A Stochastic Differential Equation for Exposure Yields a Beta Distribution

MICHAEL R. FLYNN*

CB7431 Rosenau Hall, Department of Environmental Sciences and Engineering, School of Public Health, University of North Carolina, Chapel Hill, NC 27566-7431, USA

Received 22 November 2003; in final form 20 January 2004; published online on 7 July 2003

This paper presents a stochastic differential equation for exposure based on a modified version of the standard dilution ventilation equation. An equilibrium solution is obtained with the assumption that variability in the rate of change of concentration is proportional to the product of concentration and one minus concentration. Appropriate definitions for concentration are used to ensure a physically consistent model. The probability distribution for exposure that results is the standard beta distribution. This model is supported by several exposure data sets, which fit the beta distribution well. Issues regarding parameter estimation for the beta distribution, and application of the model are presented. Recommendations are made for simultaneously collecting contaminant generation rate information, ventilation rates, and time-dependent breathing-zone tracer concentrations, in addition to the exposure data.

Keywords: beta distribution; exposure modeling; stochastic differential equation

INTRODUCTION

Two important issues for occupational hygienists are assessing the risk of disease associated with a given exposure, and the subsequent problem of implementing successful controls to minimize this risk. The common variable is exposure and hence the need for quantitative models to describe it as a function of its determinants. This is essential since it is through manipulation of the determinants that successful control is achieved. It is also important from the assessment perspective since exposure estimates must often be inferred from data available on the determinants (ventilation records, production rates, etc.) but not exposure itself. The scientific methods that hygienists rely upon to construct exposure models exist over a spectrum characterized at one end by the empirical–statistical approach, and at the other by the theoretical-deterministic view.

The empirical-statistical methodology relies primarily upon observation (measurement) and statistical characterization of the data. When used with rigorous statistical techniques, and informed by fundamental principles, this methodology can produce useful descriptive models. There are drawbacks however: prediction beyond the limits of the data is problematic, and selecting the right model to fit the data is not known a priori. In addition, the statistical model is one that describes the measurement of exposure, not exposure itself and thus experimental variability is also modeled. The major strength of this approach is that it captures the randomness of the variable in question, and a statistical distribution of exposure results, not simply a point estimate. Much of this work employs the two-parameter lognormal distribution with general linear models. Some examples of this approach are found in Rappaport et al. (2003), Lyles et al. (1997) and Symanski et al. (2001).

The theoretical-deterministic perspective relies on fundamental first principles and involves solving the differential equations governing the transport of contaminants either at the macro-scale (control-volume analysis), or the differential scale (computational fluid dynamic simulations). Recent examples of studies employing this approach include Flynn and Sills (2000, 2001), and Hyun and Kleinstreuer (2001). The strengths of this approach are that the model is known a priori, and that exposure can be predicted as a function of the primary determinants. It is not constrained by the calibration-range problem associated with the empirical-statistical approach. The main drawback is the difficulty in incorporating...
all of the random events that affect exposure. The time and expense of conducting Monte-Carlo simulations to address this deficiency, in a realistic way, is usually prohibitive.

The two approaches complement each other in that strengths of one tend to be weaknesses of the other and vice versa. It is not surprising that some exposure models incorporate elements of each. For example, attempts to integrate dimensional-analysis into an empirical-conceptual approach for exposure modeling have been made (Carlton and Flynn, 1997; Flynn et al., 1999). A desirable exposure model is deterministic in the sense that it relates concentration to the fundamental determinants of air flow and contaminant generation rate. However, it should also be capable of producing a statistical distribution of exposure to address the uncertainties and randomness of the real-world process. This describes a stochastic differential equation model of exposure in which both aspects can be integrated.

**THEORY**

The theoretical aspects of this paper are organized into three subsections that follow. The first deals with the preliminaries of definitions, and statistical issues in describing exposure, the second presents a deterministic governing equation for concentration, and the final section deals with the stochastic differential equation and its solution.

**Preliminaries**

Exposure to an airborne contaminant is defined as the time-weighted average breathing-zone concentration \( C_{TWA} \) over an interval \( (T) \), or mathematically:

\[
C_{TWA} = \frac{1}{T} \int_{0}^{T} C \, dt
\]  

(1)

Although concentration may be reported as mass per unit volume, in reality it is always a dimensionless fraction that lies between 1 and 0. This distinction is often lost when values are reported in p.p.m. or mg/m³, but it is important here with regard to the formulation of a physically consistent exposure model. All concentrations can be transformed into a standard fraction between 0 and 1. Two potential candidate distributions that fulfill this requirement, and which fit exposure data reasonably well, are the beta distribution (Flynn, 2004) and the four-parameter lognormal model.

**Governing equation**

A fundamental mass balance produces the usual differential equation describing the time derivative of the concentration \( C \) of a gas or vapor being generated at a rate \( G \) in a volume \( V \), that is ventilated at a rate \( Q' \), which has been adjusted for mixing by a factor \( K \). It should be noted that this \( K \) is purely a mixing factor and does not include a safety correction, as is often the case (e.g. Burgess et al., 1989):

\[
\frac{dc}{dt} = \frac{G - Q'C}{V}
\]

(4)

This equation is often referred to as the dilution ventilation equation (ACGIH, 1998) and a variety of approximations yield a host of solutions, although of somewhat limited value for exposure modeling. In particular, \( C \) is the average concentration in the space and the mixing factor \( K \) is difficult to determine. However, on a local scale, the equation can be thought of as a governing equation for exposure (breathing zone concentration, \( C_{bz} \)). The volume is now the breathing zone volume \( (V_{bz}) \), which is assumed to be well-mixed, the fraction of the source generation rate that enters this volume is \( f \), and the total flow that passes through the breathing zone is \( Q_{bz} \). Equation (4) can now be rewritten as:

\[
\frac{dc}{dt} = \frac{Q_{bz}(fG - C_{bz})}{V_{bz}}
\]

(6)

**Stochastic differential equation for exposure**

Equation (6) is a deterministic differential equation in the sense that all variables are considered known explicitly and a single unique solution is obtained with appropriate initial conditions. More recently an appreciation for the randomness of physical processes has led to rigorous mathematical formulations that allow for consideration of random effects within the framework of stochastic differential equations (SDEs). Solutions are now in terms of
Cobb and Thrall (1981). Following their treatment let tractable introduction with applications is given in the Ito theory of the SDE is complex, but a very probability distributions rather than point estimates. The Ito theory of the SDE is complex, but a very tractable introduction with applications is given in Cobb and Thrall (1981). Following their treatment let

\[ \xi(C_{bz}) = \frac{qG}{V_{bz}} (fG - C_{bz}) \]  

(7)

where \( \xi(C) \) is the expected rate of change of \( C \).

And

\[ \tau^2(C_{bz}) = \epsilon C_{bz}(1 - C_{bz}) \]  

(8)

where \( \tau^2(C) \) is the ‘variance’ in the rate of change of \( C \) and \( \epsilon \) is constant.

Formal mathematical definitions for \( \tau^2 \) and \( \xi \) can be found in Cobb and Thrall (1981) and are included in the Appendix. An Ito equation for concentration is:

\[ dC_{bz} = \frac{Q_{bz}}{V_{bz}} (fG - C_{bz}) dt + \sqrt{\epsilon C_{bz}(1 - C_{bz})} dw \]  

(9)

where \( w \) represents a mathematical idealization of Brownian motion known as a Wiener process. The stationary solution for a stochastic differential equation in the form of equation (9) is the beta probability density function, \( \Phi \) (Cobb and Thrall, 1981):

\[ \Phi(C_{bz}) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} C^{p-1}(1-C)^{q-1} \]  

(10)

where:

\[ p = \frac{fG}{Q_{bz}} \delta \]  

(11)

\[ q = \frac{Q_{bz} - fG}{Q_{bz} \delta} \]  

(12)

and

\[ \delta = \frac{1}{p+q} = \frac{\epsilon V_{bz}}{Q_{bz}} \]  

(13)

The mean and variance for concentration (exposure) from equation (10) are:

\[ \mu = \frac{p}{p+q} = \frac{fG}{Q_{bz}} \]  

\[ \sigma^2 = \frac{pq}{(p+q)^2(p+q+1)} = \frac{\delta}{1+\delta}[\mu(1-\mu)] \]  

(14)

Thus the stationary probability distribution for exposure that arises from a basic mass balance, and the assumption that variability in the rate of change of concentration is proportional to the product of concentration and one minus concentration, is the standard beta distribution. In what follows this assumption is examined with real exposure data, and the implications for modeling and control are discussed.

**METHODS**

The primary objective of this research is to examine the hypothesis that equations (9) and (10) define a fundamental model of human exposure to airborne contaminants. The hypothesis is supported if exposure data fit the beta distribution. If the analysis indicates this is true, then it would be useful to explore how to estimate the distribution parameters \( a \) priori, if possible, or by collecting field data in conjunction with exposure to model them empirically. As noted above, the beta distribution has been used successfully as a model to describe exposure to naphthalene vapors (Flynn, 2004). The approach is extended here to include both solid and vapor phases by exploring its application to several previously published exposure data sets.

Five different sets of personal sampling data are explored. The first two (IPA-W1 and IPA-W2) are repeat 8 h TWA measurements of isopropyl alcohol exposure on two individuals at an automobile assembly facility, described previously in Flynn and George (1996). Both individuals conducted a pre-prime wipe down operation in the same ventilated booth. The third data set consists of 8 h TWA samples for benzene exposure taken at an Italian petrochemical facility (Tolentino et al., 2003). The fourth set is comprised of personal exposure monitoring for total mass sampled during compressed air spray painting operations at a US Air Force base (Tann and Flynn, 2002). The data were collected using filters backed up with charcoal tubes and the total mass, both solid and vapor phase, was reported. Sample duration was of the order of 15 min. The final data set consists of 8 h TWA personal exposure measurements to respirable silica collected on bricklayers during masonry tuck-pointing operations (NIOSH, 1999). A statistical summary of the raw data for each of these five sets is presented in Table 1.

In each case the measured exposures were transformed into dimensionless concentrations by dividing by an appropriate normalizing (maximum possible) value. This normalization was used since the predicted beta distributions are all in standard form. For the isopropyl alcohol data, concentration was expressed as a decimal fraction by dividing each exposure measurement by the maximum possible concentration (43 421 p.p.m.) based on the saturation vapor pressure of IPA. The same approach was used with the benzene data with a saturation concentration of 98 684 p.p.m. For the spray painting studies, the concentrations were reported in milligrams per cubic meter for the total mass of contaminants (both solids and vapor). These results were transformed into...
dimensionless volume fractions using equation (2) with a paint density of 1.1325 g/cm$^3$. The respirable silica samples employed a similar transformation using the density of silica 2.6 g/cm$^3$ as the normalizing factor. The transformed data were then fit to a standard beta distribution using the method of maximum likelihood to estimate values for p and q (see Appendix for details). Once these estimates were obtained the fitted distribution was plotted against the distributions obtained from the raw data and a one-sample Kolmogorov–Smirnoff goodness-of-fit test was performed to evaluate the hypothesis that the observed distribution is a plausible sample from the fitted beta distribution.

RESULTS

The results of the parameter estimation are given in Table 2 and the graphical comparisons between the fitted distributions and the data are shown in Figs 1–5. The beta distributions are in good agreement with the data as the figures show and the Kolmogorov–Smirnov goodness-of-fit tests confirm. The beta distribution is completely defined by the parameters, p and q. Once these have been estimated the values for δ, the mean, and the standard deviation are calculated directly with equations (13) and (14). These are the results reported in Table 2. The mean and standard deviation are recovered in p.p.m. or mg/m$^3$ by multiplying by the normalizing concentration (or density), which is reported in the last line of Table 2.

The $D$ statistic reported in Table 2 for the Kolmogorov–Smirnov goodness-of-fit test is the maximum absolute difference between the theoretical (fitted) beta distribution and the empirical cumulative distribution. The one sided $P$-value is the probability of exceeding $D$ under the null hypothesis of equality. The test results indicate that for all the data sets, the observed cumulative distributions are plausible samples from a standard beta distribution. Figures 1–5 confirm this in a qualitative sense. The estimated values for the means and standard deviations using the fitted beta distribution are in very good agreement.

Table 1. Summary of raw data

<table>
<thead>
<tr>
<th>IPA-W1 (p.p.m.)</th>
<th>IPA-W2 (p.p.m.)</th>
<th>Benzene (p.p.m.)</th>
<th>Spray paint (mg/m$^3$)</th>
<th>Silica (mg/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>105</td>
<td>64</td>
<td>33</td>
<td>20</td>
</tr>
<tr>
<td>Mean</td>
<td>65.32</td>
<td>19.18</td>
<td>0.28</td>
<td>110.71</td>
</tr>
<tr>
<td>Median</td>
<td>65.65</td>
<td>17.50</td>
<td>0.186</td>
<td>101.30</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>21.56</td>
<td>7.49</td>
<td>0.312</td>
<td>71.88</td>
</tr>
<tr>
<td>Minimum</td>
<td>13.89</td>
<td>9.69</td>
<td>0.01</td>
<td>13.07</td>
</tr>
<tr>
<td>Maximum</td>
<td>113.37</td>
<td>57.04</td>
<td>1.4</td>
<td>274.66</td>
</tr>
</tbody>
</table>

Table 2. Parameters and test results from fitting a beta distribution to the raw data

<table>
<thead>
<tr>
<th>IPA-W1</th>
<th>IPA-W2</th>
<th>Benzene</th>
<th>Spray paint</th>
<th>Silica</th>
</tr>
</thead>
<tbody>
<tr>
<td>P (MLE)</td>
<td>7.49</td>
<td>8.726</td>
<td>0.96877</td>
<td>2.35138</td>
</tr>
<tr>
<td>Q (MLE)</td>
<td>4971</td>
<td>19.8275</td>
<td>344 453.14</td>
<td>24 264 149.5</td>
</tr>
<tr>
<td>δ</td>
<td>0.0002</td>
<td>0.00005</td>
<td>0.0000029</td>
<td>0.000000041</td>
</tr>
<tr>
<td>Mean</td>
<td>65.3 p.p.m.</td>
<td>19.1 p.p.m.</td>
<td>0.28 p.p.m.</td>
<td>109.7 mg/m$^3$</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.000549 or</td>
<td>0.00044 or</td>
<td>0.0000028 or</td>
<td>9.7 × 10$^{-8}$ or</td>
</tr>
<tr>
<td>KS $D$ statistic</td>
<td>0.10</td>
<td>0.09</td>
<td>0.07</td>
<td>0.11</td>
</tr>
<tr>
<td>KS one-sided p</td>
<td>0.12</td>
<td>0.32</td>
<td>0.68</td>
<td>0.58</td>
</tr>
<tr>
<td>C std</td>
<td>43 421 p.p.m.</td>
<td>43 421 p.p.m.</td>
<td>98 684 p.p.m.</td>
<td>1.1325 × 10$^9$ mg/m$^3$</td>
</tr>
</tbody>
</table>

Fig. 1. Cumulative distributions for exposure to isopropanol: worker 1.

Fig. 2. Cumulative distributions for exposure to isopropanol: worker 2.
Stochastic differential equation for exposure

with the sample values. The maximum deviation in the means is observed with the silica data where the estimated mean is \( \sim 4\% \) greater than the sample mean. The maximum difference between the standard deviations is observed with the isopropanol data on worker 2, where the sample standard deviation is \( \sim 16\% \) higher than the value obtained from the fitted beta distribution.

**DISCUSSION**

The exposure data examined here fit the beta distribution quite well. Both solid and vapor phases are included, full-shift and shorter-term sampling times are explored, and exposures for individuals and also for job classifications are included. The model defined by equations (9) and (10) is supported by these results, although interpretation of the terms depends somewhat upon whether one is looking at a group of workers or an individual. The parameter \( \delta \) is estimated using equation (13) and includes the constant \( \varepsilon \), which governs the variability. Subsequent estimation of this parameter depends upon knowledge of the breathing zone volume and the airflow into it. The airflow into the breathing zone can be estimated empirically by releasing tracer at a known flow rate and measuring the equilibrium concentration in the breathing zone (\( C_e \)):

\[
Q_{bz} = \frac{G_{tracer}}{C_e} \tag{15}
\]

In addition, if the time history of the tracer concentration in the breathing zone is also recorded then the value of \( Q_{bz}/V_{bz} \) can be obtained by using an equation of the form:

\[
C_{bz}(t) = \frac{G_{tracer}}{Q_{bz}} \left[ 1 - \exp \left( -\frac{Q_{bz}t}{V_{bz}} \right) \right] \tag{16}
\]

A value for \( \varepsilon \) can then be calculated directly from \( \delta \) with equation (13). With the breathing zone flow rate known and an estimate of the contaminant generation rate \( G \) in the space, \( f \) is estimated from the mean exposure. Thus field measurements of contaminant generation rates and time histories of tracer concentrations released in the breathing zone at a known flow rate provide supplementary data to complete the SDE model of exposure. This should provide an effective way to estimate future exposure distributions for similar processes, *a priori*, or equivalently a useful tool for retrospective exposure assessment.

For the isopropanol data sets, additional information was collected during the sampling, which permitted estimates of the generation rate at \( \sim 0.03 \) cfm of IPA vapor for each worker. It might further be hypothesized that the value of \( Q_{bz}/V_{bz} \) would not vary dramatically between the two workers since they worked in the same ventilated space, doing essentially the same job. However, fairly significant differences in mean exposure, 65.3 versus 19.2 p.p.m., and fitted values for delta, 0.0002 and 0.00005, were observed. According to the model proposed here the differences in the mean exposure would indicate a difference in \( f \) the fraction of contaminant generation rate entering the breathing zone if the breathing zone flows were similar. This is supported by observations in the original paper (Flynn and George, 1996) indicating that worker 2 used work practices that essentially assured a smaller value for \( f \) (more time outside the booth, and using cloths that had evaporated at a remote location). If the values for \( Q_{bz}/V_{bz} \) are similar for the two workers, then \( \varepsilon \) may not be a constant for a fixed job, further research is needed to explore this dependence.

The model proposed here assumes that the ‘variability’ in the rate of change of concentration is proportional to the product of concentration and one minus concentration, i.e. equation (8). Other possibilities exist to close the stochastic differential equation including that the variability is proportional to concentration alone. If this closure is invoked the probability distribution that results is a gamma distri-
bution, which is skewed with a long infinite tail to the right, much like the two-parameter log-normal model. The log-normal is often invoked as an exposure model because the variability in concentration is observed to be proportional to the mean. However, as concentration approaches either extreme, the variability in the rate of change ultimately must be constrained, and thus a strict proportionality between the variance and the mean is physically implausible. This in conjunction with the definition of exposure leads to the closure selected here; further analysis is included in the Appendix.

CONCLUSIONS

The results presented here suggest that the beta distribution may provide a fundamental and useful model of human exposure to airborne contaminants as a solution to a stochastic differential equation governing concentration. The use of a physically consistent definition for concentration as a dimensionless volume fraction facilitates the analyses for particles, gases and vapors. The connection between the probability distribution function describing exposure and a fundamental mass balance differential equation provides some basis for extrapolation to other exposure scenarios, although additional field data are required. These include estimates for contaminant generation rates, airflow, and time histories of tracer concentrations in the breathing zone.

APPENDIX

Maximum likelihood estimates for the beta distribution

Estimation of the parameters \( p \) and \( q \) for the standard beta distribution described in equation (12) can be accomplished conveniently by the method of moments as:

\[
\hat{p} = \frac{(1 - m) - c^2 m^2}{c^2} \quad \hat{q} = \frac{\hat{p} - \hat{p}}{m - \hat{p}} \quad (A1)
\]

where \( c^2 = s^2/m^2 \), where \( m \) and \( s \) are the sample mean and standard deviation, respectively. These estimates, while convenient, are not of maximum likelihood. Maximum likelihood estimates (MLE) are given (Johnson and Kotz, 1970) as solutions to the following simultaneous equations:

\[
\Psi(\hat{p}) - \Psi(\hat{p} + \hat{q}) = n^{-1} \sum_{j=1}^{n} \ln(Y_j) \\
\Psi(\hat{p}) - \Psi(\hat{p} + \hat{q}) = n^{-1} \sum_{j=1}^{n} \ln(1 - Y_j) \quad (A2)
\]

where \( Y_j \) is the transformed concentration data, and \( \Psi \) is the digamma or psi function:

\[
\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}
\]

where \( \Gamma(x) \) and \( \Gamma'(x) \) are the usual gamma function, and its derivative, respectively. Equations (A2) are somewhat cumbersome to solve, but an iterative process with an initial guess provided by equations (A1) was implemented here to obtain the MLE estimates for \( p \) and \( q \). A short FORTRAN code was written using the IMSL library for the psi function (DPSI), and an IMSL subroutine DNEQNJ for solving non-linear, coupled equations. This particular subroutine is a modified routine subroutine for an analytic Jacobian.

Closure justification

As noted in the Discussion, alternative closures for the ‘variance’ in the rate of change of concentration are possible. Variance is in quotes, since in reality it is a conditional expectation of a stochastic process, not a true variance. The precise definition for a stochastic process is (Cobb and Thrall, 1981):

\[
\chi^2(c) = \lim_{h \to 0} \frac{E\left[\frac{1}{2}\left[C(t + h) - C(t)\right]^2 \mid C(t) = c\right]}{h} \quad (A3)
\]

A similar definition for the expected rate of change is:

\[
\xi(c) = \lim_{h \to 0} \frac{E\left[\left[C(t + h) - C(t)\right] \mid C(t) = c\right]}{h} \quad (A4)
\]

The dilution ventilation equation states that the time rate of change in concentration is proportional to concentration. If the ‘variance’ in this rate of change due to random effects is proportional to a random change in concentration. Then given the definition of concentration we have:

\[
c = \frac{V_c}{V_c + V_a} \quad \text{and thus} \quad \frac{V_c}{V_a} = \frac{c}{1 - c} \quad (A5)
\]

If we assume:

\[
\frac{V_c}{V_a} = \alpha \exp[\varepsilon(z - \gamma)]
\]

where \( z \) is a random variable and \( \alpha, \gamma \) and \( \varepsilon \) are arbitrary constants, then differentiating \( c \) with respect to \( z \) gives the random change in concentration as:

\[
dc = \varepsilon c(1 - c)dz \quad (A6)
\]
Thus if the ratio of contaminant volume to pure air volume is an exponential function of a random variable, then the approximate change in concentration due to random effects will follow equation (A6). This is a rationalization for the closure selected, other formulations are possible. The key point is that the closure must produce a bounded probability distribution between 0 and the maximum possible concentration.

REFERENCES


