Atmospheric Environment 180 (2018) 138–148

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

DeltaSA tool for source apportionment benchmarking, description and sensitivity analysis

D. Pernigotti, C.A. Belis*

European Commission, Joint Research Centre, Ispra, Italy

ARTICLE INFO

Keywords:
Ensemble
Receptor models
Performance criteria
On-line

ABSTRACT

DeltaSA is an R-package and a Java on-line tool developed at the EC-Joint Research Centre to assist and benchmark source apportionment applications. Its key functionalities support two critical tasks in this kind of studies: the assignment of a factor to a source in factor analytical models (source identification) and the model performance evaluation. The source identification is based on the similarity between a given factor and source chemical profiles from public databases. The model performance evaluation is based on statistical indicators used to compare model output with reference values generated in intercomparison exercises. The references values are calculated as the ensemble average of the results reported by participants that have passed a set of testing criteria based on chemical profiles and time series similarity. In this study, a sensitivity analysis of the model performance criteria is accomplished using the results of a synthetic dataset where "a priori" references are available. The consensus modulated standard deviation *punc* gives the best choice for the model performance evaluation when a conservative approach is adopted.

1. Introduction

Despite the progress made in the latest decades, air pollution is still the primary environmental cause of premature death in Europe (Maas and Grennfelt, 2016). In order to design abatement measures, knowledge of the origin of pollutants affecting a given area is required (Directive 2008/50/EC, 2008). Source Apportionment (SA) aims to allocate shares of the measured pollutant mass to its emission sources, so called source contribution estimate (SCE). In the real-world, the actual SCEs are unknown. Due to such lack of references, a common problem in SA studies is to validate the model outputs. In the framework of the Forum for air quality modelling in Europe (FAIRMODE, 2007) the European Commission JRC launched, inter-comparison exercises for particulate matter SA among receptor models and more recently also for Chemical Transport Models. The experience gained analysing the data of such intercomparisons led to an European Guide for SA receptor models (Belis et al., 2014) and to a new methodology for evaluating SA performance (Belis et al., 2015a, 2015b: B2015 in the following).

In an intercomparison for SA (a glossary is provided in Appendix A) many practitioners run their models using the same input dataset, providing the following information for each source identified in the output (hereafter referred to as a candidate): the overall average SCE and the SCE time series (SCT) in absolute mass (e.g. $\mu g/m^3$) the source chemical profile (CP) and the contribution-to-species (the % of a given species attributed to that candidate source, C2S). In factor analytical methods the correspondence between factors with real-world sources or processes is accomplished in post processing (Hopke, 2009). Concerning source identification, particulate matter CPs measured at the source are the most reliable references. To support SA practitioners in this step, a repository for measured CPs was created. The above-mentioned SA evaluation methodology, embedded in the DeltaSA on-line tool, includes a total of 1 160 CPs from the SPECIEUROPE database developed at JRC (Pernigotti et al., 2016) and the SPECIATE database (Hsu et al., 2014). For pollutants deriving from secondary processes where measured profiles are not available (e.g. ammonium sulphate and nitrate), the stoichiometric profiles are considered.

The methodology for the intercomparison evaluation is described in B2015 and the results in Belis et al. (2015a and Belis et al. 2017. In the first steps (so called complementary and preliminary) for each source a set of screening criteria are made on the corresponding participants candidates. In real-world dataset the candidates successful to the previous test are averaged to build an ensemble reference, while this is not necessary if the reference is synthetic. In the final step each candidate is compared with the corresponding source reference to asses the participant performance. In particular, the second intercomparison is the only one performed with a synthetic dataset artificially created by the JRC where, unlike real-world data, pre-defined reference CP, SCE and

* Corresponding author.
E-mail addresses: deniperni@gmail.com (D. Pernigotti), claudio.belis@ec.europa.eu (C.A. Belis).

https://doi.org/10.1016/j.atmosenv.2018.02.046
Received 8 August 2017; Received in revised form 21 February 2018; Accepted 26 February 2018
Available online 01 March 2018

SCT are available. In all other intercomparisons no control on the reliability of the ensemble reference itself is possible.

In this paper, using the synthetic dataset, a criterion is proposed to improve the robustness of the methodology when using real-world datasets, taking into account the consensus among participants on the presence of a given source in the analyzed data. If there is a large consensus, then the model performance criteria (MPC) for that source will be more stringent. On the contrary if the consensus is low and just few participants agree on the presence of a contribution from that source, then the uncertainty will be larger and the MPC will be less stringent. The present study is divided in three sections. In the first, an updated version of the methodology for intercomparison evaluation (B2015) is summarised, as an introduction to the following sections. In the second, sensitivity tests using the abovementioned synthetic dataset to improve the model performance criteria (MPC) when using ensemble references are proposed. In the third section the Java web interface DeltaSA implementing functionalities for chemical profile similarity and model performance evaluation of the R-package with the same name is illustrated.

2. Developments in the methodology and re-evaluation of the synthetic dataset

The methodology described in B2015 has evolved as more experience with other intercomparisons was gained. In this section, the methodology for the SA model performance evaluation is summarised and the results of a sensitivity analysis using the synthetic dataset are presented. The methodology for the evaluation of the model performance comprises three steps: complementary tests, similarity tests and performance tests. In intercomparisons with real-world datasets, the objective of the first two steps is to select the candidates to be used for the ensemble reference.

2.1. The synthetic dataset

The synthetic dataset consists of artificially created PM2.5 daily average concentrations (total mass and chemical speciation for 38 species) for Milan in 2005 (Belis et al., 2015a). The 25 participants were using various receptor models (see Belis et al., 2015a for details on models): PMF (17 participants, mostly using PMF3), CMB (4), FA (2), ME2 (1), COPREM (1). They presented between 6 and 13 candidates each, with a total of 190 candidates, and 266 candidate-source couples (considering that some candidates were attributed to more than one source). There were up to four sources attributed to a single candidate while five candidates were attributed to sources that were excluded from the analysis. The most populated source was 1 (traffic), with 30 candidates.

The synthetic reference CP, SCE and SCT for each “a priori” source is shown in Fig. 1 (panels B,C and A respectively). The reference uncertainty was set to 20% for SCE and to 36% for SCT (the quadratic sum of 20%, and 30%, respectively the SCE and PM total mass uncertainties at each time step) while the CP uncertainty depends on the measurement technique.

2.2. Complementary tests

These tests provide information about the overall consistency of individual reported results. Additional checks are used to exclude participants and/or candidates whose results present macroscopic irregularities from the reference ensemble. In practice, those participants having the sum of time averaged SCT or sum of CP in absolute mass, differing by an order of magnitude from the sum of SCE or from the measured PM total mass, are excluded from the calculation of the reference. Only four candidates were excluded due to these criteria during the evaluation of the synthetic intercomparison (synthetic in the following).

In the updated methodology warnings are given for: a) participants with a difference of candidates with respect to the median for participants of more than three; b) candidates with the sum of the SCE of all the sources differing by more than 20% from the PM mass; c) candidates with SCT average total mass differing by more than 20% from SCE; d) candidates with the total reconstructed mass time series (sum of candidate SCTs) being out of the target plot (Thunis et al., 2012, in the following T2012, with the modification reported in Appendix B). Moreover, warnings are also given for candidates with less than four valid species in the CP, zero or missing SCE and/or CP.

2.3. Preliminary tests

Below we give a short summary of the tests, more extensive descriptions can be found in Appendix B, Pernigotti et al. (2016) and B2015. For each **candidate-source** couple (the couples are defined by each participant), the distances between the candidate and the source repository CPs corresponding to that source category are computed, together with the distances from all the other candidates attributed to the same source category (in the following denoted with the prefix ‘r_’ and ‘f_‘ respectively).

The distance indicators are: Pearson Distance (PD = 1-R, where R is the Pearson correlation coefficient) and the Standardised Identity Distance (SID). Only SIDcp and PDcp can be calculated against the repository CPs, while PDsct and PDc2s can be only calculated against the other candidates’ SCT and C2S. The suffix ‘-norm’ indicates that the SID has been normalised to account for the variability of the source categories (Appendix B). A source dependent coefficient $\boldsymbol{q}$ is set to the 95th percentile of the distances among repository CPs belonging to a given source category. This coefficient modulates the maximum allowed identity distance ID (MAD) for every source, so that the test is tolerant for sources with great variability in the measured chemical profiles and is stringent for those with a well-defined chemical fingerprint. In cases where $\boldsymbol{q}$ cannot be calculated (less than 3 CPs with at least 2 common species) a default value of 1 is taken. The value of $\boldsymbol{q}$ depends on the repository CPs, the considered source, as well as the intercomparison dataset, given that the calculation is only performed on the participants reported chemical species.

The values of $\boldsymbol{q}$ calculated for the synthetic dataset are: 1.15 for fuel oil, 1.12 for industrial, 1.08 for traffic, 1.08 for wood burning, 1.06 for biomass burning, 1.02 for exhaust, 1.01 for cement production, 0.94 for iron & steel production, 0.93 for road dust, 0.89 for de-icing salt, 0.88 for soil dust and 0.67 for marine aerosol. The default value of 1 is kept for the secondary sources.

The acceptability criteria for distances are $\mathbf{SIDcp\_norm} \leq \mathbf{1}$ and $\mathbf{PD} \leq \mathbf{0.4}$, where the first is given by the definition of SID_norm and the second corresponds to a Pearson coefficient above 0.6. In Fig. 2 SID distances between candidates and repository profiles are plotted for the synthetic arranged by category. The marine (12), deicing salt (66), secondary inorganic aerosol (60), ammonium nitrate (61) and ammonium sulphate (62) sources fall outside the acceptability area (green) when compared to repository profiles.

Only candidate-source couples fulfilling two out of the following three criteria pass the preliminary tests and are admitted to the ensemble reference calculation: 1) median of r-SIDcp-norm ≤ 1, 2) median of r-PDcp ≤ 0.4 (if the repository CPs are missing f distances are used) and 3) 25th percentile of f-PDsct ≤ 0.4. Criterion 3 aims at excluding candidates with an uncorrelated time trend. To avoid giving any single participant too much influence, where multiple candidates are from the same source only the candidate with the minimum r-SIDcp is kept.

In the synthetic after the application of the complementary tests all the (P = 25) participants are admitted to be ensemble members. In total 21 candidates and 50 candidate-source couples were excluded from the reference calculations.