

How to statistically analyze nano exposure measurement results: using an ARIMA time series approach

Rinke H. Klein Entink · Wouter Fransman ·
Derk H. Brouwer

Received: 17 June 2011 / Accepted: 7 October 2011 / Published online: 21 October 2011
© Springer Science+Business Media B.V. 2011

Abstract Measurement strategies for exposure to nano-sized particles differ from traditional integrated sampling methods for exposure assessment by the use of real-time instruments. The resulting measurement series is a time series, where typically the sequential measurements are not independent from each other but show a pattern of autocorrelation. This article addresses the statistical difficulties when analyzing real-time measurements for exposure assessment to manufactured nano objects. To account for autocorrelation patterns, Autoregressive Integrated Moving Average (ARIMA) models are proposed. A simulation study shows the pitfalls of using a standard *t*-test and the application of ARIMA models is illustrated with three real-data examples. Some practical suggestions for the data analysis of real-time exposure measurements conclude this article.

Keywords Nano exposure measurements · Time series · ARIMA · Environmental, health and safety (EHS) effects

Introduction

Assessment of the exposure of workers to manufactured nano particles at the workplace receives considerable attention, because the number of workers involved with nanotechnology is increasing rapidly. In the absence of sufficient scientific knowledge on the hazardous potential and possible health effects of exposure to nano particles, exposure measurements are performed to locate sources of emission and to characterize exposure in different work situations to gain knowledge on how to reduce personal exposure levels. However, there is no consensus method yet on how to (statistically) analyze and report these exposure measurement results.

Although real-time measurements in exposure assessment are found in, for instance, studying exposure to noise, measurement strategies for exposure to nano-sized particles differ from the majority of traditional integrated sampling methods by the use of real-time instruments. Common instruments for on-line measuring number concentrations or surface area concentrations of (nano) particles are the SMPS, CPC and ELPI, and diffusion chargers, respectively (Brouwer et al. 2004), which have various response times ranging between $t = 1, \dots, 180$ s. That is, every t seconds, a measurement result (number concentration) is recorded. The resulting measurement series, therefore, is a *time series*, a set of measurements collected sequentially over time. Typically, in time series the recorded set of measurement results are not

R. H. Klein Entink (✉) · W. Fransman · D. H. Brouwer
TNO, Research Group Quality and Safety,
P.O. Box 360, 3700 AJ Zeist, The Netherlands
e-mail: rinke.kleinentink@tno.nl

independent and show significant autocorrelation between subsequent samples.

The current literature on exposure to manufactured nano objects discusses several ways of statistically analyzing data obtained from real-time exposure measurements. Brouwer et al. (2004) identified the effect of work activities on particle number concentration and the percentage of ultrafine particles graphically. Although the authors showed that useful information can be retrieved, graphical analysis is limited to making qualitative inferences. Demou et al. (2008) repeatedly collected time series data on 20 days with the same production process. The analysis of the data was done by averaging the 20 time series and making a graphical analysis. Bello et al. (2009) mentions the testing of mean differences at a significance level of $P < 0.05$, but the article is unclear about what kind of test was performed. Park et al. (2010) used *t*-tests for evaluating mean differences of time series data. There are two potential problems with using *t*-tests: First, the mean of a time series only has a substantial interpretation when the time-series is stationary, that is, the time series is a random fluctuation around the mean and there are no trends in the data. Second, the autocorrelation in the measurements leads to underestimation of the variance in the data, resulting in a biased test statistic of the *t*-test. Subsequently, Park et al. (2010) modeled the data using random effects models for which they assumed a compound symmetry covariance structure. Such a covariance structure assumes that all measurements have the same correlation with each other. This implies, for instance, that two measurements with a time lag of two hours are equally correlated with each other as two measurements with a time lag of two minutes. Clearly, this assumption is not appropriate. Evans et al. (2010) emphasized the usefulness of real-time measurements, but also these authors only reported graphical inferences. Pfefferkorn et al. (2010) used (partial) autocorrelation assessments, Autoregressive Integrated Moving Average (ARIMA) time series models and first-order autocorrelation models to analyze their data. ARIMA models are well-known statistical models to deal with autocorrelated time series observations that we also propose in this article. ARIMA methods for analyzing time series data have also been used in environmental studies to (the effects of) air pollution. Recent examples are, for instance Sharma et al. (2009) and Mann et al. (2010). Note that

in this article we focus on the statistical modeling and analysis of real-time nano exposure measurements. As such, the statistical methods are only descriptive and not explanatory. For a good understanding of what is happening at a workplace, it is necessary to combine mechanistic models with data collection and statistical analysis, but this is outside the scope of this article. For a mechanistic modeling approach of exposure to nano-sized particles, see for instance Schneider et al. (2011).

The statistical analysis of time series requires special attention, since the measurements are not independent from each other. This has been addressed in the late 1980s by O'Brien et al. (1989), who recommended a check for autocorrelation for real time (dust) exposure data. When studying exposure to manufactured nano objects, people are confronted with several measurement series of different nano processes. Relevant research questions then involve testing for elevated exposure levels when a certain task is performed, or evaluating the similarity in trends and levels of repeated experiments. Other questions may relate to the identification of exposure determinants from a set of measurements. From the mentioned literature above, it is clear that exposure measurement time series have been dealt with in different ways, of which some are inappropriate for such data, and others are incomplete in checking specific (but important) assumptions of the applied models. In this article, we aim to show the difficulties of analyzing and interpreting time series data in the context of exposure assessment. First, we briefly introduce the statistical theory of time series analysis. Second, by means of simulated data examples we highlight some potential problems of dealing with nano exposure time series data. Third, we illustrate the statistical analysis with some real data examples from exposure measurements on nano-sized particles. We conclude this article with a discussion and some practical recommendations for real-time measurements and the analysis of such data.

Statistical analysis of time series

The first step in analyzing time series data is making a graph, which serves to quickly identify peaks and trends in the data. Subsequently, a potential next step is then fitting a model to the data to make quantitative inferences.

The modeling and analysis of time series data has frequent application in fields as economics (e.g., stock exchange data), geography, and engineering. A foundation for the statistical analysis of such data was Box and Jenkins (1970), who developed and applied Autoregressive Moving Average (ARMA) and ARIMA models. An introductory text is Cowpertwait and Metcalfe (2009). A Bayesian statistical approach to time series can be found in West and Harrison (1997). In this article, we have applied ARIMA models to real-time nano exposure measurements. In this section, we will first introduce ARIMA models and then give a brief overview of the model fitting procedure on the basis of a simulated example.

ARIMA models

Let $Y_t, t = 1, \dots, T$ denote a sequence of measurements of a variable Y at subsequent and equally spaced times t . The autoregressive (AR)-part of an ARIMA model refers to the regression of Y_t on time lags of itself. That is, it expresses the time series as a linear function of its past values. It is common to denote the order of the model as the number of time lags p , or AR (p). The simplest AR model is the first-order autoregressive, or AR (1), model

$$Y_t - a_1 Y_{t-1} = \mu + e_t, \quad e_t \sim N(0, \sigma^2), \tag{1}$$

where a_1 is the coefficient of the autoregression, μ is an intercept, and e_t the residual error term which follows a normal distribution with mean 0 and variance σ^2 . A value of a_1 close to 1 or -1 denotes a high autocorrelation in \mathbf{Y} , and a value of a_1 close to 0 denotes little autocorrelation in \mathbf{Y} .

The moving average (MA) part of the ARIMA model denotes the structure on the error term. The simplest MA(q) model is the first-order MA(1) model, with $q = 1$ denoting the order, given by

$$Y_t = \mu + e_t - c_1 e_{t-1}, \quad e_t \sim N(0, \sigma^2), \tag{2}$$

where c_1 is the moving average coefficient of the first order. A value of c_1 close to 1 or -1 denotes a high autocorrelation of the error term. For a value of c_1 close to 0 the model reduces to an ANOVA model with an intercept and random error component.

The integrated (I)-part refers to the order of differencing of a time series. In case of a non-stationary time series, differencing can be applied to

remove trends from the data to obtain a stationary series that subsequently can be modeled with AR and MA terms. The idea is that a trend in \mathbf{Y} can be accounted for by taking the derivative (dY/dt) of the series, which then might be stationary. Since the measurements are at discrete time steps $t = 1, 2, \dots, T$, a model of order $d = 1$ is the first-order (or I(1)) model that models the differenced series of \mathbf{Y} with lag 1: $dY/dt = Y_t - Y_{t-1}$.

From Eq (1) and (2) it can be seen that several assumptions are made in an ARMA model, of which we want to stress two explicitly: First, the ARMA model assumes that the time series is stationary. A stationary process is defined as a stochastic process whose probability distribution is not a function of time. This means that parameters like the mean and the variance of the series are constant over time. In other words, it is assumed that there is no trend or seasonal variation present in the data \mathbf{Y} . An example of a stationary process in a workplace situation would be a time series showing a constant particle concentration over time with only some random fluctuation. A non-stationary example could be an increasing particle concentration in the first hour after starting up a production process. Second, it is assumed that the error terms follow a normal distribution with constant variance. Those assumptions have important practical implications when evaluating exposure measurements. If a series of exposure measurements is not stationary, sample statistics like the mean, variance, and correlations with other variables are *not* meaningful since they are dependent on the length of the measurement series. This is quickly seen when considering a series with an increasing trend over time, e.g., the series has a positive slope over time. In that case, both the estimates for the mean and the variance will grow with sample size over time. As a consequence, neither the correlations with other variables are well-defined, nor are comparisons like a t -test of any significance. Symanski and Rappaport (1994) investigated autocorrelation and stationarity of exposure measurements (not on the nano scale) and noted that in assessing occupation exposure, summaries like the mean and variance components play an important role, illustrating the importance of assessing stationarity, see also Rappaport (1991).

Note that the distribution of the error term is assumed to be normal, it says *not* that the (empirical) distribution of \mathbf{Y} is necessarily normal. Therefore,

these assumptions are very important to check before proceeding with any inferences on the time series.

The combination of the AR(p), I(d), and MA(q) parts specifies an ARIMA(p,d,q) model, where p, d, and q denote the orders of the specific model terms. Selection of the appropriate orders is the topic of the next subsection.

Estimation and software

In most statistical software packages, standard routines are available for time series modeling with ARIMA models. For the analyses in this article, we used the `arima()` function available in the `stats-` package in the free statistical environment R (R Development Core Team 2011).

Stepwise approach for analyzing time series data

This section describes a stepwise approach for statistically analyzing time series data from nano exposure measurements using ARIMA models.

Step 1

The first step in fitting an ARIMA model is checking the stationarity of a time series. A standard procedure is to study a time series plot of the data together with the autocorrelation function (ACF, more details below). With a graph of the data one can visually check the constancy of the mean and variance (Fig. 1a). A plot of the autocorrelation of the data can also be informative (Fig. 1b). Typically, when the sample autocorrelation is high initially (>0.8) and shows a very slow decay to zero (e.g., over more than ten time lags) it is a sign for non-stationarity of the data. Then, difference the data once (corresponding to an ARIMA(0,1,0) model) and see if the differenced series appears stationary. An exponential trend in the data, alongside an increasing variance with time, is an indicator for a multiplicative relationship instead of (linear) additive growth. In that case, a log-transform of the data is useful to obtain a linear growth and to stabilize the variance.

To quantitatively test if a time series is stationary or a differencing step might be necessary first, one can test the hypothesis $H_0: a_1 = 1$ versus the alternative $H_a: a_1 < 1$. To do so, an AR(1) model can be fitted and compared to a first-order differencing or ARIMA(0,1,0)

model. In case the series is non-stationary, the AR(1) coefficient will be close to 1. For comparison of the AR(1) and the ARIMA(0,1,0) models, the AIC model fit criterion can be used (explained below in Step 2). In case the assumption of stationarity is reasonable for the data at hand, the AIC will favor the AR(1) model over the ARIMA(0,1,0) model. For a worked example, see Example 2.

When, after differencing, the series is stationary, proceed with Step 2. If it is not possible to obtain approximate stationarity of the series, ARIMA models are not appropriate. Then, qualitative graphical inferences are possible, or other statistical methods have to be considered.

Step 2

In the second step, the orders p and q of the ARMA(p,q) components need to be determined. Here, the ACF and the Partial Autocorrelation Function (PACF) play an important role. The Cross Correlation Function (CCF) is related to the ACF, but estimates the correlation between two different series.

The ACF is the correlation of a variable with itself at different times. For example, the autocorrelation at lag 2 is the correlation between Y_t and Y_{t-2} . The CCF is estimated in exactly the same way as the ACF, but with the difference that it is not the autocorrelation of the series with itself, but the correlation of two time series at lags $k = 0, 1, \dots, K$. The CCF is helpful to determine the similarity in the patterns of two time series, for instance in determining the similarity between repeated experiments (see also Example 3 in the “Empirical Examples” section). The PACF at time lag k is the correlation that remains after removing the effects of autocorrelation at shorter time lags. For example, the PACF at lag $k = 2$ is the autocorrelation that remains after correcting for the propagating effect of the autocorrelation at lag $k = 1$ (if there is an autocorrelation of 0.5 of lag $k = 0$ with lag $k = 1$, then this autocorrelation propagates to lag 2 since the samples at lag 1 and lag 2 are also correlated, resulting in an autocorrelation of lag 0 and lag 2 of $0.5 \times 0.5 = 0.25$).

To determine the orders p and q of the AR and MA model terms, plots of the ACF and PACF of the (differenced) data are typically used. It is best to look at the ACF and PACF together, and follow the following rules of thumb:

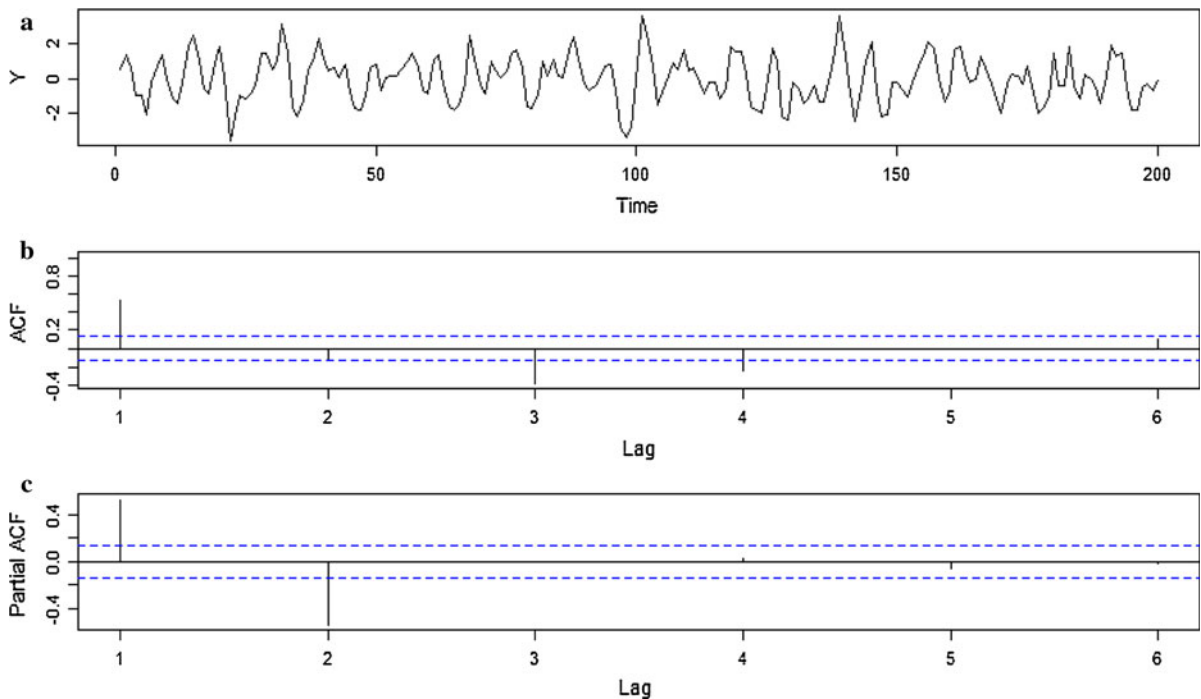


Fig. 1 Simulated time series ARIMA(2,0,0) with $n = 200$. From above to below: **a** measurement series against time, **b** autocorrelation function (ACF), and **c** the partial autocorrelation function (PACF)

- If the PACF has a sharp cut-off, then an AR term should be considered. The order of the autoregression is determined by looking at the PACF function, and checking after which time lag the PACF is approximately zero. For instance, if the PACF has significant spikes at time lags 1 and 2, but is (almost) zero at time lag 3 and higher, then an AR(2) model might be appropriate.
- For determination of the order (q) of the MA part the ACF is used in a similar way, where q is chosen based on significant (positive or negative) spikes in the ACF. The rule of thumb here is: if the ACF has a sharp cut-off, then an MA term should be considered. Again, the lag where the ACF cuts off corresponds with the order of the MA model part.

We illustrate the order selection procedure for the AR(p) component with the simulated example presented in Fig. 1. Fig. 1a shows the time series plot of an AR(2) process. In Fig. 1b the ACF is plotted, showing a significant positive spike at lag 1, and significant negative spikes at lags 3 and 4. However, it can be seen that the PACF plot shows a significant positive spike at the first time lag, and a negative spike at the second

time lag and is approximately zero afterward. Based on the PACF, it seems that after accounting for the autocorrelation at lags 1 and 2, no significant higher order terms are needed to describe this time series. An appropriate model then would be an AR(2) model.

It can sometimes be difficult in practice to select the best model based on a data sample and the estimated ACF and PACF. For instance, it can be hard to distinguish between an AR(1) or an MA(1) model by visual inspection of the ACF and PACF, and then decide which of the two models fits the data best. Therefore, information criteria like the Akaike Information Criterion (AIC) can be used to search for the best fitting model. The AIC is a relative measure of model fit that balances an improvement in model fit based on the log-likelihood with the number of added model parameters. As such, it is only useful to compare two (or more) models with each other, where the model with the lower AIC value is considered to provide a better fit to the data. An alternative to the AIC is the Bayesian Information Criterion (BIC), which differs from the AIC in how it penalizes for adding model parameters, but otherwise its interpretation is similar to the AIC. Most statistical software

packages provide such model fit statistics along with the estimated model parameters.

Step 3

ARIMA models can be extended to include covariates to explain observed effects or changes in the concentration level of a time series. For example, an indicator variable to model the effect of a process activity on exposure, where $X_t = 1$ denotes activity, and $X_t = 0$ denotes no activity. For an AR(1) model, the resulting equation then is:

$$Y_t - a_1 Y_{t-1} = \mu + (X_t - a_1 X_{t-1})\beta + e_t, \quad (3)$$

$$e_t \sim N(0, \sigma^2),$$

where β is the regression coefficient of concentration level \mathbf{Y} on indicator variable \mathbf{X} . In this case the autoregression on \mathbf{Y} also applies to \mathbf{X} . When a differencing step is necessary, note that also \mathbf{X} is differenced, i.e., the change in \mathbf{X} is related to the change in \mathbf{Y} . Effectively, this means that the ARIMA model is fitted to the errors of the regression of \mathbf{Y} on \mathbf{X} (note that without explanatory variables, the residual errors plus the mean term equal the observations \mathbf{Y}). The cross correlation between the series \mathbf{Y} and \mathbf{X} can be helpful to identify if there is a relationship between observed trends in \mathbf{Y} and a covariate \mathbf{X} .

The model in (3) can also be viewed as a linear regression model that accounts for the serial correlation in the measurements. In Eq. 3, the serial correlation is modeled by an AR(1) model, but a regression model with an MA structure on the error terms can also be used. However, it is important to evaluate the stationarity of the series before making inferences and conclusions from ARMA regression models, since both ARMA- and regression models (and their combination) assume stationarity with constancy of variance of the error terms. Non-stationary series typically violate such assumptions and may bias the estimated coefficients.

Step 4

When the time series appears stationary and the appropriate orders of the ARIMA model have been determined, model assumptions as the normality of the residuals and residual autocorrelation and have to be

checked. This step finalizes the model fitting and model checking procedure. Subsequently, interpretation of the results remains.

Testing for mean differences: standard t -test or ARIMA?

To show the influence of autocorrelation between measurements on statistical testing for mean differences (i.e., a standard t -test and an ARIMA regression model), a simulation study was performed. All data were simulated from an AR(1) model with $\mu = 0$ and $\sigma^2 = 1$, see Eq (1), for four different values of the autocorrelation, $a_1 = 0.3, 0.5, 0.7,$ and 0.9 . The length of the series was $N = 200$ samples. A switch in the mean level of the series occurred at $t = 101$, and two tests for the difference in mean level between the first 100 and the second 100 samples were performed. The first test was a standard t -test for a difference in means. In the second test, an AR(1) model was fitted, where an indicator variable modeled the mean difference between the second and the first half of the data. From the fitted AR(1) regression model, a t -test for the estimated mean difference was obtained from the estimated coefficient and its standard deviation for the indicator variable, corresponding to a regular t -test but with the important difference that autocorrelation between the samples was accounted for. For each condition (mean difference and autocorrelation) 100 data sets were simulated.

Figure 2 summarizes the results, where the averages over the 100 simulated data sets of the estimated t -values were plotted against the simulated mean differences. It can be seen that the t -test overestimated the statistical significance of the mean difference in all cases, where the overestimation was greater for increasing autocorrelation between the samples. Although the mean, variance, and sample size of the simulated data sets were chosen arbitrarily, the principle of this result stands for any data set where there is autocorrelation between subsequent samples. It shows that when testing for mean differences, autocorrelation in the data cannot be ignored. Neglecting autocorrelation can lead to false significant results, especially when mean differences are small or autocorrelation between samples is high.

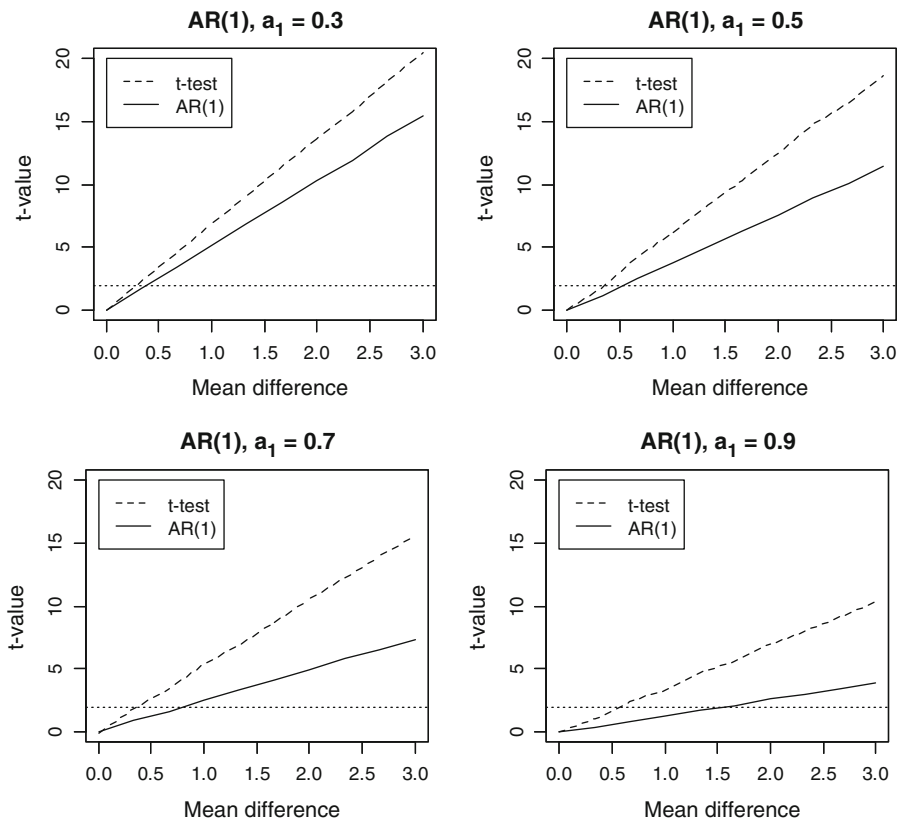


Fig. 2 Estimated *t*-value (y-axis) for mean differences (x-axis) using an AR(1) model (solid line) and a *t*-test (dotted line). AR(1) autoregression coefficient denoted by a_1 . Horizontal dots denote

$t = 2$, which corresponds to the lower bound for a statistical significant test result at the $\alpha = 0.05$ level

Empirical examples

In this section, we present three real data examples to illustrate the use of ARIMA models for (statistically) analyzing time-series of nano exposure measurement results.

Example 1: Testing for an effect of an activity on the particle number concentration level

A measurement series of a certain activity or task resulted in a time series of 1500 subsequent measurements of number concentration of particles smaller than 100 nm, using an ELPI on-line measurement device with a response time of 1 s. For simplicity, we refer to this measurement series as “Example 1” from now on. A time series plot of the data is shown in Fig. 3a, in which a dotted line denotes the region where the task or activity was performed during the measurement period. The research question was to investigate a potential rise

in exposure to nano-sized particles (<100 nm) during performance of the activity compared with the non-activity period.

Step 1

First, the stationarity of the series is evaluated. From Fig. 3a, there are no suggestions for a non-stationary process, since the measurements seem to fluctuate around a constant value. In addition, Fig. 3b shows that the ACF quickly decays to zero (in 4 time lags). Therefore, from the time series plot and ACF it was concluded that it is safe to assume that the time series is stationary.

Step 2

Since the ACF (Fig. 3b) indicates that there is significant autocorrelation in the data, the order of the ARMA components has to be selected. The ACF

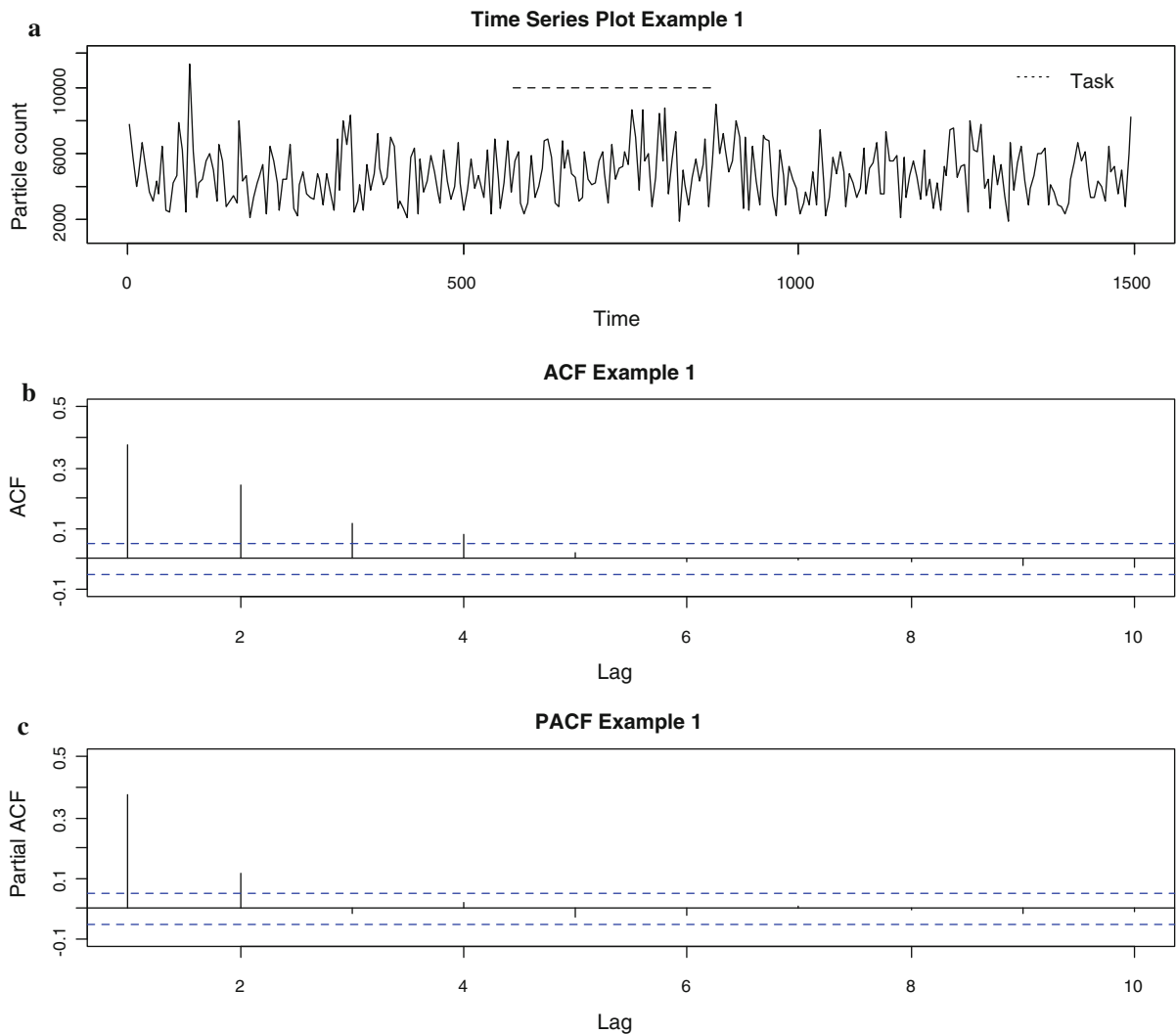


Fig. 3 From above to below **a** measurement series against time **b** autocorrelation function (ACF), and **c** the partial autocorrelation function (PACF)

shows significant spikes at the first four time lags. A possible MA model of order $q = 4$ could therefore be fitted to the data. However, the PACF suggests that there is a significant first-order autocorrelation, and probably a moderate second order component, after which the PACF is approximately zero. Therefore, the ACF and PACF suggest either an MA model of order 4, or an AR model of order 2. Since the AR model needs 2 parameters less than the MA model to fit the data, the AR model seems to be the most parsimonious choice. Therefore, two models were subsequently fitted to the data, an AR(1) and an AR(2) model. The AIC was then used to select the order of the AR model. The AIC favored the AR(2) model

(AIC = 37462) over the simpler AR(1) model (AIC = 37483). For illustrative purposes, also an MA(3) model was fitted. The AIC for the MA(3) model was estimated at 37468, and thus also indicates that the simpler AR(2) model is more appropriate than the MA(3). Based on the combined information from the (P)ACF function and the AIC, it was assumed that the AR(2) model was most appropriate for this data.

Step 3

The cross correlation function of the indicator variable for the task (which equaled 1 at the times the task was performed and 0 otherwise) with the particle count was

estimated. The CCF showed a small positive correlation between the indicator variable and the particle count measurements, which was strongest at a lag of 26 time steps into the future. That is, the particle count responded 26 time steps later on a change in the task.

Step 4

Model diagnostics were evaluated and showed that the standardized residuals were approximately normal and that there was no significant residual autocorrelation.

Based on the above assumptions, we fitted an ARIMA(2,0,0) model to the data of Example 1, extended with the indicator variable for task as a covariate with a time lag of 26 s. The estimated effects, given in Table 1, show a statistically significant effect of the activity on exposure, indicating that exposure to nano-sized particles increases on average with 364 particles (<100 nm) when the activity is performed, with a time lag of 26 s.

Example 2: Testing for an effect of an activity on the particle number concentration level

As in Example 1, the interest is in testing the effect of an activity on the particle number concentration level. The time series plot, ACF and PACF of a measurement series of 260 samples of nano-sized particle concentration taken with a CPC with a response time of 20 s are shown in Fig. 4.

Step 1

The stationarity of the series was evaluated. The ACF showed a high autocorrelation at lag 1 of 0.83, declining slowly to 0.4 at lag 20. This is an indicator for non-stationarity of the series. However, the Partial ACF suggested this was a first-order process where the

Table 1 Estimated model parameters for the AR(2)-regression model for Example 1

Parameter	Estimate	SD	P value
Intercept	4660.4	79.4	<0.01
a_1	0.33	0.03	<0.01
a_2	0.12	0.03	<0.01
Task effect	364.4	176.7	0.04
Residual variance	2327827		
AIC	26258		

autocorrelation at lag 1 propagates through subsequent samples (Fig. 4c). It was debatable if a first-order differencing, or an AR(1) model was most appropriate for this data set. To investigate this, the hypothesis test described under Step 1 was used. Both the AR(1) and the ARIMA(0,1,0) model were fitted and their AIC-values were compared. The AIC's were 3157.7 and 3159.01, respectively, a small difference but favoring the AR(1) model. Also note that the first-order differencing corresponds to a (non-stationary) AR(1) model with coefficient $a_1 = 1$, while the estimated AR(1) coefficient was 0.83, with SD = 0.03, and thus was significantly different from 1 and indicates a stationary time series. Therefore, we assumed the AR(1) model was appropriate for this data set.

Step 2

It was investigated if there was residual autocorrelation remaining after the AR(1) model was fitted. However, the ACF of the residuals of the AR(1) model showed no significant autocorrelation anymore. Therefore, no additional AR or MA terms were needed.

Step 3

Testing for the influence on the particle concentration level of an activity that was performed at different times during the measurement period using the ARIMA(1,0,0) regression model, a positive but non-significant ($P > 0.05$) result was found (Table 2).

Step 4

Residual analysis showed no aberrant patterns in the residuals and there no significant residual autocorrelation was apparent.

To compare these results with results of a standard t -test, the estimated parameters are shown in Table 2 together with the results for a t -test for mean differences where the indicator variable for the activity was used as the grouping variable. For both tests a positive effect for the activity on particle number concentration of nano-sized particles was found. However, the ARIMA model (which accounted for the autocorrelation in the series) gave a non-significant result (at the level $P < 0.05$), whereas the t -test (which ignores the autocorrelation in the series) suggested a statistically significant difference ($P < 0.05$) between activity and

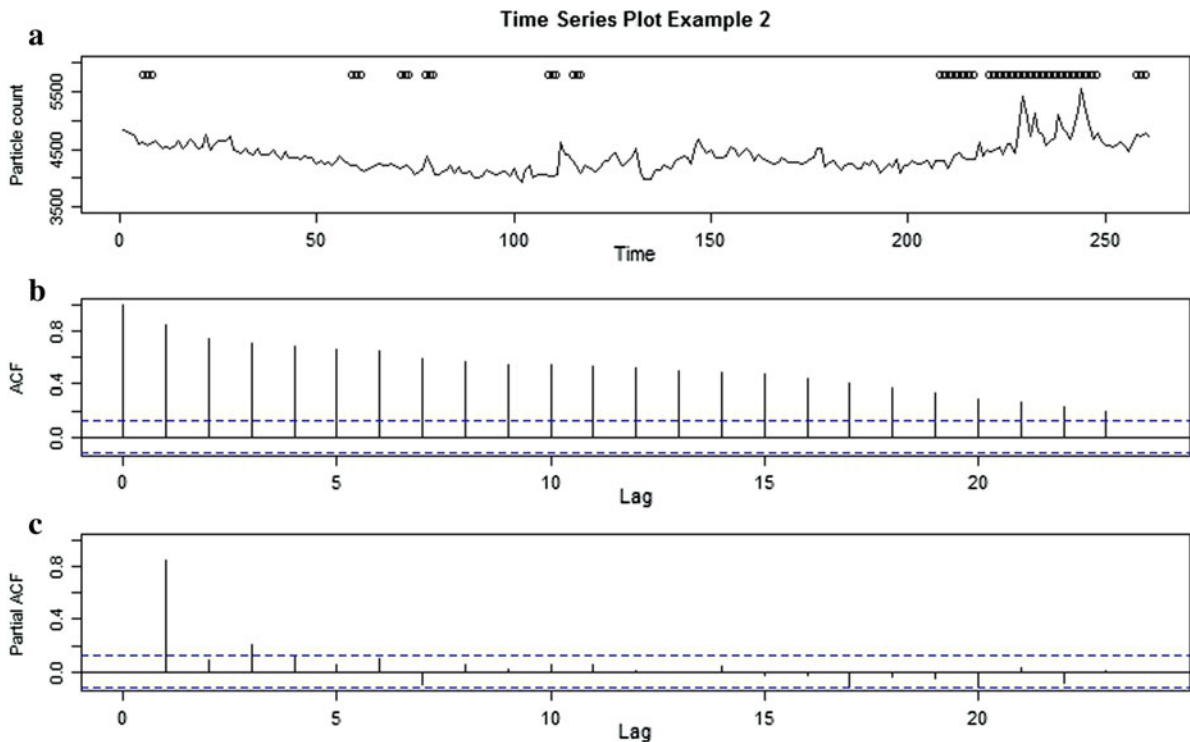


Fig. 4 **a** Time series plot Example 2, *dots* indicate activity. **b** ACF-plot Example 2 and **c** PACF Example 2

non-activity. This is in line with the results from our simulated time series presented in Fig. 2, which showed that the standard *t*-test overestimated the statistical significance of the mean differences in comparison with the ARIMA approach, when there are serial correlations in the data.

Example 3: Testing for a difference between three repeated experiments

In this example the data of an experiment that was repeated three times was analyzed. In a closed room, a

deodorant was sprayed for 3 s. Subsequently, the particle number concentration was measured during 12 min with a diffusion charger type of device (Nanotracer[®]) with a response time of 16 s. The experiment was repeated when the particle number concentration was back at the original background level. This resulted in three time series, which are plotted together in subsequent order in Fig. 5.

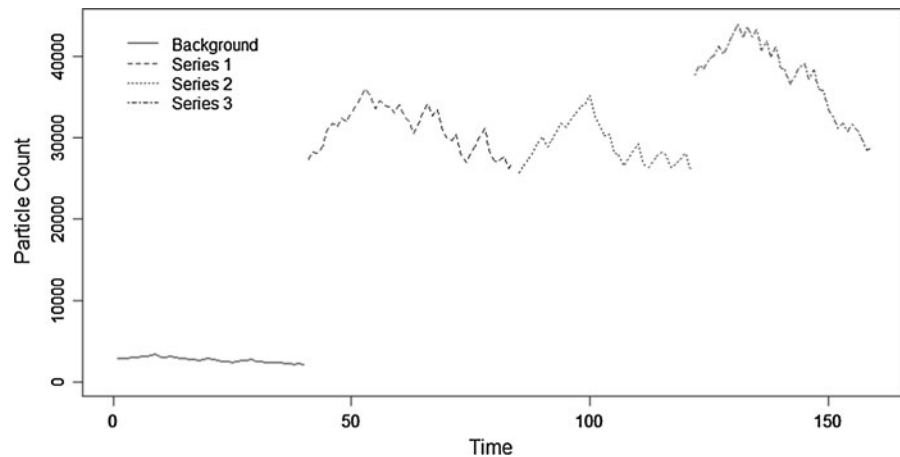
When investigating the similarity between two time series, two things are of main importance: 1) Do the series follow a similar pattern, and 2) Are there substantial absolute mean differences between the measurements?

To answer the first question, a graphical comparison was made first. From Fig. 5, it can be seen that the first two experiments show similar behavior, but that the third experiment (series 3) seems to differ. The CCF gives a quantification of the similarity of the series, and was estimated for all combinations of the series. The estimated CCF for series 1 with series 2 was 0.76, showing a substantial association between the trends of those two series. The estimated CCF between series 1 and 3, and the CCF between series 2 and 3, were estimated to be 0.66 and 0.69, respectively. These

Table 2 Estimated parameters for Example 2

	Parameter	Estimate	SD	<i>P</i> value
ARIMA-model	Intercept	4391.9	58.8	<0.01
	a_1	0.87	0.03	<0.01
	Task effect	36.9	31.4	0.25
	Residual variance	16124		
<i>T</i> -test for equal means	<i>t</i> -Value	−2.33		0.02

Fig. 5 Example 3: Time series plot of the three repeated experiments and background measurement



values are somewhat lower, but correspond with the observation that all series first show an upward trend (an increase in particle number concentration after spraying with the deodorant), followed by a decreasing particle number concentration. Therefore, it can be concluded from the CCF and the graphical inspection that the three series follow a similar pattern.

However, similar trends do not have to correspond with similar concentration levels. To test this (and answer question 2), an ARIMA-regression model was fitted to the data with two dummy variables with series 1 being the reference series with which the mean levels of series 2 and 3 were compared.

Step 1

The stationarity of the series was investigated before fitting the ARMA regression model. The ACF between the measurements was very high, showing a slowly decreasing ACF with an autocorrelation of 0.93 at lag 1. Taking into account the increasing trend of the three series, followed by the decrease, this points at non-stationarity of the series. Therefore, a first-order differencing step was taken.

Step 2

The ACF and a residual plot of the differenced series were inspected, to see if there was autocorrelation remaining. The residual plot and its ACF are shown in Fig. 6. From these plots it can be seen that there is no significant autocorrelation remaining, and the residuals appear to be randomly. Therefore, it was concluded that no additional AR or MA terms were needed.

Step 3

The regression model was fitted to the (differenced) data. The estimates of the parameters are given in Table 3, which demonstrate that series 2 tends to show a lower particle concentration level than series 1, however, this is not statistically significant. Series 3 on the other hand, clearly shows a positive offset compared with series 1, indicating that the measured number concentrations in the third repeat of the experiment were substantially higher.

Step 4

The residual analysis did not indicate deviations from normality of the residuals and no significant residual autocorrelation was apparent.

In conclusion, the results suggest that the first two repeats of the experiment were very similar, both in absolute particle number concentration levels as well as observed pattern. However, the third repeat of the experiment differed from the other two with a statistically significant ($P < 0.05$) higher particle number concentration of on average $10,278 \text{ \#/cm}^3$ in experiment 3 compared with experiment 1 (Table 3).

Discussion

This article showed that statistically analyzing results of on-line measuring particle number concentration levels of manufactured nano objects in real-time (i.e., a time series) needs considerable attention. Typically in time series, the recorded set of measurement results

Fig. 6 Residual plot and ACF of the time series (Example 3) after a first-order differencing step

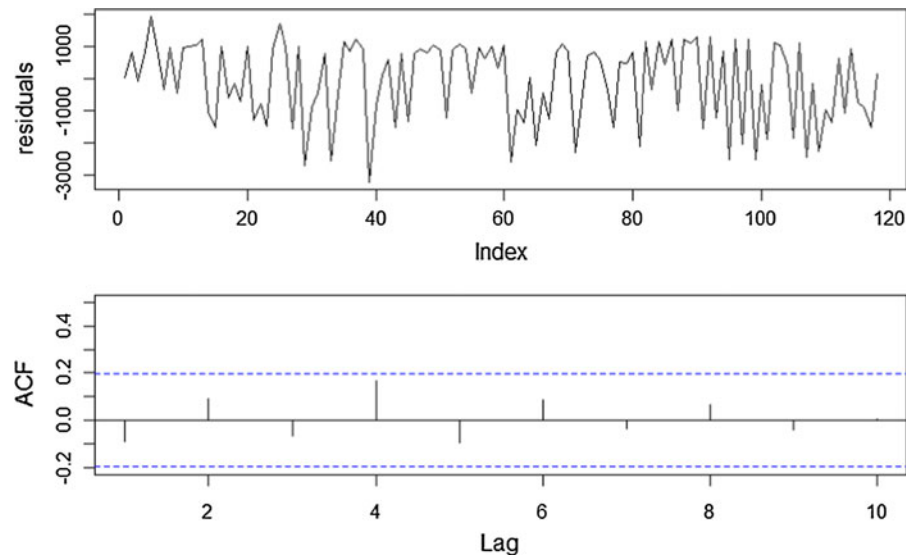


Table 3 Estimated model parameters for the AR(2)-regression model for Example 3

Parameter	CCF	Estimate	SD	<i>P</i> value
Difference series 1–2	0.76	−1336.0	1218.0	0.76
Difference series 1–3	0.66	10278.0	1722.5	<0.01
Difference series 2–3	0.69			

are not independent and show significant autocorrelation between subsequent samples. The aim of this article was to raise awareness for the potential statistical difficulties when analyzing real-time measurements for exposure assessment to manufactured nano objects. It was shown how ARIMA models can be applied to make inferences from real-time exposure measurements, where the dependencies between subsequent measurements were accounted for. This offers a more sophisticated approach compared to calculation of the difference between a predicted exposure revealed from a regression equation minus the observed exposure and an evaluation of the residual values, as proposed by O’Brien et al. (1989). Our simulations demonstrated that the autocorrelation in time series data results in an overestimation of the statistical significance of the mean difference when using a standard *t*-test. To account for the autocorrelation in the time series data, we propose a stepwise approach for statistically analyzing time series nano exposure measurement data using an

ARIMA model. As was shown in the empirical examples, these models can be used to identify the effect of a specific activity on concentration levels, or to compare the concentration level and/or pattern of multiple (repeated) experiments.

Although often applied to time series problems, ARIMA models are not the only methods for analyzing time series data. In signal processing and analysis, often modeling is done within the frequency domain instead of in the time domain. Also, more advanced dynamic models exist, that allow model parameters to vary over time (West and Harrison 1997, Prado and West 2010). These models are more flexible in describing, for instance, non-stationary series.

Sample size influences the accuracy of estimation. In principle, for a model with number of parameters m , at least $m + 1$ samples are required for estimation (Hyndman and Kostenko 2007), however this will generally be too small a sample size for a reasonable accuracy of estimation. It is difficult to give simple guidelines for sample size, since these will depend on both model complexity and the amount of random variation in the data (the error term). Nevertheless, based on a simulation study we performed (not reported here), we recommend a sample size of at least 40 for proper estimation of an AR(2) regression model. In our real data examples, sample sizes were large (>100) and therefore sample size was not of concern. The aim of the study should also be considered. When interest is in testing for a specific

activity effect, the total measurement time should be long enough for the process to reach a new steady state after an induced change.

In time series data, the response time of the real-time measurement instrument directly influences the properties of the series, where shorter response times lead to greater sample sizes and higher autocorrelations between samples. Typically, this results in a greater overestimation of the statistical significance when using a standard *t*-test for testing mean differences. By taking into account the autocorrelation in the data by using an ARIMA model, this overestimation can be overcome. Also, it is important to know if the output of the measurement instrument is the raw data, or that a transformation step has been performed. For instance, when the instrument averages over measurements, this reduces the observed variance in the data.

The proposed use of ARIMA models for analysis of real-time nano exposure measurements has the drawback of requiring more statistical experience from the researcher. However, these models are implemented in most standard statistical software packages. Identifying the appropriate order for an ARIMA(p,d,q) model for the observed data may be a difficult task, but for instance the *forecast*-package for use in the free R-statistical environment provides the opportunity of automated model selection. What cannot be done automatically, however, is a graphical evaluation of the collected data by the researcher. Making a time series plot of the data should always be the first step in an analysis, since it reveals trends and peaks in the data. The observed pattern should make sense to the researcher before proceeding with fitting any model or making inferences. To facilitate the statistical analysis of nano exposure real-time measurement results by using ARIMA models, we proposed a stepwise approach to assist researchers in applying these types of models on time series data. Currently, the authors are working on a practical guideline for analyzing real-time exposure measurements that will be made freely available.

In exposure assessment, interest is often in scenario-wide assessments where differences between companies and workers are studied. It is possible to extend ARIMA models with a random-effects structure to specify between-company and between-worker variance components. This is something we want to explore in the near future. Finally, the applications in

this article all relate to task based measurements. However, the proposed methodology can equally well be applied to shift-based measurements.

Acknowledgments We thank two anonymous reviewers for their helpful comments that improved an earlier version of this article.

References

- Bello D, Wardie BL, Yamamoto N, Guzman deVilloria R, Garcia EJ, Hart AJ, Ahn K, Ellenbecker MJ, Hallock M (2009) Exposure to nanoscale particles and fibres during machining of hybrid advanced composites containing carbon nanotubes. *J Nanopart Res* 11:231–249
- Box G, Jenkins G (1970) *Time series analysis: forecasting and control*. Holden-Day, San Francisco
- Brouwer DH, Gijsbers JHJ, Lurvink MWM (2004) Personal exposure to ultrafine particles in the workplace: exploring sampling techniques and strategies. *Ann Occup Hyg* 48: 439–453
- Cowpertwait PSP, Metcalfe AV (2009) *Introductory time series with R*. New York, Springer
- Demou E, Peter P, Hellweg S (2008) Exposure to manufactured nanostructured particles in an industrial pilot plant. *Ann Occup Hyg* 52:695–706
- Evans DE, Ki Ku B, Birch ME, Dunn KH (2010) Aerosol monitoring during carbon nanofiber production: mobile direct-reading sampling. *Ann Occup Hyg* 54:514–531
- Hyndman RJ, Kostenko AV (2007) Minimum sample size requirements for seasonal forecasting models. *Foresight* 6:12–15
- Mann JK, Balmes JR, Bruckner TA, Mortimer KM, Margolis HG, Pratt B, Hammond SK, Lurmann FW, Tager IB (2010) Short-term effects of air pollution on wheeze in asthmatic children in Fresno, California. *Environ Health Perspect* 118:1497–1502
- O'Brien DM, Fischbach TJ, Cooper TC, Todd WF, Gressel MG, Martinez KF (1989) Acquisition and spreadsheet analysis of real time dust exposure data: a case study. *Applied Industrial Hygiene* 4:238–243. doi:10.1080/08828032.1989.10388570
- Park JI, Ramachandran G, Raynor PC, Eberly EB, Olson G (2010) Comparing exposure zones by different exposure metrics using statistical parameters: contrast and precision. *Ann Occup Hyg* 55:1–14
- Pfefferkorn FE, Bello D, Haddad G, Park J-I, Powell M, McCarthy J, Bunker KL, Fehrenbacher A, Jeon Y, Virji MA, Gruetzmacher G, Hoover MD (2010) Characterization of exposures to airborne nanoscale particles during friction stir welding of aluminium. *Ann Occup Hyg* 54(5):486–503. doi:10.1093/annhyg/meq037
- Prado R, West M (2010) *Time series: modelling. Computation and Inference*. Chapman & Hall/CRC, New York
- R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna. <http://www.R-project.org>
- Rappaport SM (1991) Assessment of long-term exposures to toxic substances in air. *Ann Occup Hyg* 35:61–122

- Schneider T, Brouwer DH, Koponen IK, Jensen KA, Fransman W, van Duuren-Stuurman B, van Tongeren M, Tielemans E (2011) Conceptual model for assessment of inhalation exposure to manufactured nanoparticles. *J Expo Sci Environ Epidemiol* 21:450–463
- Sharma P, Chandra A, Kaushik SC (2009) Forecasts using Box–Jenkins models for the ambient air quality data of Delhi city. *Environ Monit Assess* 157:105–112
- Symanski E, Rappaport SM (1994) An investigation of the dependence of exposure variability on the interval between measurements. *Ann Occup Hyg* 38:361–372
- West M, Harrison PJ (1997) *Bayesian forecasting and dynamic models*, 2nd edn. Springer, New York