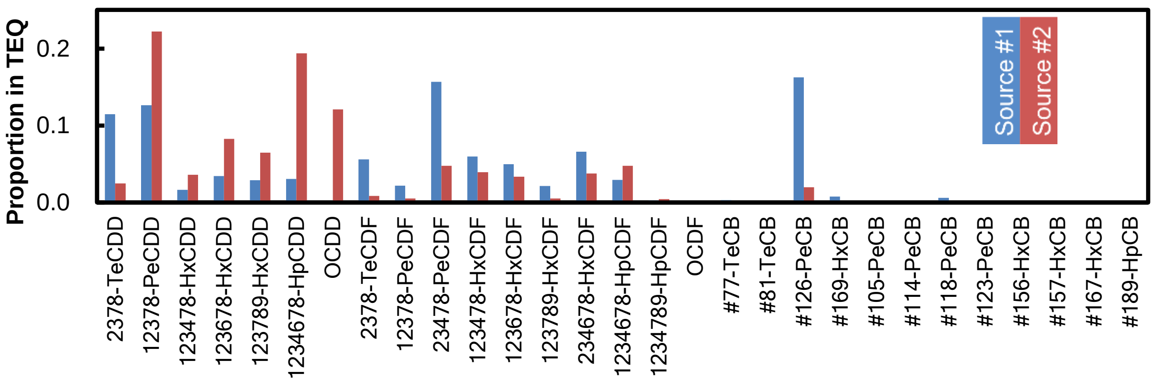

Fig. 2 TEQ based congener profiles in the two pollution sources estimated by NMF

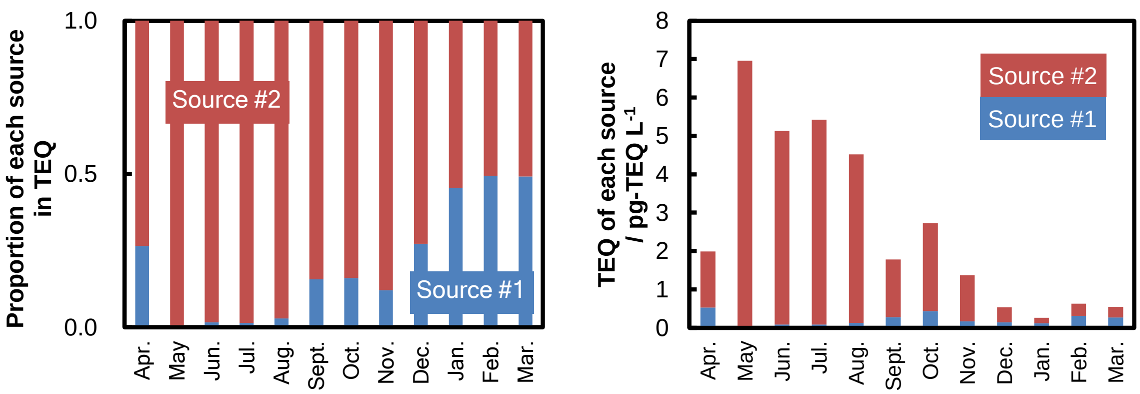

Fig. 3 Annual contributions of the two pollution sources estimated by NMF to TEQ

Pollution origins

Dioxins in Japanese environment are mainly influenced by waste incineration (combustion), PCB products, PCP formulations and CNP formulations. In previous studies, we proposed a method to estimate the TEQs of these four pollution origins based on the concentrations of five-indicative congeners: 1,2,3,7,8-PeCDD, 1,2,3,4,6,7,8-HxCDD, 2,3,4,7,8-PeCDF, #105-PeCB and #126-PeCB [2, 5], namely 'the indicative congener method'. This method has been applied to identify specific dioxin-sources in the environmental samples in Japan [2, 6, 7]. The results of contribution analysis to the pollution sources #1 and #2 by the indicative congener method are shown in Fig. 4. The sum of the TEQs from the four pollution origins was found to be almost one, which indicating that the congener profiles estimated by NMF can be explained by the four-origins. The main TEQ contributor to the source #2 was PCP formulation, followed by CNP formulation, and combustion. Therefore, the pollution source #2 was considered to be dioxins transferred from paddy fields to the river. This was consistent with the result discussed above. On the other hand, pollution source #1 is considered to be derived from dioxins in atmosphere flowing into the river by runoff water, which has a high proportion of TEQ originating from combustion, followed by PCB products.

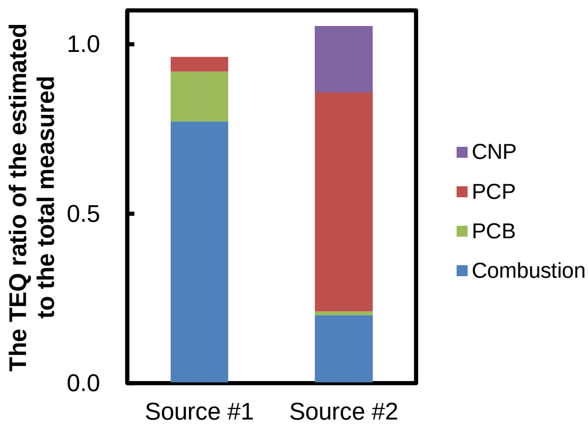

Fig. 4 The TEQ ratios of four origins estimated by the indicative congener method to the two pollution sources obtained by NMF

We collected 61 bulk (dry and wet) deposition samples from February 2012 to February 2013 in Kazo city, Saitama Prefecture, and the samples were analyzed for dioxins [6]. The TEQ based contributions in the deposition samples and the pollution source #1 are shown in Fig. 5. The two profiles were generally similar except a few congeners having high TEFs, together with low concentrations, such as 2,3,7,8-TeCDD, 1,2,3,7,8-PeCDD and #126-PeCB. This is consistent with the above speculation that pollution source #1 was due to dioxins in atmospheric deposition into the river via rain.

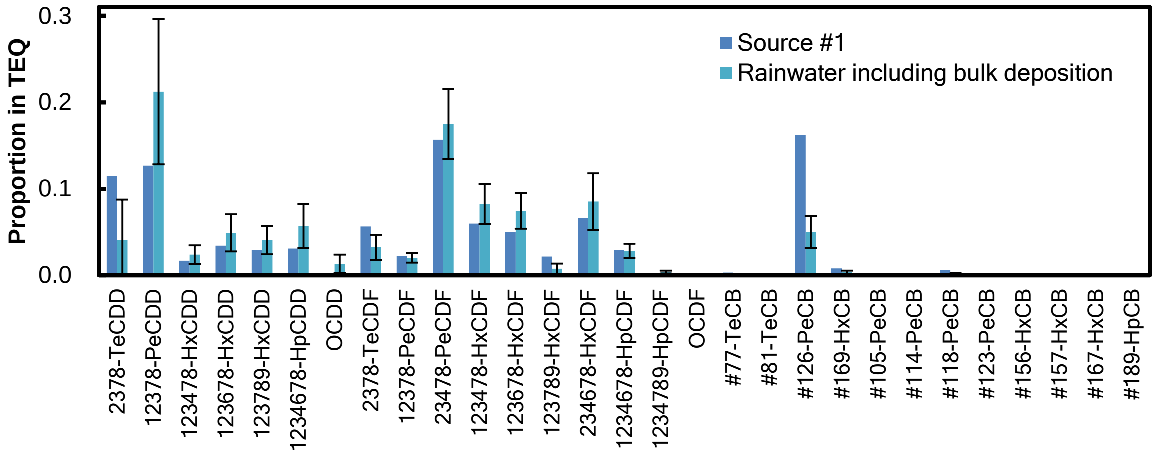

Fig. 5 TEQ based congener profiles in the pollution source #1 and rainwater including bulk deposition samples
The error bars represent standard deviation

In addition, NMF analysis was carried out for dioxin concentrations in other river water samples (Niigata River, Saitama Prefecture), which were collected in the same period to the Ayase River. We also found that the number of pollution sources (factors) and congener profiles for pollution sources obtained for Niigata River were similar to those of Ayase River.

As described above, both NMF and the indicative congener methods are effective as source analysis of dioxins, and have different characteristics of the acquired information. The indicative congener method gives information concerned with the four pollution origins. On the other hand, NMF method gives information concerned with pollution sources including unknown origins. NMF can be also applied to the environmental samples that are strongly influenced by unknown origins other than the four origins.

Acknowledgements

This study was supported by JSPS KAKENHI Grant Number JP26241026 and JP15K00573.

References

1. Lee DD and Seung HS (1999) *Nature*, **401** 788-791.
2. Minomo K, Ohtsuka N, Hosono S, Nojiri K and Kawamura K (2011) *Chemosphere*, **85** 188-194.
3. Hsieh CJ and Dhillon IS (2011) *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1064-1072.
4. Masunaga S, Takasuga T and Nakanishi J (2001) *Chemosphere*, **44** 873-885.
5. Minomo K, Ohtsuka N, Nojiri K, Hosono S and Kawamura K (2010) *Chemosphere*, **81** 985-991.
6. Minomo K, Ohtsuka N, Nojiri K, Matsumoto R and Kohata K (2014) *Organohalogen Compounds*, **76** 838-841.
7. Kishida M (2013) *Journal of Environmental Sciences*, **25** 441-451.