Exploratory study on a statistical method to analyse time resolved data obtained during nanomaterial exposure measurements

Frédéric Clerc, G.-H. Njiki-Menga, Olivier Witschger

EXPLORATORY STUDY ON A STATISTICAL METHOD TO ANALYSE TIME RESOLVED DATA OBTAINED DURING NANOMATERIAL EXPOSURE MEASUREMENTS

F. Clerc[1]**, G-H Njiki-Menga, O.Witschger**

INRS, Département Métrologie des Polluants, 1 rue du Morvan, 54500 Vandoeuvre les Nancy, France

E-mail: frederic.clerc@inrs.fr

Abstract. Most of the measurement strategies that are suggested at the international level to assess workplace exposure to nanomaterials rely on devices measuring, in real time, airborne particles concentrations (according different metrics). Since none of the instruments to measure aerosols can distinguish a particle of interest to the background aerosol, the statistical analysis of time resolved data requires special attention. So far, very few approaches have been used for statistical analysis in the literature. This ranges from simple qualitative analysis of graphs to the implementation of more complex statistical models. To date, there is still no consensus on a particular approach and the current period is always looking for an appropriate and robust method. In this context, this exploratory study investigates a statistical method to analyse time resolved data based on a Bayesian probabilistic approach. To investigate and illustrate the use of the this statistical method, particle number concentration data from a workplace study that investigated the potential for exposure via inhalation from cleanout operations by sandpapering of a reactor producing nanocomposite thin films have been used. In this workplace study, the background issue has been addressed through the near-field and far-field approaches and several size integrated and time resolved devices have been used. The analysis of the results presented here focuses only on data obtained with two handheld condensation particle counters. While one was measuring at the source of the released particles, the other one was measuring in parallel far-field.. The Bayesian probabilistic approach allows a probabilistic modelling of data series, and the observed task is modelled in the form of probability distributions. The probability distributions issuing from time resolved data obtained at the source can be compared with the probability distributions issuing from the time resolved data obtained far-field, leading in a quantitative estimation of the airborne particles released at the source when the task is performed. Beyond obtained results, this exploratory study indicates that the analysis of the results requires specific experience in statistics.

---

[1]   To whom any correspondence should be addressed.

1. Introduction

Manufactured nanomaterials are becoming more and more common in a wide range of domains: material, food industry, healthcare, energy or transportation industry [1]. The manufacturing of such materials is realized by workers who might be exposed to aerosols containing manufactured nanoparticles (MNP). Even if the danger related with MNP is not always known, the risk always exists and consequently, the reduction of exposures should be an objective.

Three routes of exposure exist: skin permeation, ingestion and inhalation. Concerning MNP, the inhalation is considered as the main route of exposure. For reducing this exposure, a relevant and reliable quantitative characterization of exposure at workplaces has to be developed.

Background aerosols from natural and incidental sources in or outside the work environment are ubiquitous and present a major issue to overcome when using time resolved instruments that are not specific for MNP; as indeed any other aerosol instrument used in industrial hygiene to characterize larger airborne particles (i.e. photometers or optical particle counters). Adequate characterization of exposures to aerosols of MNP cannot be accomplished without successfully distinguishing them from background aerosols. One of the approaches that can be used when using time-resolved instruments is based on a combined analysis of time series with and without activity and spatial analysis.

With respect to aerosol instrumentation, almost all the studies used the commonly type of instrument for counting airborne particles: condensation particle counter (CPC). The CPC is a time-resolved size-integrated instrument which detection range varies with the specific type but usually goes from about 2 nm up to about 1 µm. With this instrument, so called time series of particle number concentrations can be subject to statistical studies.

In other fields of science and technology, time resolved data is a common tool used for monitoring activity. Beyond signal processing, focused on mathematical methods, the field of economic sciences makes an extensive use of statistical methods used for analysing time resolved data, in particular ARIMA methods [2]. In most cases, the aim is to forecast the (near) future of time series. The input time series is often long term data possibly presenting cycles or seasonality. Apart exceptions, these requirements do not match with occupational exposure to chemical substances objectives. Indeed, is often consists of making a reliable comparison with a limit value; while in most cases, the duration of the measurement is a shift or less. This is why novel methods should be developed.

In most case studies, graphical representations of time series are provided and commented, with the support of common statistical indicators [3, 4, 5]. Recently, advanced statistical methodology for estimating exposure to MNP has been presented; it uses ARIMA time series analysis [6]. In the present study, a different statistical approach is proposed for assessing potential exposure based on time-resolved particle concentration measurements.

2. Methodology

2.1. Time-resolved data used

The time-resolved data used in this exploratory work are from a recent workplace related exposure study that investigates the potential for exposure via inhalation from cleanout operations by sandpapering of a reactor producing nanocomposite thin films embedded with silver nanoparticles [7]. In this academic research laboratory, small stainless steel samples were coated with nanocomposite thin films using a plasma deposition process reactor. During the experimental phases which could last several consecutive days, the cleaning of the reactor was a daily operation. This operation was done by hand with sandpaper; the objective being to polish all the internal walls of the reactor to recover the

original color of stainless steel. As all sanding operations of this type, emissions of airborne particles could be suspected but almost no emission or exposure data are available in the literature so far. Therefore, the objective was to characterize the emitted aerosol in terms of particle sizes, particle concentrations and morphology, and also to look how the element silver, signature of the embedded nanoparticles in the thin films produced, could be found in the workplace atmosphere in the near field or far field environment of the reactor. The measurement strategy was based on the recent one developed in France [8]. Among aerosol instruments used were two handheld CPCs. These are the data from these two CPCs that are used in this exploratory study. One of the two was located close to the source; the distance to the source was about 10 to 20 cm. The second was located at 5 to 6 meters away in the same room (see Fig. 1). The only person performing a task was the one that worked on the cleaning of the reactor. No other activity or task was running in the laboratory room during the measurement period. Prior, during and after the cleaning operation, no incidental sources of nano-sized airborne particles in or outside the work environment has been identified. Also, the doors were left open, and the room was not equipped with special built ventilation compartment or local exhaust ventilation.
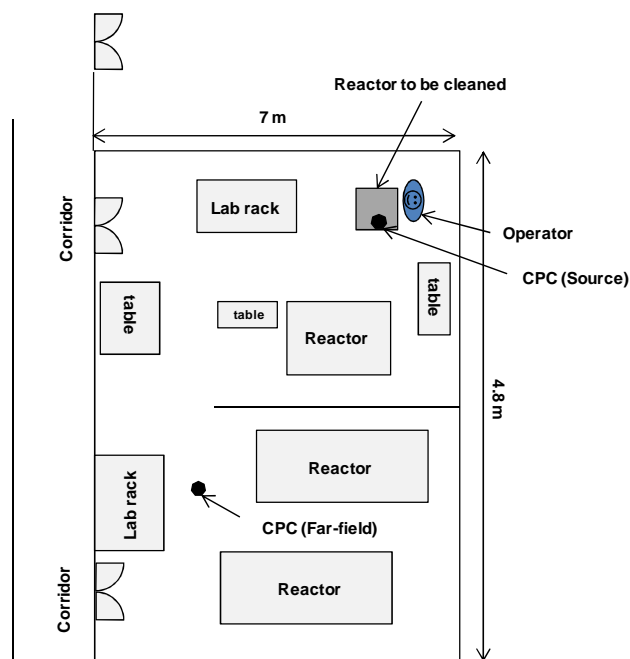

Figure 1. Floor plan of the laboratory room where the measurements took place.

Both CPCs recorded the particle number concentration during a total time of 68 minutes. The different tasks of the cleaning operation are labeled A to F; these correspond to different parts that were sanded inside the reactor like the walls, the silver target etc.

In the initial strategy, operators in charge of the measurements have used two different frequencies for each of the two CPCs: the "far field" CPC was recording a value every 30 seconds while the "source" CPC was recording the number concentration every 5 seconds. For the purposes of statistical analysis, a frequency identical to the two instruments is required. This is why the first action was to equalize the two time series at the same frequency of 30 seconds Thus, each set of source values corresponding to a single far field value are identified. The average of these values is then computed.

Finally, the time series analyzed are composed of 51 values as shown on Fig. 2.
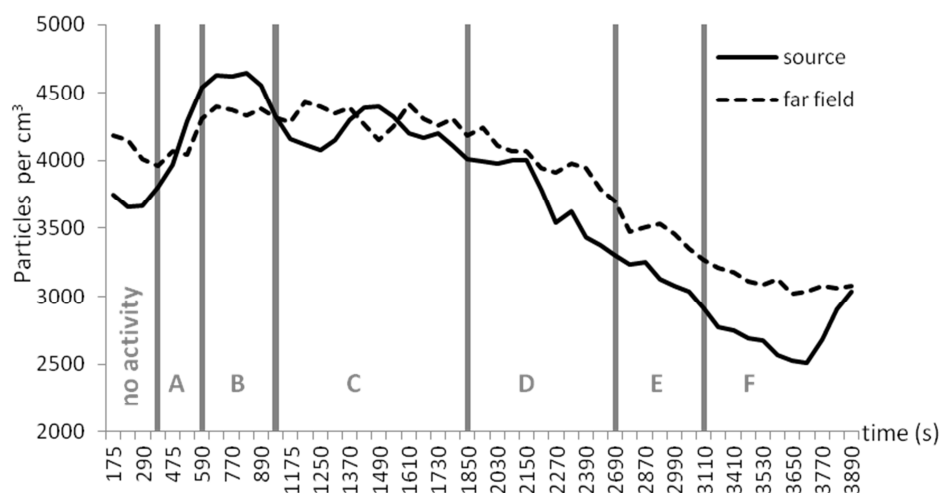
Figure 2. Particle number concentration profiles over time measured at the source and far field during a cleaning operation. The letters above the X-axis correspond to different parts that were sanded inside the reactor. During the time period 0-350 s there was no activity or task performed.

2.2. Statistical analysis

Three hypothesis are assumed

(1) Both source (black curve) and far field (dotted curve) CPCs record the same aerosol when there is no activity. The measured values include a systematic difference due to the devices characteristics.

(2) When the cleaning operation is realized, the CPC close to the source of emission also record the airborne particles released by the hand sandpapering.

(3) There was no specific ventilation in the laboratory room during the activity: the background aerosol behaves similarly around both CPCs, while the released airborne particles remain in the near field of the source.

Based on these assumptions, the principle is composed of four steps (see Fig. 3)

(1) Computing the systematic difference between measurement devices. The systematic difference is measured when no activity is realized. It is represented as a probability distribution of possible values.

(2) Including the systematic difference in the "source" time series thus computing a corrected source time series. This corrected source time series is comparable with the "far field" time series. Since the systematic difference is a probability distribution, each point of the corrected source time series comes with uncertainty.

(3) Computing the time series of emitted particles. For each data point, the far field value is subtracted from the source value with uncertainty. The result is the time series of extra

particles measured at the source of emission, with a global uncertainty coming from corrected source time series.

(4) Computing the quantity of particles counted during each task, in the form of probabilistic distributions taking into account the uncertainty related with the systematic difference and the time variation.
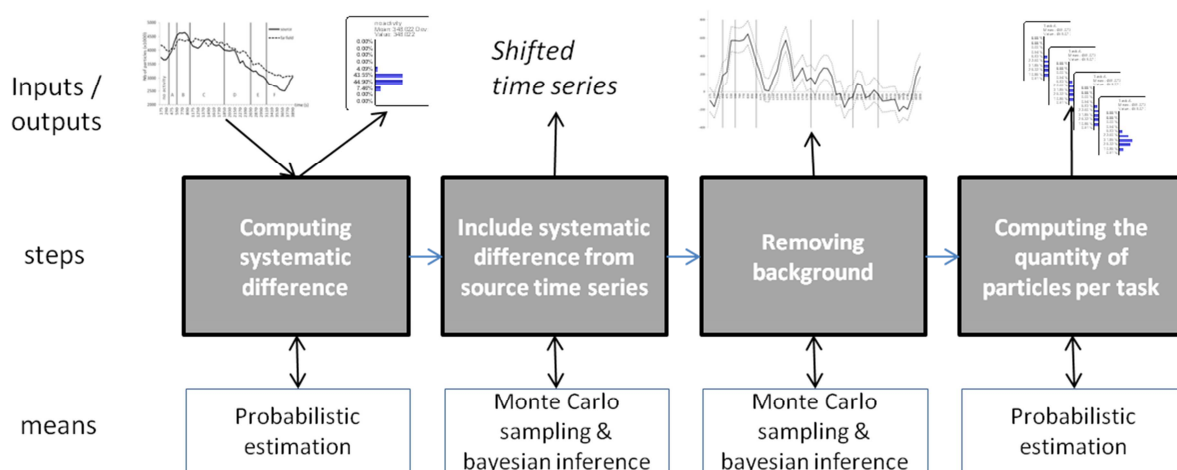

Figure 3. Overview of statistical analysis realized in this study.

3. Results

3.1. Computing the systematic difference

The systematic difference can be considered as a single value such as the difference of averages. Nevertheless, the number of values supporting the calculation of this difference is very small (4 data points). Consequently, it is more relevant to use a probability distribution as a systematic difference. In the same manner, the number of particles emitted by each task is presented as a probability distribution. This type of probabilistic computation can be done by sampling through Markov Chain Monte Carlo methods [9] (supported for instance by BUGS software [10]). It can also be done by exact computation through the use of a Bayesian network [11].

The aim of a Bayesian network is to provide a formal and graphical framework for representing knowledge and to support probabilistic calculations: Bayesian inference. A Bayesian network is an oriented graph whose nodes represent variables and whose arcs represent relationships between variables. The relationships between variables are quantified with conditional probabilities. The software used for Bayesian network design and inference is BayesiaLab [12]. If there are two arcs from variables "far field" and "source" to variable "no activity", the conditional probability table of "no activity" represent the distribution of values of "no activity" as a function of "far field" and "source". In the example (see Fig. 4), the function is "far field"-"source", the conditional probabilities of "no activity" are computed by sampling.

source	<=3705	<=3757	>3757
	50%	25%	25%

far field	<=4033	<=4107	>4107
	50%	0%	50%

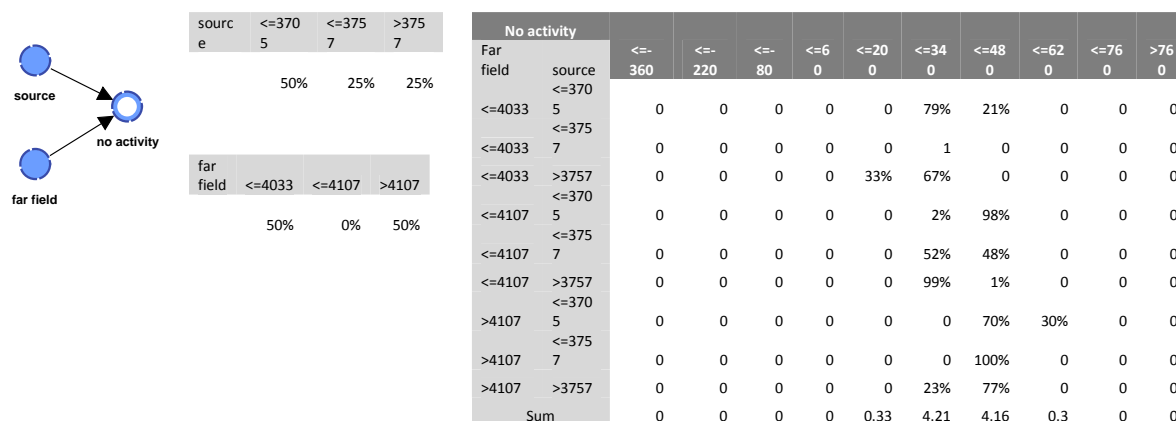| No activity		<=-360	<=-220	<=-80	<=60	<=200	<=340	<=480	<=620	<=760	>760
Far field	source										
<=4033	<=3705	0	0	0	0	0	79%	21%	0	0	0
<=4033	<=3757	0	0	0	0	0	1	0	0	0	0
<=4033	>3757	0	0	0	0	33%	67%	0	0	0	0
<=4107	<=3705	0	0	0	0	0	2%	98%	0	0	0
<=4107	<=3757	0	0	0	0	0	52%	48%	0	0	0
<=4107	>3757	0	0	0	0	0	99%	1%	0	0	0
>4107	<=3705	0	0	0	0	0	0	70%	30%	0	0
>4107	<=3757	0	0	0	0	0	0	100%	0	0	0
>4107	>3757	0	0	0	0	0	23%	77%	0	0	0
	Sum	0	0	0	0	0.33	4.21	4.16	0.3	0	0

Figure 4. Bayesian network used in the study (left), Conditional Probability Tables of "source" and "far field" nodes (middle), Conditional Probability Table of "no activity" node: systematic difference.

The nodes representing "far field" and "source" are darker because their conditional probabilities distributions come from the data; the difference "no activity" has a lighter patch because it is computed. The computation or *Bayesian inference* consists of transforming the "no activity" conditional probability table into a so call *marginal* distribution. In this simple case, it is the sum of all probabilities per column, which is normalized afterwards for obtaining the marginal probability distribution of differences when there is no activity (see Fig. 5).
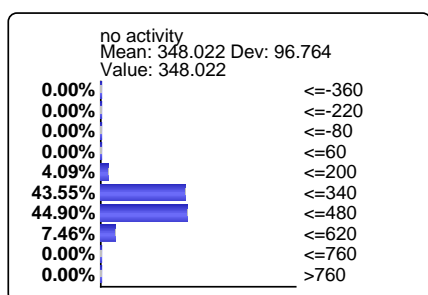
no activity	
Mean: 348.022 Dev: 96.764	
Value: 348.022	
0.00%	<=-360
0.00%	<=-220
0.00%	<=-80
0.00%	<=60
4.09%	<=200
43.55%	<=340
44.90%	<=480
7.46%	<=620
0.00%	<=760
0.00%	>760

Figure 5. Marginal probability distribution of "no activity" node. This is the distribution of systematic difference between "source" and "far field" observed when no activity is realized.
When there is no activity, the concentration measured at far field is higher than at source, which fits with the graphical observation that can be made from Fig. 2.

3.2. Removing systematic difference from source time series & computing time series of emitted particles

During the "no activity" period, the number concentration at the source is greater than that obtained in the far field of 348 $1/cm^3$ on average (mean indicator in Fig. 5). The probability of observing a number concentration within the range]200; 340] is 43,38%. This can be explained by that fact that the particles size ranges of the CPCs used were not exactly the same. Based on the "no activity" distribution i.e. the systematic difference between devices, the background aerosol part of the source time series can be removed. Graphically, with regards to Fig. 2, this means shifting the black curve upwards. With respect to the hypothesis, the number concentration C due to the activity is the number concentration measured at the source C_{source}, where the number concentration measured at far field $C_{far field}$ is removed and the systematic difference "no activity" is added. For adding the "no activity" systematic difference, 200 values are sampled out of the distribution for each time step. The real "far

field" and "source" time values are used. A total of 51 time steps are defined, which means that 51*200=10200 data points are sampled in total. The following algorithm describes this process.

For each time step t
              For i=1 to 200
              Sample d_i out of "no activity" distribution
              $C_i(t) = C_{source}(t) - C_{far\ field}(t) + d_i$
              End for
End for

The result C provided by using this approach is in the form of probability distributions (one per time step) based of the 200 C_i values. With this, it is possible to draw the time series of number concentration of released particles, considering the median and 5th percentile and 95th percentile as confidence interval for each time step as shown on the Fig. 6.

Figure 6. Number concentration of particles emitted at the source after the procedure of removal of the aerosol background as median value with 95%-confidence intervals.

On the basis of this time series, it is possible to quantify the number concentration of released airborne particles during each task through the probability distributions. By observing the graph, it seems that tasks A and B are generating a surplus of airborne particles, while the peak observed during task C may also create an average surplus of airborne particles. On the contrary, tasks D, E and F seem to have no impact. For quantifying these assumptions from graph data, the same methodology allowing building probability distributions of particle counts is applied. For each task, the number concentration of released particles at the source is depicted in Fig. 7. The statistical analysis shows that the mean number concentrations of released airborne particles are about 460 1/cm^3 and 540 1/cm^3 for task A and task B respectively. For the task C it is about 250 1/cm^3, while for the tasks D, E and F our analysis shows that no airborne particles was released.
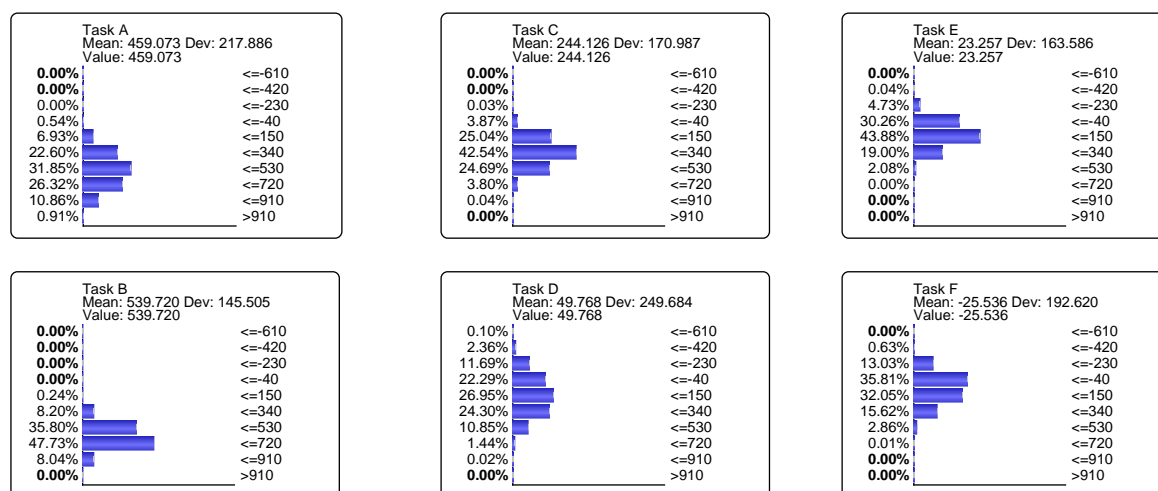
Figure 7. Probability distribution of the number concentration of release airborne particles for tasks A and B, C, D, E and F.

4. Discussion

It is important to note first of all that the measurement strategy used for collecting data was not established for this type of statistical analysis. This is why data present some specificity that complicates here the statistical analysis. For example, because of the difference in sampling frequency of CPCs, the data had to be combined, which dramatically reduced the number of concentration data on which the statistical analysis could be performed.

The results of the statistical analysis suggests that, overall, there is a positive effect of the cleanout operation by sandpapering of the reactor on the number concentration of particles measured by the CPC close to the source. However this effect is not positive or significant for all tasks, which correspond to sanding different parts of the reactor. For example, task A shows a mean number concentration of released airborne particles of about 460 $1/cm^3$, with a peak over 910 $1/cm^3$ (p=0.91%). For task C, the mean number concentration of released airborne particles is about 250 $1/cm3$ with peaks superior to 720 $1/cm^3$ (p=0.04%). The amount of airborne particles counted (and therefore released) during the following D, E and F tasks is around zero.

It is important to remember that the statistical analysis focused on measurement data using CPCs which are time resolved instruments giving no information on the nature of the detected particles.

5. Conclusion

Since measurements strategies [8, 12, 13, 14] published so far to characterize workplace exposure to manufactured nanoparticles put emphasis on the use real-time instrumentation for task-based evaluation of (size-resolved) aerosol concentrations, appropriate analysis of the results obtained from these instrumentation is a key factor.

In this exploratory study, it is assumed that the so-called background concentration is similar for source and far field sampling location: the far field concentration is not affected by the released

aerosol at source. A statistical method based on a Bayesian probabilistic approach was developed to analyze time resolved data. The Bayesian probabilistic approach allows a probabilistic modeling of data series, and the observed task is modeled in the form of probability distributions. The probability distributions issuing from time resolved data obtained at the source can be compared with the probability distributions issuing from the time resolved data obtained far-field, leading in a quantitative estimation of the airborne particles released at the source when the specific task is performed.

To illustrate the use of this statistical method, particle number concentration data from a workplace study that investigated the potential for exposure via inhalation from cleanout operations by sandpapering of a reactor were used. The results suggests that, overall, there is a positive effect of the cleanout operation by sandpapering of the reactor on the number concentration of particles measured by the CPC close to the source.

What is interesting in this approach is that, in the perspective of future limit values based on measurement results obtained in real time, it becomes possible to use a probabilistic indicator as the 95th percentile.

Beyond obtained results, this exploratory study indicates that the analysis of the results requires specific experience in statistics, and therefore highlights the need for an association between statisticians and specialists in aerosol metrology.

6. References

[1]   *(In french)* M. Lahmani and C. Brechignac 2006 *Les nanosciences. Tome 2 : Nanomateriaux et nanochimie* (Paris, FR:Belin)
[2]   G. Box and G. Jenkins 1976 *Time series analysis: forecasting and control* (San Francisco: Holden-Day)
[3]   D.H. Brouwer, J.H.J. Gijsbers and M.V.M. Lurvink 2004 *Ann Occup Hyg.* **48** 439
[4]   E. Demou, P. Peter and S. Hellweg 2008 *Ann Occup Hyg.* **52** 695
[5]   D.E. Evans, B. Ki Ku, M.E. Birch and K.H. Dunn 2010 *Ann Occup Hyg.* **54** 514
[6]   R.H. Klein Entink and W. Fransman 2009 *J Nanopart Res.* **13** 6991
[7]   Witschger O., Bau S., Bianchi B., Wrobel R. and Matera V. 2013. Journal of Physics: Conference Series This issue:
[8]   *(In french)* O. Witschger, S. Le Bihan, M. Reynier, C. Durand, A. Marchetto, E. Zimmermann and D. Charpentier 2012 *Hygiène et Sécurité du Travail* **226** 41
[9]   W. Gilks, S. Richardson and D. Spieghalter 1996 *Markov Chain Monte Carlo in Practice* (Chapman & Hall/CRC)
[10]   D. Lunn, D. Spieghalter, A. Thomas and N. Best 2009 *Statistics in Medicine* **28** 3049
[11]   D. Koller and N. Friedman 2009 *Probabilistic Graphical Models* (MIT Press)
[12]   BayesiaLab *www.bayesia.com* [Online].
[13]   D. Brouwer, B. van Duuren-Stuurman, M. Berges, E. Jankowska, D. Bard and D. Mark 2009 *J Nanopart Res.* **11**, 1867
[14]   G. Ramachandran, M. Ostraat, D. E. Evans, M. M. Methner, P. O'Shaughnessy, J. D'arcy, C. L. Geraci, E. Stevenson, A. Maynard and K. Rickabaugh 2011 *J. Occup Environ Hyg* **8** 673