

## ***Interactive comment on “Application of Factor and Cluster Analyses to Determine Source–Receptor Relationships of Industrial Volatile Organic Odor Species in a Dual-Optical Sensing System” by Jen-Chih Yang et al.***

**Jen-Chih Yang et al.**

renayang@itri.org.tw

Received and published: 11 June 2019

Response to Interactive discussion: 'amt-2018-344', Anonymous Referee #1, 15 April 2019

Q1: General comments: The simultaneous measurement of different gases is very suitable for the “Factor analysis” and the determination of different emissions profiles produced by different processes; the “Cluster analysis” as well as the meteorological data confirm these findings. This approach which allows to identify the different

C1

sources and determine contributions to the odor quantitatively is interesting, meets the scope of the journal, is in general well written and therefore suitable for publication in AMT after some revision in more detailed described below and after providing some more detailed information. The strength of the article is the nice and clear concept using FTIR measurements which allows for the measurement of 16 species simultaneously, measure “emission profiles “ of these species at the potential sources using cell measurements and detect these emission profiles at the receptor site using open path measurements and compensate the dilution using an increased path length. The statistical methods are chosen in a proper way and the results are clear. In addition the study also is able to determine different industrial processes, which are occurring at the three odor producing sites, and the meteorological conditions, which confirm the identification of the origin of the odor of different events. The structure of the article might be logical, but the references to the figures in the text is sparse and might be missing sometimes and the use of T1 and TA1 as well as Figure 1 y Figure A1, without adding an Appendix with a text document is confusing and make it a bit difficult to understand the article. Therefore work is required to improve the manuscript and its readability before publication. 1.) Two set up are used but to my understanding it would work also with just one FTIR, as the measurements at the potential sources are realized independently and also the open path measurements at the two sites do not have to be simultaneously.

Reply: Thanks for the comment. Yes, the measurement at emission sources would work with only one set of CC-FTIR. Also the OP-FTIR measurement does not have to be operated simultaneously at the two sites as well.

2.) Factor and cluster analysis are two independent methods: Are they used to confirm each other, compliment some aspects or as “combined” analysis

Reply: Thanks for the comment. Factor and cluster analysis are used to confirm each other. The concurrent trends between different species measured by the CC-FTIR can be analyzed using both factor and cluster analysis. In order to gain insight into the

C2

underlying emission source characteristics, odor contaminants with concurrent patterns were grouped together as a factor. Cluster dendrograms provide linkage paths between groups of chemicals to offer more information about the characteristics of different emission sources.

3.) Which role-plays the meteorology in the analysis? The meteorological data might also be used in the cluster analysis or are only used to confirm the other method in the different cases.

Reply: Thanks for the comment. A meteorological station was operated simultaneously with an OP-FTIR system to collect continuous wind data that can enable identification of the incoming direction of different odor contaminants, and provide spatiotemporal measurement of odor pollutants. In other words, the meteorological data was used to confirm the factor analysis in the way that the incoming direction of each factor (representing a group of chemicals) may be different according to the locations of each potential odor sources.

Q2: 0. Abstract : Line 6 page 2 “Continuous monitoring” alone is a bit misleading, if you talk about both receptor site and source emissions, because it implicitly say “simultaneously”, which is not the case. Maybe you can add just the periods where sources and where receptor sites are “measured continuously”. Or just add “during different periods”

Reply: Thanks for the comment. The phrase “continuous monitoring” in Page2 (Line 6) was rewritten as “Both receptor and source monitoring data were collected to characterize the emission sources of various odorous substances”.

Q3: 2. Materials and Methodology (2.1 Site description and sampling techniques): 1.) Page 4, line 14-page 5 line 19. I would recommend to separate “Site description” line 16-26 and “sampling techniques” line 27-19. Please could you add a map with all sites and distances?

C3

Reply: Thanks for the comment. Page4 (Line 16-26) has been separated from Page4 (Line 27) -Page5 (Line 19), and the subtitles were renamed subsequently: 2.1 site description; 2.2 sampling techniques; 2.3 chemical analysis methods; and 2.4 qualitative receptor modeling. A map with all sites and distances was added in Figure 2.

2.) Measurement method: the end of the introduction already gives description of the method, for me that is ok, but please add there or here the very important information about the path length in the “Closed Cell” the used pressure in the cell and an estimation of the gas flow through the cell. Did you try to reduce the water vapour to decrease interference with H<sub>2</sub>O absorption or is it not necessary.

Reply: Thanks for the comment. The following sentence has been added to Page 5 (Line 14-18) –“A 10-m (path length) gas cell with the inner pressure of 720 mm Hg, and an estimated gas flow rate of 0.37 Liter/Sec. was used for the CC-FTIR multi-reflection gas measurements. The water vapour was mostly removed by using an impinger connected to the inlet of the gas cell to decrease interference with H<sub>2</sub>O absorption in the FTIR spectra”.

Q4: 2. Materials and Methodology - 2.2 Chemical analysis methods: 1.) There is no always a link to the figures and tables, but from the table 1 it can be seen that you found 17 relevant species, this is more interesting than the list of species in the library. I would recommend move the parts of the description of the technique from the Instrument manual to the introduction (e.g. more than 300 species) and the concrete chosen settings and the indeed used species in the method- section. It would be nice to get the complete set of 17 micro windows, where the 17species are retrieved also in a Table as Figure 3A has very small characters at the x axes and shows only 16 species.

Reply: Thanks for the comment. The following sentences – “The IR “fingerprints” of over 300 compounds were established on the basis of information from the US Environmental Protection Agency (USEPA) and the FTIR software developers” has been

C4

moved from Page 5 (Line 26-28) to the introduction section on Page 3 (Line 29-30). The following sentences were added to the method section – “The unique fingerprint characteristics of each chemical compound brought identification of gaseous pollutants possible through comparing the shape, position and relative peak height of each measured spectrum with reference spectra.” on Page 5 (Line 32) - Page 6 (Line 2). The sentence “any gaseous compounds absorbed in the IR region (approximately 2.5–25 microns) were potential candidates for monitoring using FTIR technology” originally on Page 5 (Line 25-26) was moved to Page 5 (Line 28-29) for a better explanation of the analytical techniques. All characters in the x and y axes have been enlarged and all 16 species have been included in the Figure 4.

2.) How the rolling background is calculated and used should be explained more detailed.

Reply: Thanks for the comment. The rolling background was collected using the first spectrum as a background to create an absorbance spectrum from the second spectrum, using the second spectrum as a background for the third spectrum and so on. The integral values of concentrations are calculated to obtain time series data for each compound. The advantage of using the rolling background is that it will have the best correction for water vapor, detector and instrument response, and the lowest residual error.

Q5: 2. Materials and Methodology - 2.3 Qualitative receptor modelling: 1.) It is not very clear described. What is the index in  $X_i$ ,  $X_1$ ,  $X_2$  in Eq 1-4. Or the first index in the coefficient  $a_{11}$ . I would assume the index  $X_1 = X_{t1}$  and describe the time of the observations, as the origin of the Factors  $F_1, \dots$  is not stated it seems to be taken from the OP-observations. I would like something like  $F_{1\_op}$  (open path 1) and maybe an corresponding Factor  $F_{1\_CY}$  or  $F_{1\_NS}$ ,  $F_{1\_KS}$ . So it is very clear. And please report the factors at the receptor site and the source sites in a comparable way. Maybe complete the factors in table 3 as they are in table 2 just by adding 0.0. And please add the Factors from the site NS to table 3.

C5

Reply: Thanks for the comment. In order to explain the meaning of each index in Eq 1-4  $X_1 = a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + e_1$  Eq. (1)  $X_2 = a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + e_2$  Eq. (2)  $X_p = a_{p1}f_1 + a_{p2}f_2 + \dots + a_{pm}f_m + e_p$  Eq. (3)  $X = (X_1, \dots, X_p)'$ ,  $f = (f_1, \dots, f_m)'$ , and  $e = (e_1, \dots, e_p)'$  Eq. (4) The following sentences have been added to Page 6 (Line 13-15) – “where  $X_i$  = the  $i$ th chemical species with mean 0 and unit variance,  $i = 1, \dots, p$ ;  $a_{i1}$  to  $a_{im}$  = the factor loadings for the  $i$ th chemical species;  $f_1$  to  $f_m$  =  $m$  uncorrelated common factors, each with mean 0 and unit variance;  $e$  = the error terms indicating the residual part of  $X_i$  that is not in common with the other variables”. In Eq 1-4, “ $a_{11}$ ” represents the factor loading of factor 1 ( $f_1$ ) for the first chemical species, “ $a_{12}$ ” represents the factor loading of factor 2 ( $f_2$ ) for the first chemical species; “ $a_{21}$ ” represents the factor loading of factor 1 ( $f_1$ ) for the second chemical species.  $X_1$  describes the communality or common variance of the first chemical species; whereas,  $X_2$  represents the communality or common variance of the second chemical species. The factor # based on OP-FTIR (e.g.  $F_{1\_OP}$ ) and the factor # based on CC-FTIR with its corresponding source name (e.g.  $F_{1\_CY}$ ,  $F_{1\_NS}$ ,  $F_{1\_KS}$ ) have been added to Table 2, Table 4, Figure 6 and Figure 7; the factors at the receptor site and the source sites are now presented in a comparable way. The factors from the NS stacks has been added to table 4 (originally labeled as table 3) on Page 17 and the following sentences were added to Page 10 (Line 12-14) – “The chemicals from the NS stacks were mainly inorganic materials (nitrous oxide, silane, ammonia, nitrous acid, and nitrogen dioxide) that were commonly used in the solar cell production (Table 4e), all of which were not corresponded with the organic odorous solvents identified in the receptor sites”.

2.) Maybe you could also report a table with the scalar-products of  $\langle F_{1\_CY}, F_{2\_OP} \rangle$  for the 4 open path factors  $F_{i\_OP}$  with all source factors. Just taking Table 2 and table 3 (after adding the NS factors) would be a  $4 \times (2+2+\text{factors NS})$  Table/Matrix. It would even be interesting, if the factors of the different sources NS, CY, KS are more or less orthogonal or have a strong overlap.

Reply: Thanks for the comment. A scalar product is a scalar value that is the result of

C6

an operation of two vectors with the same number of components. Given two vectors A and B each with n components, the scalar product is calculated as:

$$A \cdot B = A_1B_1 + \dots + A_nB_n \text{ .Eq.A1}$$

The scalar-products of 4 open path factors (table 2) with all source factors [table 4 (originally labeled table 3)] were calculated using Eq. A1. In order to perform the calculation of scalar –products, the factor loadings (in table 2 & 4) were replaced by the eigenvectors generated by the SAS programs. The outcomes of a 4 x 6 table/matrix was shown in table A below(see also the attached 'Table A' on the last page of this document). It is suggested that none of the scalar-products was orthogonal except for the one calculated from F4\_OP  $\hat{\cdot}$  F2\_NS, indicating that F4\_OP was absolutely irrelevant with F2\_NS, meaning that F4\_OP –“solvent use for paint remover” was absolutely irrelevant with F2\_NS–“HNO3 thermal decomposition”. The top 3 scalar-products were calculated from F1\_OP  $\hat{\cdot}$  F1\_CY, F1\_OP  $\hat{\cdot}$  F1\_KS, and F4\_OP  $\hat{\cdot}$  F2\_KS with scalar-product values of 0.963, 0.727 and 0.570, respectively; indicating that the angles between these three pairs of vectors were far less than 90 degrees, meaning that the relationships between these three pairs of factors were strongest following the order of F1\_OP  $\hat{\cdot}$  F1\_CY > F1\_OP  $\hat{\cdot}$  F1\_KS > F4\_OP  $\hat{\cdot}$  F2\_KS in comparison with the rest of the pairs in this 4 x 6 table/matrix. Therefore, it would suggest that F1\_OP–“paint thinner for surface coating” was highly related to F1\_CY–“plastic paint thinner”; whereas F1\_OP  $\hat{\cdot}$  F1\_KS–“paint thinner for surface coating” was closely related to F1\_KS–“metal paint thinner”; F4\_OP –“solvent use for paint remover” was also related to F2\_KS–“cleaner or others”. The results of the scalar-products demonstrated that the factors at both the receptor site and the source sites were intercomparable, which were consistent with the findings in the previous sections of this manuscript.

Table A: The scalar-products of 4 open path factors with all source factors (a 4 x 6 matrix) F1\_OP  $\hat{\cdot}$  F1\_CY F1\_OP  $\hat{\cdot}$  F2\_CY F1\_OP  $\hat{\cdot}$  F1\_KS F1\_OP  $\hat{\cdot}$  F2\_KS F1\_OP  $\hat{\cdot}$  F1\_NS F1\_OP  $\hat{\cdot}$  F2\_NS 0.963 -0.130 0.727 0.303 0.003 -0.002 F2\_OP  $\hat{\cdot}$  F1\_CY F2\_OP  $\hat{\cdot}$  F2\_CY F2\_OP  $\hat{\cdot}$  F1\_KS F2\_OP  $\hat{\cdot}$  F2\_KS F2\_OP  $\hat{\cdot}$  F1\_NS

C7

F2\_OP  $\hat{\cdot}$  F2\_NS -0.072 0.001 -0.076 0.005 -0.179 0.074 F3\_OP  $\hat{\cdot}$  F1\_CY F3\_OP  $\hat{\cdot}$  F2\_CY F3\_OP  $\hat{\cdot}$  F1\_KS F3\_OP  $\hat{\cdot}$  F2\_KS F3\_OP  $\hat{\cdot}$  F1\_NS F3\_OP  $\hat{\cdot}$  F2\_NS 0.026 -0.026 0.040 -0.044 -0.309 0.344 F4\_OP  $\hat{\cdot}$  F1\_CY F4\_OP  $\hat{\cdot}$  F2\_CY F4\_OP  $\hat{\cdot}$  F1\_KS F4\_OP  $\hat{\cdot}$  F2\_KS F4\_OP  $\hat{\cdot}$  F1\_NS F4\_OP  $\hat{\cdot}$  F2\_NS -0.092 0.349 -0.310 0.570 0.029 0.000 Note: (1) If A and B are orthogonal (at 90 degrees to each other), the result of the scalar product will be zero; (2) If the angle between A and B are less than 90 degrees, the scalar product will be positive (greater than zero); (3) If the angle between A and B are greater than 90 degrees, the scalar product will be negative (less than zero)

Q6: 3. Results and Discussion - 3.2 Ambient data from receptor path: 1.) Please add a Figure with the time-series which are the basis for the calculation of r\_phi and r\_pb, OP-FTIR measurements and indicate, when odor was reported. Figure 2: shows no values, when the factor might be negative, please correct it, even if the contribution of the factor is negative it has to be reported. There should be errors in the coefficients, which explain negative values as least in a small range. Caption Figure 2: Time-series pattern -> Diurnal time-series pattern. Missing Figure: Could you add a complete time series of the 4 factors found at the receptor site, which is not a diurnal pattern.

Reply: Thanks for the comment. The time-series pattern of chemical species (used as the basis for the calculation of r\_phi and r\_pb) detected at the receptor site by the OP-FTIR has been added in Figure 5; the yellow highlights in the figure indicated the periods when odor was reported. The following sentences were added to Page 8 (Line 7-9) – “A complete time series pattern of chemical species found at the receptor site that were used as the basis for the calculation of r\_phi and r\_pb was shown in Fig. 2, in which the periods when odor was reported were highlighted”. The negative values in Figure 6a to 6d (originally labeled as Figure 2a to 2d) have been added to the diurnal time-series trends. Caption Figure 6 (originally labeled as Figure 2): “Time-series pattern” has been revised as “Diurnal time-series pattern”. A complete time series pattern (not diurnal pattern) of the four factors found at the receptor site has

C8

been added in Figure 6e, which suggested that the proportion of the factor scores in negative values were in a relatively small range.

Q7: 4. Conclusions: Is a bit short and very arbitrary and a little redundant: 1.) p.10 line.17: I would replace “dual-optical sensing system” by “FTIR- spectroscopic measurements.” And clarify less ambivalent how the meteorological data and cluster analysis was used. Maybe something similar as: “This study developed an alternative investigative framework for detecting air pollution sources of odor nuisance by measuring 17 gas species simultaneously using FTIR spectroscopic measurements and factor analyses to identify and characterize emission sources of multiple air contaminants. Meteorological data and Cluster analyses were employed to proof the identification of the major odor emissions” Maybe you could add some numbers how often the odor occurs which originate from CY,KS,NS and the different processes.

Reply: Thanks for the comment. “Dual-optical sensing system” has been replaced by “FTIR- spectroscopic measurements” in the conclusion section. The role of meteorological data and cluster analysis on Page10 (Line 17-19) were clarified by using the sentences suggested by the referee 1. The overall content of the conclusion has been rewritten as follows: “This study developed an alternative investigative framework for detecting air pollution sources of odor nuisance by measuring 16 gas species simultaneously using FTIR spectroscopic measurements and factor analyses to identify and characterize emission sources of multiple air contaminants. Meteorological data and cluster analysis were employed to proof the identification of the major odor emissions. Different industrial processes were related to a specific combination of different pollutants, and this combination was obtained using the two statistical methods of factor analysis and cluster analyses. Factor and cluster analyses were employed to improve the quality and completeness of the source profiles. A field study used FTIR spectroscopic measurements to determine the source of the emission of volatile organic odor species near an industrial park in southern Taiwan demonstrated the feasibility of this proposed method. The major odor emission source was identified through qualita-

C9

tive source apportionment of factor and cluster analyses. With enhanced efficiency in odor investigation methodology, future emission reduction plans can be developed and overall air quality can be improved”.

Q8: Figures and Table: 1.) Please do not use Appendix if you refer to the figure or table in the main text and ensure that all figures and tables are mentioned in the text and keep the order how they are used in the text.

Reply: Thanks for the comment. The “10. Appendix” section has been removed from the manuscript. All tables and figures originally in “10. Appendix section” were rearranged to either “8. Table” or “9. Figure” sections, in which they were reordered according to how they are used in the text to ensure that all figures and tables are mentioned in the text.

Please also note the supplement to this comment:

<https://www.atmos-meas-tech-discuss.net/amt-2018-344/amt-2018-344-AC1-supplement.pdf>

---

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-344, 2019.

C10

Table A: The scalar-products of 4 open path factors with all source factors (a 4 x 6 matrix)

F1_OP · FI_CY	F1_OP · F2_CY	F1_OP · FI_KS	F1_OP · F2_KS	F1_OP · FI_NS	F1_OP · F2_NS
<b>0.963</b>	-0.130	<b>0.727</b>	<b>0.303</b>	0.003	-0.002
F2_OP · FI_CY	F2_OP · F2_CY	F2_OP · FI_KS	F2_OP · F2_KS	F2_OP · FI_NS	F2_OP · F2_NS
-0.072	0.001	-0.076	0.005	-0.179	0.074
F3_OP · FI_CY	F3_OP · F2_CY	F3_OP · FI_KS	F3_OP · F2_KS	F3_OP · FI_NS	F3_OP · F2_NS
0.026	-0.026	0.040	-0.044	<b>-0.309</b>	<b>0.344</b>
F4_OP · FI_CY	F4_OP · F2_CY	F4_OP · FI_KS	F4_OP · F2_KS	F4_OP · FI_NS	F4_OP · F2_NS
-0.092	<b>0.349</b>	<b>-0.310</b>	<b>0.570</b>	0.029	<u>0.000</u>

Note: (1) If A and B are orthogonal (at 90 degrees to each other), the result of the scalar product will be zero; (2) If the angle between A and B are less than 90 degrees, the scalar product will be positive (greater than zero); (3) If the angle between A and B are greater than 90 degrees, the scalar product will be negative (less than zero)

Fig. 1.