Relative congener scaling of polychlorinated dibenzo-*p*-dioxins and dibenzofurans to estimate building fire contributions in air, surface wipes, and dust samples

Joachim D. Pleil[1*] and Matthew N. Lorber[2]

[1] Human Exposure and Atmospheric Sciences Division, National Exposure Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711

[2] National Center for Environmental Assessment, Headquarters, U.S. Environmental Protection Agency Washington, DC 20460

Supporting Information for the paper in Environ. Sci. Technol.

1 August, 2007

Contents:

- Part 1: Imputation of missing values
- Part 2: Table describing data sets and TEQ's
- Part 3: Discussion of patterns of dioxin measurements from other studies

Supporting Information (Part 1).

Imputation of Missing data:

There are two basic scenarios for which data are imputed for subsequent calculations:

1. entries are missing entirely at random from a large data set, typically due to loss to follow up in medical studies, or laboratory error, contamination, lost sample, etc.
2. entries are missing because the amount of material measured is below a laboratory's detection limit; that is, the true value is between zero and some known threshold.

The first scenario above is referred to as "MAR" or "missing at random" and is defined to occur when the probability of a missing value does not depend on the magnitude of the value (Junninen et al. 2004). This is not the case for our work.

Often in trace environmental measurements, the need for imputation is created by the second scenario: some of the samples are simply below the laboratory detection limit. Such entries are listed as "below detection limit or "BDL" in the data tables. The BDL value could be further qualified due to interference with other co-eluting compounds (Lubin et al. 2004). In our case, the missing values occur due to BDL issues.

The WTC disaster was an unanticipated transient event, and therefore all available samples are unique and precious; they cannot be repeated or re-sampled. As such, we wished to use as many available samples as possible for our pattern analysis effort even if a few of the 17 congeners were below the analytical detection limit in any given sample. We did, however, require a non-zero entry for each congener to perform the subsequent calculations. Because all dioxin congeners are generally present in a fire source sample, substituting a zero for a non-detectable value is likely wrong; this is referred to left censored data where it is expected that there are indeed true non-zero values below a known threshold. Therefore, we wish to impute some reasonable value that lies between zero and the analytical limit of quantitation (LOQ). We note that the missing values (below LOQ) are not due to errors in the laboratory but are a result of the difficulties encountered in collecting enough sample volume under the hazardous conditions of the WTC rescue efforts.

As discussed in the main body of the paper, we constructed a basis set of 29 WTC samples that had at least <80% sample coverage of the 17 measured congeners (missing no more than 3 congener values). Of the 29 samples, 22 were complete, one was missing 3 values, three were missing 2 values and three were missing 1 value. This required imputing 12 of 493 (2.4%) values in the basis sample set. We used a single value imputation method for its simplicity and because it was mathematically reasonable for the underlying structure of the available data. Below is a discussion of the rationale for the choice made.

One of the simplest methods of assigning a value to left-censored entries is single value imputation wherein a value related to the LOQ is inserted into the data for each missing value. Hornung and Reed (1990) compared methods of such imputation for reconstructing means of data sets with missing values. Their "gold standard" for comparison was the Hald method wherein maximum likelihood estimates are directly calculated for the mean and standard deviation of the data set requiring a complex calculation involving all non-zero entries. The comparison methods invoked a single value imputation using either LOQ/2 or LOQ/√2 and made the empirical calculation from the completed data set. They found that the two simpler imputation methods ***"…are desirable and sufficiently accurate..."*** subject to the conditions that the distributions are log-normal and that less than 50% of the data are missing. Furthermore, they suggest that the latter (LOQ/√2) substitution is preferable if the ***"…frequency in the first or second interval is less than one or more than of the subsequent intervals…"*** in a histogram of the detectable data. Other schemes for single value imputation have been proposed for multiple laboratory scenarios where the value corresponding to the median percentile below each laboratory's detection limit is substituted or through use of predicted values of structural equation models gleaned from the non-censored data (Succop et al. 2004). This was not necessary for our work.

Although there other, more complex methods developed to deal with a variety of missing value scenarios, these are developed primarily for the MAR scenario as discussed by Junninen et al. (2004). For environmental (left-censored) data, the next level of complexity is multiple value imputation wherein randomly generated values (below the LOQ) are substituted for the BDL entries and the mathematical calculations (means, standard deviations, modeled parameters, etc.) are made repeatedly using different sets of imputed values according to Lubin et al. (2004). Typically, these methods are more appropriate when many values (as many as 30%, or so) are missing. Comparison among such trials can then demonstrate the impact of the imputed values and assign very good estimates for errors induced by the imputation process (Liu et al. 1997, Lynn 2000, Lubin et al. 2004).

For this project, we chose single value imputation with LOQ/√2 substitution as the most appropriate method. We found that 16 of the 17 congeners individually demonstrated log-normal behavior using Shapiro-Wilkes and all 17 passed Kolmogorov-Smirnov test with $p>0.10$. We further found that the histograms of individual congeners fulfilled the empirical parameters set forth by Hornung and Reed (1990). As such, the single value imputation method using LOQ/√2 is confirmed as appropriate by the available peer reviewed literature; more complex methods are unnecessary.

References (Part 1).

Junninen, H, Niska, H, Tuppurainian, K, Ruuskanen, J, Kolehmainen M. 2004. Methods for imputation of missing values in air quality data sets, *Atmos. Environ.* 38:2895-2907.

Lubin JH, JS Colt, D Camman, S Davis, JR Cerhan, RK Severson, L Bernstein, P Hartge. 2004. Epidemiological evaluation of measurement data in the presence of detection limits. *Env. Health Persp*. 112: 1691-1696.

Hornung, R.W.; Reed, L.D. Estimation of average concentration in the presence of nondetectable values. *Appl. Occup. Environ. Hyg.* **1990**, 5 (1), 46-51.

Succop, PA, Clark, S, Chen, M, Galke, W. 2004. Imputation of data values that are less than a detection limit. *J. Occup. Environ, Hyg.* 1:436-441.

Liu, S., Lu,JC, Kolpin, DW, Meeker, WQ. 1997. Analysis of environmental data with censored observations, *Environ. Sci. Technol*. 31:3358-3362.

Lynn, HS, 2001. Maximum likelihood inference for left censored HIV RNA data, *Statist. Med.* 20:33-45.

Supporting Information (Part 2).

Table 1. Description of data sets including average and range of TEQ concentrations

Sample Set	N	Reference; Description; Units	Average, Range of TEQ
I. Associated with WTC			
WTC air (basis)	29	Santella, 2002; Air concentrations from samplers on-site and just off-site Ground Zero, Sep 16 to Oct 31, 2001; pg/m^3	53, 0.3 – 170
WTC window film, impacted	6	Rayne, 2005; window film wipe samples from 6 buildings near Ground Zero, ng/m^2	1.22, 0.43 – 3.29
OSHA Personal air	8	Ferrario, 2002; samples from OSHA personal samplers from workers on pile in Sep, 2001, pg/m^3	13, 2 – 79
Deutsche Bank dust wipe	8	Santella, 2005; wipe samples from Deutsche Bank taken in 2003; highly impacted building is located at 4 Liberty St, bordering Ground Zero, ng/m^2	28, 13 - 52
Street bulk dust a (WTC)	5	Ferrario, 2002; dust samples taken at Ground Zero and nearby by EPA's ORD on 9/22, pg/g	114, 13 - 330
Street bulk dust b (WTC)	3	Lioy et al, 2002; dust samples taken on 9/16 & 9/17 from streets near Ground Zero, pg/g	101, 96-104
WTC window film, background	2	Rayne, 2005; window film samples taken at locations distant from Ground Zero, including Brooklyn and NYU, ng/m^2	0.024, 0.010, 0.037
II. Other Fire-Related Data			
Philadelphia soot	9	Kominsky & Freyberg, 1992; soot samples from building fire in Philadelphia, pg/g	40,419, 14,281 – 103,392
Philadelphia indoor wipes	8	Kominsky & Freyberg, 1992; wipe samples from impacted locations in Philadelphia building fire, ng/m^2	10, 3-20
Binghamton office air	1	NYSDOH, 1989; transformer fire in office building in Binghamton, NY; 1 air sample from in electrical room where fire occurred; pg/m^3	12.3,
Binghamton office wipes	11	NYSDOH, 1989; wipe samples from light fixtures, floor, stairs, windows, ng/m^2	9, 3 – 14
Columbus incinerator stack emissions	5	Schaum, 1994; stack emission measurements from incinerator in Columbus, OH, emitting large amounts of dioxin, ng/dry standard cubic meter	155, 63 - 228
Columbus impacted air downwind of incinerator	1	Lorber, et al., 1998; average profile from 2 air measurements in predominant downwind direction from incinerator, pg/m^3	0.28,
III. Background Data			
CARB urban air (basis)	29	CARB, 2005; a random selection of data from the Los Angeles and Oakland areas from the CARB air toxics monitoring program from years 2001 through 2002; pg/m^3	0.027, 0.009 – 0.066
NDAMN air	12	Cleverly, et al, 2001; Cleverly, 2005; Subset of the background NDAMN data; quarterly 2001 data for sites in MD, PA, and IL, pg/m^3	0.017, 0.008 – 0.040
Columbus background urban soil	1	Lorber, et al, 1998; Average profile from 12 samples taken in urban background locations in Columbus, OH, pg/g	4.6,

Columbus background urban air	1	Lorber, et al., 1998; average profile from 6 air measurements upwind of incinerator representing urban background in Columbus, OH, pg/m^3	0.05
CARB (San Jose) air	20	CARB, 2005: data from 2001 and 2002 from the CARB air toxics monitoring program; pg/m^3	0.026, 0.008 – 0.069
CARB (Riverside) air	21		0.027, 0.007 – 0.16
CARB (Sacramento) air	12		0.026, 0.007 – 0.065
CARB (Livermore) air	25		0.034, 0.006 – 0.19

References.

Santella, D. 2002. Personal communication from D. Santella, EPA Region 2, New York City, to M. Lorber, Office of Research and Development, US EPA, Washington, DC.

Rayne S, MG Ikonomou, CM Butt, ML Diamond, J Truong. 2005. Polychlorinated dioxins and furans from the World Trade Center Attacks in Exterior Window films from Lower Manhattan in New York City. Env. Sci Tech. 39: 1995-2003.

Ferrario, JB. 2002. Personal communication from J. Ferrario, Office of Pesticide Programs, US EPA, to M. Lorber, Office of Research and Development, US EPA. Unpublished data from personal samplers worn by OSHA workers in the September/October 2001 time frame, and from dust samples taken by US EPA ORD.

Santella, D. 2005. Personal communication from D. Santella, EPA Region 2, New York City, to M. Lorber, Office of Research and Development, US EPA, Washington, DC. Reports provided by counsel for Deutsche Bank on results of wipe sampling for dioxins and other contaminants that were taken in the Deutsche Bank Building, New York City.

Lioy PJ, CP Weisel, JR Millette, S Eisenreich, DVallero, et al. 2002. Characterization of the dust/smoke aerosol that settled east of the World Trade Center (WTC) in lower Manhattan after the collapse of the WTC 11 September 2001. Env. Health Persp. 110: 703-714.

Kominsky JR, RW Freyberg. 1992. Environmental Characteristics of PCB, PCDD, and PCDF Contamination in a 38-Story Office Building in Philadelphia, PA. Report prepared by Environmental Quality Management, Cincinnati, OH, for E/R Associates, Philadelphia, PA.

NYSDOH. 1989. Comparison of PCBs and PCDD/PCDFs in the Air and on Surfaces of the Binghamtom State Office Building and Utica State Office Building. Center for Environmental Health and Wadsworth Center for Laboratories and Research. New York State Department of Health, Albany, New York.

Schaum J. 1994. Personal communication from John Koval, Air Toxics Unit, Ohio EPA, to John Schaum, Office of Research and Development, US EPA, describing Ohio EPA's emission testing, modeling, and assessment of dioxin emissions from the Columbus Solid Waste Incinerator, Columbus, OH.

Lorber M, P Pinsky, P Gehring, C Braverman, D Winters, W Sovocool. 1998. Relationships between dioxins in soil, air, ash, and emissions from a municipal solid waste incinerator emitting large amounts of dioxins. Chemosphere, Volume 37:2173-2197.

CARB. 2005. California Air Resources Board. Data on dioxin air monitoring results from the CARB Air Toxics Program. Data publicly available and accessible at, http://www.arb.ca.gov/toxics/toxics.htm.

Cleverly D, D Winters, J Ferrario, et al. 2001. The National Dioxin Air Monitoring Network (NDAMN): Measurement of CDDs, CDFs and Coplanar PCBs at 15 Rural and 6 National Park Areas of the US June 1998 - December 1999. Organohalogen Compounds 51:1-4.

Cleverly D, 2005. Personal communication. Congener-specific air monitoring data from 2001 from three NDAMN sites in PA, MD, and IL; 4 data points per site corresponding to four annual quarters (samples taken in March, May, Aug, and Nov).

Supporting Information (Part 3).

Discussion of the model's relationships to other patterns reported in the literature.

Upon closer evaluation of individual congeners in the data sets from other fire sources listed in Table 1, the fire-related profiles are very different from each other. This is most apparent when more closely examining the Binghamton Office fire data set. In this profile, the dioxin congeners are virtually absent and the lower chlorinated furan congeners dominate the profile. OCDD has a D7 value of 0.05, and F1, the fraction of 2,3,7,8-TCDF in the profile, is the highest of the furan congeners at 0.17. It is this very high fraction of 2,3,7,8-TCDF, F1, that causes such a strong signal for model 2 (OPT = 1.62). In fact, the first four furan congeners of the Binghamton profile account for 0.63 (63%) of the entire profile. This is clearly distinct from other combustion profiles, where the first four furan congeners contribute no more than 0.40 of the profile and 2,3,7,8-TCDF in the range of 0.02 – 0.10.

The Binghamton office building fire was the result of a transformer explosion, and the burning of PCBs has been attributed to the resulting high concentrations of CDD/CDFs found and the CDD/CDF profile (NYSDOH 1989). However, another consideration for the anomalous 2378-TCDF finding could be analytical chemistry. Rayne et al. (2005) discuss the importance of a second column confirmation for this congener, as they identify 5 additional non-dioxin-like furan congeners which co-elute with 2,3,7,8-TCDF. Therefore, a single column analysis could overestimate the presence of 2,3,7,8-TCDF. This same trend was identified by Lorber (2005). In a preliminary and qualitative evaluation of WTC trends, Lorber noted that a limited second column confirmation of WTC air samples for 2,3,7,8-TCDF resulted in a reduction of the first column concentration by 85%. While Lorber corrected the WTC air profiles in the analysis he conducted (a precursor to the evaluations in this paper), no such data manipulations were attempted with the diverse sets of CDD/CDF data collected for this effort. Even without this large reported 2,3,7,8-TCDF value in the Binghamton office fire data set, the prevalence of other lower chlorinated furans in that profile suggest that it is distinct from the WTC fire profile. Examination of other combustion source profiles would likely unearth similar differences. There have been other CDD/CDF profiles reported in the literature that have benefited from the fact that concentrations of the CDD/CDF congeners were very high in the sampled matrices, including stack gas, animal feeds, a geologic clay, and even food products. They also had the advantage of focusing on an unambiguous source of the dioxins and furans. Neither characteristic (high levels, or unambiguous source) is generally available when trying to understand and interpret background ambient air concentrations of CDD/CDFs.

The use of dioxin congener profiles to "fingerprint" sources is fairly common in the dioxin literature, although most efforts rely on qualitative observations and graphical methods, rather than quantitative, predictive methodologies, as in this paper. Ferrario et al (2000) identify a very unique pattern of dioxin-like compounds in ball clay – they found extremely elevated levels of dioxin congeners but a virtual absence of furan congeners. Their study of ball clay was prompted by an earlier anomalous finding of elevated 2378-TCDD in an EPA survey of dioxin-like compounds in poultry (Ferrario, et al, 1997), which eventually was traced to the presence of ball clay as an anti-caking agent in the poultry feed. Van Larebeke et al (2001) document a similar food contamination episode in Belgium, where poultry, pork, and other food products were found to have elevated PCBs and PCDD/Fs based on contamination of animal feed by waste oils containing PCBs. In that case, they identify a clear predominance of furan congeners over dioxin congeners, which they conclude is compatible with dioxin contamination of PCBs in contrast to the pattern from thermal combustion, which is dominated by OCDD. Buekens et al (2000) conducted principal component analysis on stack emission profiles of dioxins from 6 combustion sources including smelters, furnaces, and municipal solid waste incinerators. Among other conclusions, they found that while there were important similarities among the combustion sources, they were interesting differences, such as the fact that 2378-TCDF dominated emissions from a lead smelter. Congener profiles of 18 different combustion sources were provided in Cleverly, et al (1997). Among other observations that are drawn in this compilation, Cleverly notes that while OCDD is the dominant congener in several sources including municipal solid waste incinerators, various furan congeners dominate other sources such as hazardous waste incinerators, secondary lead smelters, and other sources.   All of these efforts and others benefited from the fact that concentrations of the dioxin and furan congeners were very high in the sampled matrices, including stack gas, animal feeds, a geologic clay, and even food products. They also had the advantage of focusing on an unambiguous source of the dioxins and furans.

As noted in the introduction of the main paper, urban air dioxins are likely the result of numerous disperse sources. Also, levels are generally low, background data tend to be sparse, and ambient air could be dominated by an individual compound. The very high levels of dioxins in the air near the unambiguous WTC fire source, coupled with the availability of a robust dataset of background dioxin concentrations from the California Air Resources Board, allowed for the development of statistically sound multivariate regression models. Further, we found that a form of cluster analysis can reduce independent variables (i.e., individual congeners) into groups based on their covariance structure.   Based on such practical (but not necessarily elegant) mathematical procedures, we found that we could distinguish between fire dominated and urban source dominated dioxin congener patterns with a simple model that focuses primarily on one dominant congener. Upon removing the dominant compound and calculating a second model, we found an additional potential advantage for distinguishing among various nominal background data sets albeit at the expense of some additional complexity and uncertainty over the original model.   As such, our work advances the state of the knowledge by accommodating more subtle changes in patterns, the occasional below detection limit congener, and the scarcity of environmental measurements.

References (Part 3):

NYSDOH. 1989. Comparison of PCBs and PCDD/PCDFs in the Air and on Surfaces of the Binghamtom State Office Building and Utica State Office Building. Center for Environmental Health and Wadsworth Center for Laboratories and Research. New York State Department of Health, Albany, New York.

Rayne S, MG Ikonomou, CM Butt, ML Diamond, J Truong. 2005. Polychlorinated dioxins and furans from the World Trade Center Attacks in Exterior Window films from Lower Manhattan in New York City. Env. Sci Tech. 39: 1995-2003.
Lorber

Schaum J. 1994. Personal communication from John Koval, Air Toxics Unit, Ohio EPA, to John Schaum, Office of Research and Development, US EPA, describing Ohio EPA's emission testing, modeling, and assessment of dioxin emissions from the Columbus Solid Waste Incinerator, Columbus, OH.

Ferrario, JB. 2002. Personal communication from J. Ferrario, Office of Pesticide Programs, US EPA, to M. Lorber, Office of Research and Development, US EPA. Unpublished data from personal samplers worn by OSHA workers in the September/October 2001 time frame, and from dust samples taken by US EPA ORD.

Ferrario JB, CJ Byrne, and DH Cleverly. 2000. 2,3,7,8-Dibenzo-p-dioxins in mined clay products from the United States: evidence for possible natural origin. Env. Sci. and Tech 24: 4524-4532.

Van Larebeke N, L Hens, P Schepens, et al. 2001. The Belgian PCB and dioxin incident of January-June 1999: exposure data and potential impact on health. Env. Health Persp. 109: 265-272.

Buekens A, E Cornelis, H Huang, T Dewettinck. 2000. Fingerprints of dioxin from thermal industrial processes. Chemosphere 40: 1021-1024.

Cleverly D, J Schaum, G Schweer, J Becker, D Winters. 1997. The congener profiles of anthropogenic sources of chlorinated dibenzo-p-dioxins and chlorinated dibenzofurans in the United States. Organohalogen Compounds 32: 430-435.