Application of Cross-Section Filtration to Detect and Correct Errors Caused by Technical Problems in SMPS.

Wagner, Zdeněk
2019

APPLICATION OF CROSS-SECTION FILTRATION TO DETECT AND CORRECT ERRORS CAUSED BY TECHNICAL PROBLEMS IN SMPS

Zdeněk WAGNER[1], Pavel KOVANIC[2], Naděžda ZÍKOVÁ[1], Jakub ONDRÁČEK[1], Jaroslav SCHWARZ[1], Vladimír ŽDÍMAL[1]

[1] Institute of Chemical Process Fundamentals of the CAS, v. v. i., Prague, Czech Republic, wagner@icpf.cas.cz
[2] Retired from Institute of Information Theory and Automation of the CAS, v. v. i., Prague, Czech Republic

Keywords: Scanning mobility particle sizer, Particle size distribution, Cross-section filtration, Mathematical gnostics, Residual entropy

INTRODUCTION

The scanning mobility particle sizer (SMPS) is one of the instruments used for measurement of particle size distribution function in atmospheric aerosols. The particles are classified according to their mobilities in a flow of a gas (usually air) and an electric field. The particles are then counted. At a constant gas flow and voltage the particles at the output of the classifier have equal electrical mobility. Certain percentage of particles bear a single elementary charge but a known portion of larger particles bear two or more elementary charges. Such particles exit the classifier at the same voltage of the electric field as smaller particles bearing a single charge. In order to estimate the particle size distribution, correction to multiple charges has to be applied. However, this correction can have a negative impact to the estimated distribution. A technical problem can occur in a particular size bin. If measurement of large particles is affected by such a problem, the error propagates to the corrected number of smaller particles yielding in an extreme but not so rare case to a negative count of particles. Our idea is that the errors caused by the technical problems should have a character differing from the properties of common experimental errors. The aim of the work is thus development of a robust algorithm for detection of such errors and their fully automatic correction. The algorithm should be fast so that it could be applied in real time.

DATA

In this work data from two SMPS instruments were used, namely the one located in National Atmospheric Observatory Košetice (NAOK), station code CZ0003R, GAW-ID KOS, instrument name SMPS_KOS, measurement started on May 1, 2008, and the one located in Prague Suchdol, station code CZ0004B, GAW-ID PRG, instrument name SMPS_Suchdol, measurement started on April 6, 2012. Time resolution of measurement is 5 min in both instruments. With the exception of data presented in Fig. 2, raw mobility size distributions, extended version, were used to calculate inverted values by TROPinv v. 2.13 (2013-09-19) described in TROPinv (2019). We are interested in analysis of errors propagation during the multiple charge correction, therefore corrections for particle losses and CPC (condensation particle counter) efficiency were not applied. The instruments have different sheath flows that was taken into account for the evaluation of the DMA transfer function area.

Distributions with negative concentrations were identified in the **.inv* files by the standard UNIX utility *grep*. Afterwards the instrument logbooks were consulted. Such events often occur during switching between measurement and calibration. These distributions are always flagged as invalid data and were therefore discarded from further analysis. Only data considered as valid were used.

STATE OF THE ART

According to the specification issued jointly by EBAS and GAW WDCA (2019), the data are submitted to the international databases in three forms derived from the NASA Ames 1001 format (ESPO Data Archive, 2019). Level 0 file contains the raw data and technical variables without any corrections. Level 1 file contains inverted data which are in case of SMPS corrected to multiple charges, particle losses and CPC efficiency. Level 2 file contains hourly means of data converted to the standard pressure and temperature. A set of standard flags is defined for marking distributions affected by various types of atmospheric and technical conditions which may reduce the reliability of the data but the data are considered valid as well as flags which invalidate the data and provide the reason of invalidation. So far there is no established method for automatic detection and/or correction of errors caused by the technical problems. Such data cannot be easily spotted by human eye of a person responsible for QA/QC. It is therefore highly probable that such data remain unnoticed and are submitted without any flag meaning that these incorrect data are considered valid by the data originators and submitters.

It could be anticipated that averaging can reduce the impact of distributions damaged by technical problem. However, Fig. 1 shows that this is not the case. Here two distributions were affected by technical problems. The magenta line displays the mean calculated from all data. If the physically impossible negative values are replaced with zeros prior to averaging, the resulting mean distribution is at least physically meaningful but still incorrect. The red line shows the average of distributions corrected by the algorithm explained later in the article.

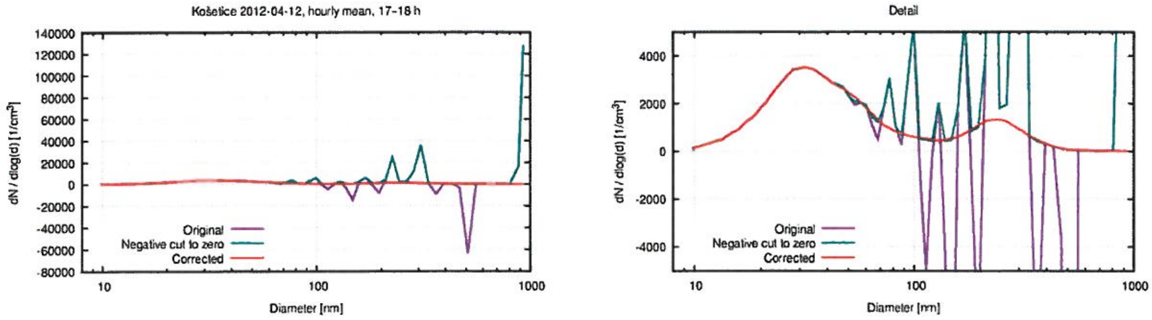

Fig. 1: Sample hourly mean distribution with a detailed view

In our automatic system the algorithms described by Ždímal *et al.* (2008) are used. Fig. 2 shows that even in case where filtration in the time domain is used prior to estimation of the particle size distribution false modes located at large diameters are found.

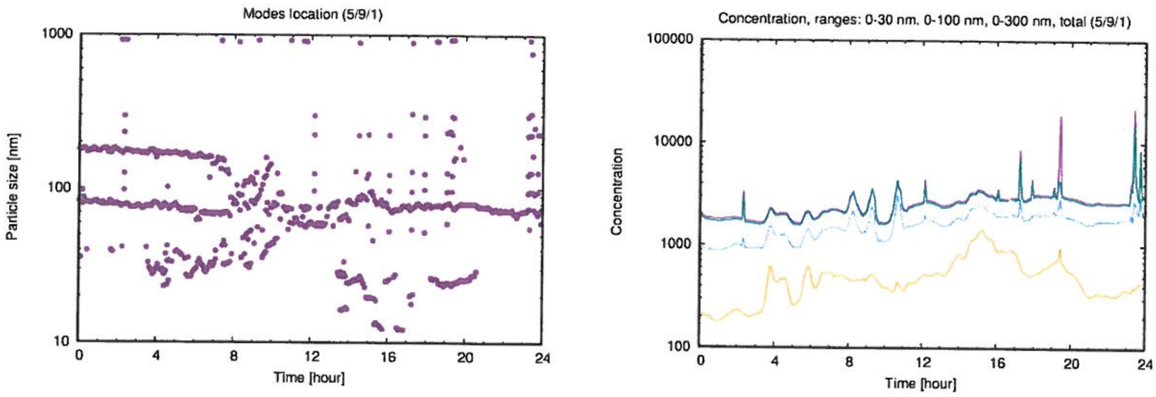

Fig. 2: Modes locations (left) and concentrations (right) estimated from inverted data filtered in the time domain

METHOD

Both SMPS instruments used in our stations measure size distributions in two scans, an upscan with increasing voltage and hence increasing particle diameter, followed by a downscan with decreasing voltage. The particles are counted in predefined size bins forming a geometric series. The instruments are equiped with software provided by TROPOS (TROPinv, 2013) which calculates arithmetic averages from the two scans (upscan and downscan) for each size bins and then makes use of these averages for inversion (multiple charges correction etc.). If particle count in one of the scans for a particular size bin cannot be obtained, only the count from the other scan is used.

We have found that if a technical problem occurs, only one of the scans is affected. We can therefore omit this number from averaging. This means that we do not insert any artificial artifact, we thus do what the software would do by itself if instead of measuring the wrong value the particle count in the corresponding scan were not obtained.

Particle counts in the scans can differ for various reasons: a common measurement error, a change of the aerosol during the time between the scans, a technical problem of the instrument. The aim of the additional data processing is to detect and correct instrument problems only while leaving the other differences intact. We thus have to distinguish technical problems from other sources of differences in particle counts between the scans. The algorithm must be adaptive because the aerosol can change considerably in time. Statistical methods cannot be used because the type of the distribution of the particle count differences between the scans is unknown and there is no theoretical background which could be used. In addition, the number of size bins affected by technical problems is always small, hence statistically insignificant. Useful methodology in such cases is offered by mathematical gnostics (Kovanic *et al.*, 1995).

Mathematical gnostics is a nonstatistical paradigm of uncertainty rooted in the theory of measurement. The condition of measurability implies that the measured values and uncertainties must be modelled by Abelian groups with neutral and inverse elements. A value is not considered a number. Instead, an isomorphism to an Abelian group of real numbers is only one of numerous isomorphisms. Abstract analysis of possible models of uncertainty allows us to find analogies with the special theory of relativity and with thermodynamics. We can thus assign entropy to each individual datum. Making use of equations derived in the information theory we can further calculate information and probability of each individual measured value. The equations including the derivations are given in the above cited book. The book also explains what the ideal gnostic cycle is. It consists of two different paths, quantification and estimation. The entropy is related to their lengths which is measured in different geometries, namely in the Euclidean and Minkowskian geometry. The difference between these two entropies is called *residual entropy* which should not be confused with the same term used in thermodynamics.

Detection of errors caused by the technical problems is based on analysis of dependence of the residual entropy on the value of an added datum. We evaluate the first two derivatives which enables us to split the set of entropies to kernels. It was found that data representing the differences caused by technical problems have their own kernel which can easily be identified. In the size bins identified by this algorithm as subject to technical problems of the instrument we accept the count which is closer to the previous corrected distribution and the count measured in the other scan is discarded. In all remaining size bins we preserve the counts measured in both scans and the inversion software will use their arithmetic means.

RESULTS

The proposed algorithm was tested on 2880 distributions from NAOK and 2016 distributions from the Prague-Suchdol station. The data from the whole days were selected in such a way that they contained negative concentrations in the uncorrected inverted data and there was no comment in the logbook for the corresponding day meaning that there was no apparent reason to flag the data as invalid. Since the data from whole days were analyzed, the test set contained both distribution that were subject to technical problems and undisturbed distributions where corrections should not be applied, as the technical problems typically affects only some distributions during the day. Calculation was performed by a preliminary version of the gnostic software which can only run in a single thread and a distribution was typically processed within 6 seconds. The algorithm can thus be applied in real time even on the same computer which is used for controlling the measurement device.

Two types of technical problems were observed. Both types can occur simultaneously in the same distribution but such case is not frequent. In the first type a particle count in one of the scans for a smaller particle size is missing but the acquisition software replaces it as a valid zero. The arithmetic mean of these scans thus yields a small value and after multiple charge correction the resulting concentration in the size bin can be negative. The error caused by the second type of a technical problem is more pronounced. The measured count of large particles is too high, sometimes even several orders of magnitude higher than the real number. After application of multiple charge correction the error propagates to particle concentrations in the lower size bin. Fig. 3 shows that several consecutive distributions can be affected by such technical problems.

This explains why neither averaging nor filtration in the time domain helps in obtaining reasonable particle size distribution as demonstrated in Figs. 1 and 2.

Fig. 4 Displays the event caused most probably by a car running nearby. As the car approaches, the downscan measures higher concentration, afterward the larger concentrations are observed in the upscan. The proposed algorithm properly recognizes this effect as a real process and does not correct the data.

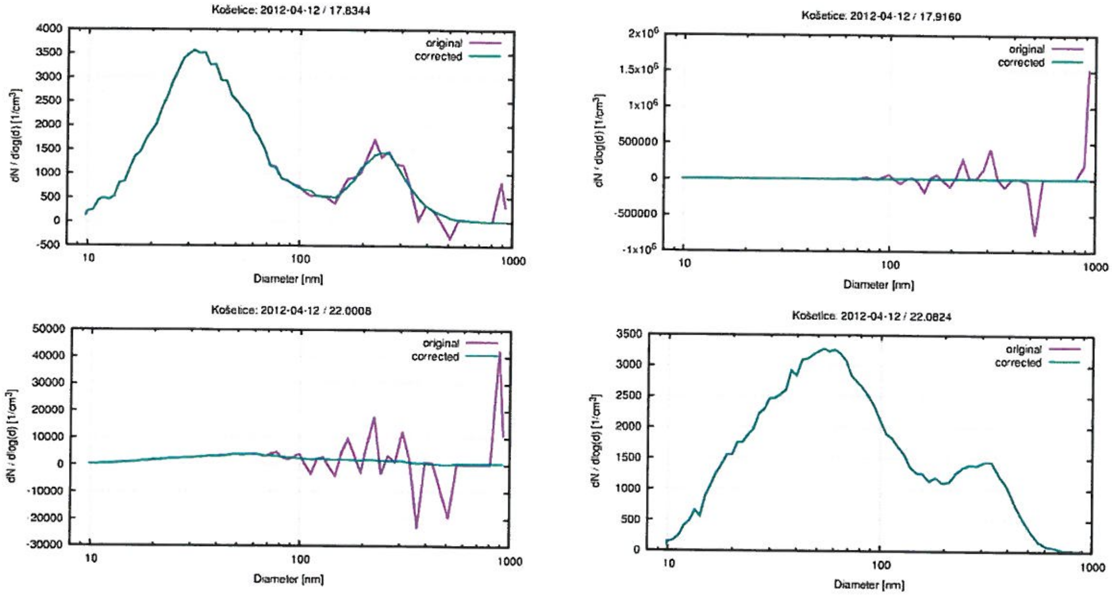

Fig. 3: Original and corrected distributions

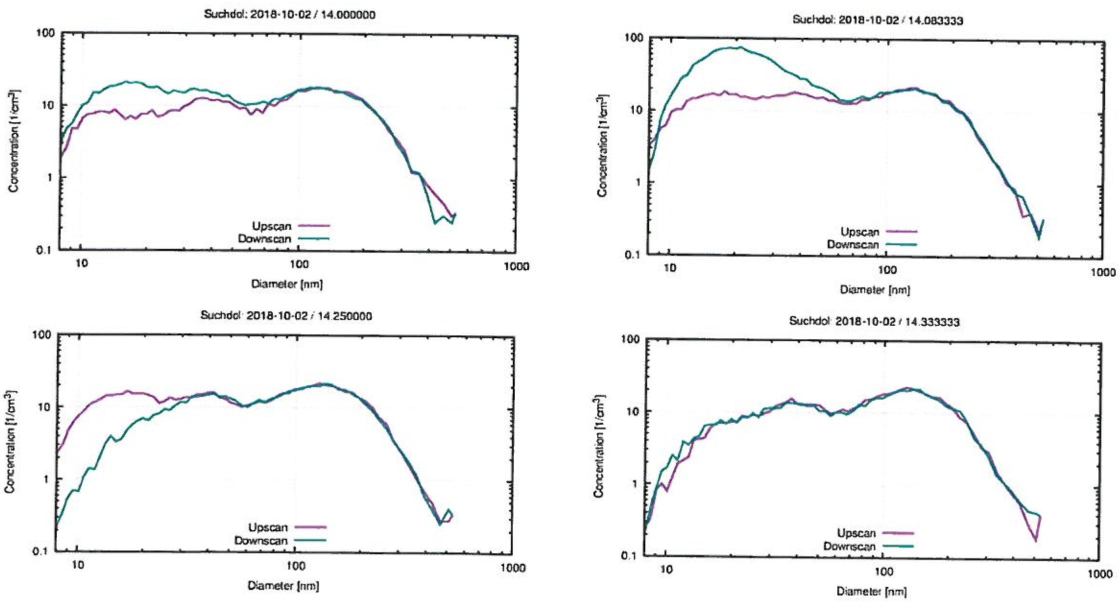

Fig. 4: Fast change of the particle concentration due to a mobile source (traffic), no correction will be applied before inversion

CONCLUSION

This work proves that utilization of knowledge of physical principles of the process leads to better results than application of purely mathematical solution based on abstract assumptions that cannot be verified. The recent results are based on data from a rural background and an urban background station. Analysis of data of other types of stations are needed before they can be reliably generalized.

We also agree that even such mathematical method is a tool which should be used but not abused. We must not try to play tricks with poor quality measurement. However, we often need to carry out measurement at the very limits. In such cases we can achieve our goal using a combination of correct experimental techniques with advanced data analysis.

ACKNOWLEDGEMENT

This work was supported by the Czech Science Foundation under grant number 17-19798S and grant EU H2020, ACTRIS-2, number 654109.

REFERENCES

ESPO Data Archive: Ames Format Specification (v2.0). Available only online at https://espoarchive.nasa.gov/content/Ames_Format_Specification_v20 [cit. 2019-09-12]

GAW WDCA: WMO Global Atmosphere Watch, World Data Centro for Aerosols. Available only online at https://www.gaw-wdca.org/Submit-Data [cit. 2019-09-12].

Kovanic P., Humber M. B.: The Economics of Information—Mathematical Gnostics for Data Analysis, book, TXu001594192, 717 pp., 2015. Available only online at https://www.math-gnostics.eu/books/ [cit. 2019-09-11].

TROPinv (2013). Available only online at http://wiki.tropos.de/index.php/TROPinv [cit. 2019-09-11].

Ždímal V., Brabec M., Wagner Z.: Comparison of Two Approaches to Modeling Atmospheric Aerosol Particle Size Distributions. (Eng) Aerosol Air Quality Res. 8(4), 392-410 (2008).