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.

Anonymous Referee #2

Received and published: 9 July 2019

Application of Factor and Cluster Analyses to Determine Source–Receptor Relationships of Industrial Volatile Organic Odor Species in a Dual-Optical Sensing System

Jen-Chih Yang, Pao-Erh Chang, Chi-Chang Ho, Chang-Fu Wu

Summary

The authors present the results of a field study carried out over 10 days in March, 2015, whose goals were to identify and characterize odours reported by commuters at an intersection in an industrial park in southern Taiwan. Most likely candidates were

assumed to be industrial VOC emissions from a nearby sunglass factory ("CY") and a metal casings factory ("KS"), which emit organic solvents used in plastic and metal surface coating. Additionally, a nearby solar cell factory ("NS") is a source of inorganic materials used in high-temperature glass sintering.

The experimental approach involved measuring the stack emissions directly with a multi-path closed cell ("CC") connected to an FTIR spectrometer (effectively a point measurement), as well as indirectly at the intersection with an open-path FTIR system (effectively a path-average measurement over 143 m, one way between FTIR spectrometer and retroreflector array). FTIR measurements at all sites were made at 1 cm-1 resolution, averaging 64 IR scans over 5 minutes, and yielding 2,911 consecutive spectra in the case of the open-path system running over 10 days. It is not clear whether the closed cell measurements at the three factories were done at the same time or separately, but there is 1 day of measurements from CY, 10 days from KS and 4 days from NS. Wind speed and direction measurements were also made near the open-path measurement.

Spectral data analysis to derive VOC species concentrations used multicomponent classical least squares with "rolling backgrounds" in the open-path spectra (subject to changing meteorological conditions) and a "fixed reference method" in the closed-cell spectra (which are sampling stack emissions directly).

Subsequent data analysis consisted of performing a factor analysis involving 4 factors (surface coating, incomplete car engine combustion, solar cell production, solvent use) based on the chemical species observed by open-path FTIR at the receptor site (∼16, but there are inconsistencies in the text about which species are being measured and analyzed). The 4 factors (eigenvalues >1) were used in factor loading calculations; chemical species with loadings > 0.4 were considered as influential variables. Factor scores were presented as a function time, separating the one weekend from the remaining weekdays. Additionally, factor scores were presented as windroses.

Species detected at the stacks were compared to ambient open-path data, as well as a database of vehicle exhaust emissions. Finally factor and cluster analysis was performed on measurements inside the stacks, yielding 2 factors with eigenvalues > 1. One of the factors was identical between stack "CY" and the ambient OP-FTIR data Factor 1, leading the authors to conclude that this was the major odour source.

Major Comments

The study is well motivated, the experimental setup and data set is valuable, and the data is presented relatively clearly (e.g., dominant factors in bold/red in tables), however, the paper cannot be published without addressing the comments below.

First, the paper suffers from a number of distracting inconsistencies between what is described in the text and what is found in the tables and figures. For example, chemical species are referred to in the text that are not in in the figures or tables (e.g., cyclohexane, methanol, sulfur hexafluoride, isopropanol, dichloromethane) and the number of species is given as 17 when it is actually 16. Another example involves the discussion of wind directions in Figure 3 regarding OP-FTIR factors, where the named winds do not correspond to the windroses. Another example is a text reference to between 24 and 72 hours of measurements at the stacks but then it states that 2907 spectra were obtained from the KS stack at 5-minute intervals, i.e., a full 10 days of observations.

Second, the study claims to develop an "alternative approach" and to demonstrate the feasibility of an "alternative investigative framework", however, a clear description of the details of the methodology, with references to the literature, is insufficient for an un-initiated reader to make use of this "alternative approach". This applies to 1) the spectral analysis, 2) the factor and cluster analysis of the resultant chemical species concentrations, and 3) the grouping of the data as a function of time and wind direction. The paper is quite short so an expansion is warranted. Figure and table captions are quite sparse, and tables, in particular make improper use of footnote numbering to give information without the number applying to anything in particular in the table (see Table

1 and 2).

Third, there are a number of issues with the spectral data analysis. The detection limits in Table 1 are mostly sub-ppb, which seems high for the OP-FTIR technique in general (see, e.g., Jarvis, 2003, "Open Path Spectrophotometry" in the Instrument Engineer's Handbook), more so given the relatively low number of co-added spectra in a 5 minute period, and the relatively high absolute humidity levels (not given in the paper but inferred). The description of the detection limit calculations in the Table 1 footnote is not clear. Moreover, it is not clear why CO and N2O cannot be retrieved from these FTIR spectra. The issue of "rolling backgrounds" is mentioned, but CO and N2O are far from regions contaminated by water and have been retrieved routinely by others using OP-FTIR, though not with CLS (e.g., Paton-Walsh et al., 2014; Akagi et al., You et al., 2017). Smith et al (2011, AMT) have shown that classical least squares analysis yields inaccurate results at high concentrations (c.f. high concentrations in Table A1 in this study). The nitrous oxide shown in this study in Figure 4, panel (e), has a concentration below 20 ppb (7.9 in Table A1), which is impossible given the ambient concentration of ∼330 ppb. Is the accuracy of other retrieved gases similarly affected?

Fourth, the authors concede that the first factor might not be limited to one source (P8L18), especially when confronted with Figure 3a, which shows high factor scores from NNE, ESE, and also WNW (last direction not discussed though highest factor score from here). Why the sharp features at WNW and ESE as compared to the more spread out feature at NNW,N, NNE, NE? Also, on P8L30 it is stated that six species coexist at both the CY and KS stacks, but only acetone is truly common, the other 5 are emitted ∼1000x more at CY – can that be used to separate them?

Fifth, the authors name CY as the source of the odours based on an identical Factor 1 composition as compared to OP-FTIR. Does this check out with a windrose plot of butyl cellosolve and PGMEA, which are unique to CY based on Table A1? For that matter, can it be verified that when winds blow from stack NS (nearly from the East, not NE as given on P7L34 and P8L1) cyclohexane, acetylene and ethylene increase?

There should also be toluene and xylene. I can understand why no correlation plots (like Figure 5) are shown for KS given that the correlation is discussed as below 0.1, but why is NS not discussed in this way at all?

Sixth, why is Factor 2 (incomplete engine combustion) clustered to the NW and SE (Figure 3b, not discussed), whereas the source directions are indicated as "all directions" in Table 2? It also appears like there are roads all around the open-path in Figure 1. Are some roads more major than others? Seventh, the time-series of factor scores do not correspond to factory working hours in an obvious way. Do the plants run 24 hours a day? Does the solar cell plant run on weekends? Why is there a traffic peak from 6-8 on the weekend? Why does the traffic factor look similar to the solvent factor?

Seventh, the time-series of factor scores do not correspond to factory working hours in an obvious way. Do the plants run 24 hours a day? Does the solar cell plant run on weekends? Why is there a traffic peak from 6-8 on the weekend? Why does the traffic factor look similar to the solvent factor?

C5