Journal of Exposure Science and Environmental Epidemiology (2009) 19, 149–171
© 2009 Nature Publishing Group All rights reserved 1559-0631/09/$32.00
www.nature.com/jes

Reconstructing population exposures to environmental chemicals from biomarkers: Challenges and opportunities

PANOS G. GEORGOPOULOS[a], ALAN F. SASSO[a], SASTRY S. ISUKAPALLI[a], PAUL J. LIOY[a], DANIEL A. VALLERO[b], MILES OKINO[c] AND LARRY REITER[b]

[a]*Environmental and Occupational Health Sciences Institute (EOHSI), A Joint Institute of UMDNJ-RW Johnson Medical School & Rutgers University, Piscataway, New Jersey, USA*
[b]*National Exposure Research Laboratory, USEPA, Research Triangle Park, North Carolina, USA*
[c]*National Exposure Research Laboratory, USEPA, Las Vegas, Nevada, USA*

A conceptual/computational framework for exposure reconstruction from biomarker data combined with auxiliary exposure-related data is presented, evaluated with example applications, and examined in the context of future needs and opportunities. This framework employs physiologically based toxicokinetic (PBTK) modeling in conjunction with numerical "inversion" techniques. To quantify the value of different types of exposure data "accompanying" biomarker data, a study was conducted focusing on reconstructing exposures to chlorpyrifos, from measurements of its metabolite levels in urine. The study employed biomarker data as well as supporting exposure-related information from the National Human Exposure Assessment Survey (NHEXAS), Maryland, while the MENTOR-3P system (Modeling ENvironment for TOtal Risk with Physiologically based Pharmacokinetic modeling for Populations) was used for PBTK modeling. Recently proposed, simple numerical reconstruction methods were applied in this study, in conjunction with PBTK models. Two types of reconstructions were studied using (a) just the available biomarker and supporting exposure data and (b) synthetic data developed via augmenting available observations. Reconstruction using only available data resulted in a wide range of variation in estimated exposures. Reconstruction using synthetic data facilitated evaluation of numerical inversion methods and characterization of the value of additional information, such as study-specific data that can be collected in conjunction with the biomarker data. Although the NHEXAS data set provides a significant amount of supporting exposure-related information, especially when compared to national studies such as the National Health and Nutrition Examination Survey (NHANES), this information is still not adequate for detailed reconstruction of exposures under several conditions, as demonstrated here. The analysis presented here provides a starting point for introducing improved designs for future biomonitoring studies, from the perspective of exposure reconstruction; identifies specific limitations in existing exposure reconstruction methods that can be applied to population biomarker data; and suggests potential approaches for addressing exposure reconstruction from such data.
Journal of Exposure Science and Environmental Epidemiology (2009) **19,** 149–171; doi:10.1038/jes.2008.9; published online 26 March 2008

Keywords: *biomarker interpretation, biomonitoring, exposure biology, exposure reconstruction, inversion, PBTK.*

Introduction

It has been recently acknowledged that the "ability to generate new biomonitoring data often exceeds the ability to evaluate whether and how a chemical measured in an individual or population may cause a health risk or to evaluate its sources and pathways for exposure" (NRC, 2006). It is in fact now widely recognized that these biomonitoring data can be used not only as early indicators of a biological effect for assessing health risks but that, under certain circumstances, they may also be used to identify contributors to exposures, thus allowing for rational health risk management planning. However, the use of biomonitoring data to date has been limited to assessing the effectiveness of pollution controls for relatively straightforward exposure scenarios, such as those involving inert and persistent chemicals with relatively long biological half-lives and well-defined sources and pathways of exposure. For example, a well-known successful use of biomonitoring data has been in relation to lead exposures (Pirkle et al., 1995; Needham et al., 2007), which are easily attributable to relatively few source–route combinations (e.g., ingestion and inhalation.) For complex exposure scenarios, involving multiple routes of entry into the body the use of biomonitoring data in designing and evaluating exposure reduction strategies may require significant amounts of supporting or complementary exposure information (e.g. variability in source activities and in background concentrations, and multimedia dynamics of the chemicals). Table 1 lists examples of available population biomarker databases and the extent of supporting exposure-related information they provide in each case, from the

1. Address all correspondence to: Professor Panos G. Georgopoulos, EOHSI, 170 Frelinghuysen Road, Piscataway, NJ 08854, USA.
Tel.: +1 732 445 0519. Fax: +1 732 445 0915.
E-mail: panosg@ccl.rutgers.edu
Received 2 November 2007; accepted 22 January 2008; published online 26 March 2008

Table 1. Biomarker databases for population exposure reconstruction. Databases that include complementary exposure data can be used to evaluate exposure reconstruction approaches.

	Organophosphate Pesticides			Pyrethroid Pesticides			Metals					
Program/study	Chlorpyrifos	Diazinon	Malathion	Permethrins	Cyfluthrin	Cypermethrin	As	Cd	Cr	Hg/MeHg	Pb	Location; number of subjects
CHAMACOS (1999–2000) (Castorina et al., 2003)	bd	bd	bd									CA; 600 pregnant women
CTEPP (2000–2001) (Wilson et al., 2004)	ac	ac		ad	ad							NC, OH; 257 children (1.5–5 yrs)
MNCPES (1997) (Quackenboss et al., 2000)	ac	ac	ac									MN; 102 children (3–12 years)
NHANES-III (1988–94) (Hill et al., 1995)	c							c			bc	US; 1000 adults (20–59 yrs)
NHANES (1999–2000) (CDC, 2005)	cd	cd	cd					c		c	bc	US; 9,282 subjects (all ages)
NHANES 2001–2002 (CDC, 2005)	cd	cd	cd	c	cd	c		c		c	bc	US; 10,477 subjects (all ages)
NHANES 2003–2004	cd	cd	cd	c	cd	c	c	c		c	bc	US; 9,643 subjects (all ages)
NHEXAS-AZ (1995–97) (Robertson et al., 1999)	ac	ac	ac				ac	ac	ac		ac	AZ; 179 subjects (all ages)
NHEXAS-MD (1995–96)	ac		ac				ac	ac	ac		ac	MD; 80 subjects (above 10 yrs)
NHEXAS-V (1995–1997) (Whitmore et al., 1999)							ac	ac	ac	c	ac	EPA Region V; 251 subj. (all ages)

a: measurements of multimedia concentrations (indoor, outdoor, and personal air; drinking water; duplicate diet; dust; and soil);
b: partial measurements of environmental concentrations (e.g. outdoor air concentrations, pesticide use, etc.);
c: specific metabolites;
d: nonspecific metabolites.

perspective of interpreting biomonitoring data. It is clear that the available complementary or supporting exposure information is quite variable. Thus, source or contact inferences would be difficult in most cases. Selected data sets from the National Human Exposure Assessment Survey (NHEXAS) and National Health and Nutrition Examination Survey (NHANES) studies are in fact examined in demonstration applications within the present analysis.

Assessing biological doses and their effects using exposure measurements constitutes a "forward" mode of analysis, while estimating or reconstructing exposures from biomarkers requires an "inverse" mode of analysis. The forward analysis can be accomplished through the direct application of environmental (microenvironmental) exposure, toxicokinetic and toxicodynamic models, either empirical or mechanistic (i.e. physically and biologically based), whereas the reconstruction analysis requires application of numerical model inversion techniques applied in conjunction with the toxicokinetic and/or toxicodynamic models, and with available complementary data for interpretation. The magnitude of error and uncertainty introduced by simplifying assumptions in the "forward" modeling process can be substantial, in relation to the objectives of particular applications (see e.g. Isukapalli et al., 2008).

A computational framework that can address both the forward and inverse modeling aspects of exposures to multiple chemicals is essential for enhanced interpretation of biomarkers and their sources and routes of exposure. Physiologically based toxicokinetic (PBTK) and biologically based dose–response (BBDR) models in conjunction with numerical inversion techniques and optimization methods should form major components of such a framework, as shown in Figure 1. The figure shows how available supporting or complementary exposure data can be used to develop prior estimates of exposures for individuals and populations. These estimates can then be improved by using PBTK modeling and inversion techniques along with corresponding biomarker data. The approach presented in Figure 1 involves straightforward application of inversion via brute-force sampling, with significant computational requirements and consequent needs for performance improvements, as discussed in the following sections. A prerequisite for implementing such a computational framework is a systematic evaluation of available methods and computational tools that can be used to "merge" existing forward models and biomarker data for exposure reconstruction. This will allow for the identification and recommendation of appropriate methods and development needs for different types of biomarker and exposure timescales and chemical characteristics. This study examines these tools and provides demonstration applications. Eventually, a comprehensive exposure reconstruction framework should address reconstructions involving aggregate (i.e. from multiple exposure routes) and cumulative (i.e. for multiple chemicals) exposures, and provide user-friendly computational tools for use by the exposure/risk assessment and management communities.

Background

Various computational techniques have been employed in the literature for numerical model inversion in general and for exposure reconstruction from biomarkers in particular. Georgopoulos et al. (1994) used the maximum likelihood estimation (MLE) method in conjunction with PBTK modeling for reconstructing short-term (30 min) exposure

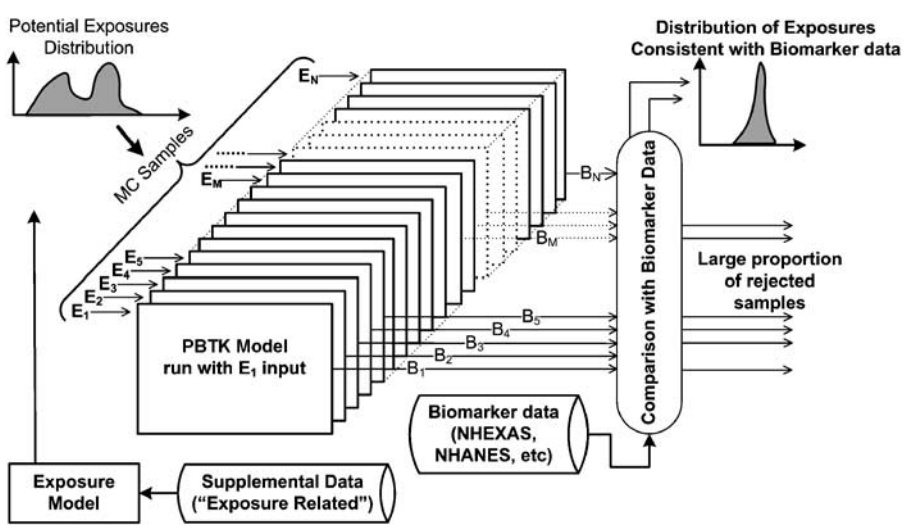

Figure 1. A simplified schematic of a computational framework for exposure reconstruction showing major components and processes. Available exposure-related data can provide "prior estimates" of exposures, which in turn can be used in conjunction with biomarkers and PBTK modeling to obtain improved estimates of exposures and doses.

to chloroform, and to resolve the total dose between two routes of uptake (i.e. inhalation and dermal absorption). Furthermore, Roy and Georgopoulos (1998) used the combined MLE–PBTK modeling approach with synthetic biomarker data and demonstrated that it is mathematically feasible to reconstruct longer term exposures to VOCs. The synthetic data were comprised of outputs from a PBTK model with known parameters and were used to represent exhaled breath concentration data while additional random "noise" was added to represent measurement error, with the rationale being that it is easier to evaluate the inversion method with synthetic data, because all the exposure and PBTK model parameters are already known. Rigas et al. (2001) used urinary biomarker data and the inverse solution of a simple, two-compartment toxicokinetic (PK) model for chlorpyrifos (CPF) to estimate the magnitude and timing of doses, based on the Minnesota Children's Pesticide Exposure Study (MNCPES) (Quackenboss et al., 2000; see Table 1). Some recent studies on population-level exposure reconstruction focused on data sampled from distributions of biomonitoring studies such as NHANES using direct deconvolution of biomarker distributions assuming a linear response (Tan et al., 2006), or a brute-force Monte Carlo sampling approach (Sohn et al., 2004; Tan et al., 2007). However, simplifying assumptions such as linearity are known to produce erroneous exposure characterizations in the forward mode of analysis and therefore they should be considered with great caution in the inverse mode of analysis. An "interim" step in interpreting biomonitoring data is provided by biomonitoring equivalents, which employ forward modeling to determine exposure levels that correspond to the reference doses of the chemicals of concern (Hays et al., 2007). This indeed can provide useful insight in relating biomarker levels to regulatory requirements under the simplifying assumption of chronic, steady exposures. However, as Hays et al. (2007) also implicitly recognize, these techniques are not applicable to reconstructing real-world exposure scenarios, because such scenarios typically involve non-steady, transient exposures with variable frequencies, durations, and magnitudes.

Computational inversion can be formulated as a problem where the objective is to identify the specific input combinations or distributions that best explain the observed outputs while minimizing an "error metric." In the case of exposure and dose models, the inputs involve spatial and temporal information on microenvironmental media concentrations of contaminants as well as corresponding information on human activities that result in intakes of these contaminants, whereas the outputs are observed biomarkers, and the error metric can be defined in terms of population variation, random error, etc.

Computational inversion/optimization techniques can be broadly classified into "deterministic" and "stochastic" (see e.g. Moles et al., 2003). These methods in general utilize a systematic exploration of the input space to identify the "global" minimum of the error metric and the corresponding values of inputs. Stochastic methods randomly sample the input space and do not necessarily guarantee convergence to the best solution to the inversion problem. Deterministic methods can, in principle, guarantee the convergence to a global minimum, though not necessarily in a finite time, but they often lead to a prohibitive "exponential" increase in computational resources as the number of inputs increases. In contrast, stochastic methods can provide a reasonable solution with relative efficiency; this solution is in practice often the best available for modest computation times (Moles et al., 2003). Furthermore, these methods are usually quite simple to implement and use, especially for "black box" models. Many specialized monographs are available covering a wide range of both deterministic and stochastic inversion techniques (Vogel, 2002; Aster et al., 2005; Tarantola, 2005). A selective list of representative deterministic and stochastic inversion methods, that can be used for exposure reconstruction, is presented in Tables 2 and 3.

Deterministic Inversion The exposure reconstruction problem can be formulated as a global minimization problem that involves finding possible exposures $\mathbf{x}$ by minimizing a "cost function" J, based on observed biomarker data $\mathbf{b}'(t_i)$ at each time point t_i (a total of N_{meas} measurements), and a forward model for estimating biomarkers: $\mathbf{b} = \mathbf{m}(\mathbf{x}, t_i)$. Additionally, constraints can be included in the form of

- bounds on possible exposures ($\mathbf{x}^{\text{L}} \leq \mathbf{x} \leq \mathbf{x}^{\text{U}}$),
- equality constraints on the model ($f(\mathbf{x},\mathbf{b},t) = 0$), and
- inequality constraints ($g(\mathbf{x},\mathbf{b},t \leq 0$).

Typical examples of J include

$$J = \sum_{i=1}^{N_{\text{meas}}} (\mathbf{b}'(t_i) - \mathbf{m}(\mathbf{x},t_i))^{\text{T}} (\mathbf{b}'(t_i) - \mathbf{m}(\mathbf{x},t_i)) \quad (1)$$

(least square minimization)

and

$$J = -\text{L}(\mathbf{m},\mathbf{x},\mathbf{b}') = -\prod_{i=1}^{N_{\text{meas}}} f_{\mathbf{x}}(\mathbf{b}'(t_i)|\mathbf{x}, \mathbf{m}(\mathbf{x},t_i)) \quad (2)$$

(maximum likelihood estimation)

where the likelihood L is expressed as a product of likelihood of the data at each point through function $f_{\mathbf{x}}$, which depends on the assumptions regarding the distribution of "errors" (i.e. the differences between observations and model estimates).

Stochastic Inversion/Bayesian Approach A general probabilistic framework for inverse problem solution is provided by the Bayesian approach, which is based on

Table 2. Representative "deterministic" inversion methods, example environmental and biological applications and software.

Method/technique	Example application area	Example available software
Regression methods		
Non-linear least squares (Dennis et al., 1981; Yano et al., 1989; Kaltenbach and Vistelle, 1994; Bender and Heinemann, 1995)	Estimation of PK parameters (Kaltenbach and Vistelle, 1994; Muzic and Christian, 2006)	MOT (Mathworks, 2006c); NIST (NIST, 2006); Boomer (Bourne, 1989); MultiFilt (Yano et al., 1989)
Singular value decomposition (SVD) (Paige, 1986; Anderson, 1999)	Spectrum reconstruction from dose (Armbruster et al., 2004); pathway analysis of gene expression (Tomfohr et al., 2005)	Numerical recipes (NR) (Press, 2002); Netlib-TOMS (ACM, 2006)
Extreme (LeBlanc et al., 2006)/logistic	Survival analysis (LeBlanc et al., 2006)	Matlab toolboxes
Projection pursuit (PPR) (Friedman and Stuetzle, 1981; Lingjaerde and Liestol, 1998)	Phenol toxicity mechanism (Ren and Kim, 2003)	R (RFSC, 2006); XTAL (Cherkassky and Gehring, 1996)
Multivariate adaptive (MARS) (Friedman and Roosen, 1995); Bayesian MARS (Denison et al., 1998; Holmes and Denison, 2003; Nott et al., 2005)	Phenol toxicity mechanisms (Ren and Kim, 2003); GI absorption (Deconinck et al., 2005); subgroup disease–risk relationships (York et al., 2006)	Netlib-Misc (Netlib, 2006)
Gradient methods		
Automatic differentiation (Hovland et al., 2005; Ringrose and Forth, 2005)	Uncertainty in PBPKs (Isukapalli et al., 2000)	ADIFOR (Bischof et al., 1996); ADIC (Bischof et al., 1997); MAD (Kharche and Forth, 2006)
Steepest descent (Marquardt, 1963; Conway et al., 1970)	Inverse treatment planning (Hristov et al., 2002)	NR (Press, 2002); TOMS (ACM, 2006)
Likelihood methods		
Maximum (MLE); restricted MLE (Elzo, 1994; Jamshidian, 2004); approximate likelihood (Albertini et al., 2006); NONMEM (Overgaard et al., 2005; Tornoe et al., 2005)	Short- and long-term exposure reconstruction using PBPK for chloroform (Georgopoulos et al., 1994; Roy et al., 1996)	Matlab (MathWorks Inc., 2006); NR (Press, 2002); TOMS (ACM, 2006); NONMEM (Beal, 2002)

Bayes' Theorem

$$p(\mathbf{x}|\mathbf{b}) = \frac{p(\mathbf{b}|\mathbf{x})\,p(\mathbf{x})}{\int p(\mathbf{b}|\mathbf{x})\,p(\mathbf{x})\mathrm{d}\mathbf{x}} \tag{3}$$

where $\mathbf{x}$ and $\mathbf{b}$ are defined as in the previous subsection and the probability densities $p(.)$ are interpreted as representing available knowledge/information rather than as representing intrinsic properties of the variables.

If prior knowledge information on exposure attributes is represented by $p_{\mathrm{prior}}(\mathbf{x})$, and the theoretical ("model") knowledge of the relation between $\mathbf{x}$ and $\mathbf{b}$ is represented by $p_{\mathrm{theory}}(\mathbf{x}|\mathbf{b}')$,

$$\begin{aligned} p_{\mathrm{theory}}(\mathbf{b}|\mathbf{x}) &= p_{\mathrm{model}}(\mathbf{b}|\mathbf{x}) \\ p(\mathbf{x},\mathbf{b}) &= p_{\mathrm{theory}}(\mathbf{b}|\mathbf{x})p_{\mathrm{prior}}(\mathbf{x}) \end{aligned} \tag{4}$$

For a specific set of biomarker measurements $\mathbf{b}'$, the posterior (inferred) distribution of exposure attributes will be $p_{\mathrm{inferred}}(\mathbf{x}|\mathbf{b}')$ and

$$p_{\mathrm{prior}}(\mathbf{x},\mathbf{b}') = p_{\mathrm{inferred}}(\mathbf{x}|\mathbf{b}')p_{\mathrm{prior}}(\mathbf{b}');\ \text{for fixed } \mathbf{b}' \tag{5}$$

Overall,

$$p_{\mathrm{prior}}(\mathbf{x}) = \int p_{\mathrm{prior}}(\mathbf{x},\mathbf{b})\mathrm{d}\mathbf{b} = p_{\mathrm{prior}}(\mathbf{x}) \int p_{\mathrm{theory}}(\mathbf{b}|\mathbf{x})\mathrm{d}\mathbf{b} \tag{6}$$

Therefore,

$$\underbrace{p_{\mathrm{inferred}}(\mathbf{x}|\mathbf{b}')}_{\text{Posterior distribution}} = \frac{p_{\mathrm{theory}}(\mathbf{b}'|\mathbf{x})p_{\mathrm{prior}}(\mathbf{x})}{\int p_{\mathrm{theory}}(\mathbf{b}'|\mathbf{x})p_{\mathrm{prior}}(\mathbf{x})\mathrm{d}\mathbf{x}} \tag{7}$$

If $p_{\mathrm{model}}(\mathbf{m},\mathbf{x})$ denotes the probability density of the "true" model output being $\mathbf{m}$ for inputs $\mathbf{x}$,

$$p_{\mathrm{theory}}(\mathbf{b}|\mathbf{x}) = p_{\mathrm{error}}(\mathbf{b}|\mathbf{m})p_{\mathrm{model}}(\mathbf{m}|\mathbf{x})\mathrm{d}\mathbf{m} \tag{8}$$

where $p_{\mathrm{error}}(\mathbf{b}|\mathbf{m})$ is the probability of measuring $\mathbf{b}$ when the true value is $\mathbf{m}$; i.e. it is the distribution of "measurement error" and not "model error." Therefore

$$p(\mathbf{x},\mathbf{b}) = \frac{p(\mathbf{x}) \int p_{\mathrm{error}}(\mathbf{b}|\mathbf{m})p_{\mathrm{model}}(\mathbf{m}|\mathbf{x})\mathrm{d}\mathbf{m}}{\int p(\mathbf{x}) \int p_{\mathrm{error}}(\mathbf{b}|\mathbf{m})p_{\mathrm{model}}(\mathbf{m}|\mathbf{x})\mathrm{d}\mathbf{m}\mathrm{d}\mathbf{x}} \tag{9}$$

- For a "deterministic" model, $p_{\mathrm{model}}(\mathbf{m}|\mathbf{x})$ is a delta function centered on $\mathbf{m}(\mathbf{x})$.
- For a "deterministic" model with "model error" (or uncertainty), $p_{\mathrm{model}}(\mathbf{m}|\mathbf{x})$ represents the distribution of uncertainty in $\mathbf{m}(\mathbf{x})$.
- For a "stochastic" model, $p_{\mathrm{model}}(\mathbf{m}|\mathbf{x})$ represents the distribution of predictions that can be obtained for a specific (fixed) set of inputs $\mathbf{x}$.

Methods for Exposure Reconstruction from Population Biomonitoring Studies

There are several Bayesian and non-Bayesian approaches for inversion of biomarker data for obtaining exposures and PBTK model parameters. Georgopoulos et al. (2008) discuss the relative advantages of Bayesian and non-Bayesian approaches for population parameter estimation using PBTK

Table 3. Representative "stochastic" inversion methods and statistical pattern methods, along with example environmental and biological applications and software.

Method/technique	Example application area	Example available software
Evolutionary methods		
Tabu search (Glover, 1994; Glover et al., 1995; Youssef et al., 2001) Genetic algorithms (Kazarlis et al., 2001; Youssef et al., 2001; Coello and Pulido, 2005)	Cancer classification (Tahir et al., 2005); protein structure prediction (Blazewicz et al., 2005); ordering microarray data (Moscato et al., 2006); PBPK parameter estimation (Holmes et al., 2000); outlier detection (Vankeerberghen et al., 1995)	OpenTS (Abramov et al., 2005); OptQuest (OptTek Systems, 2006); Matlab toolbox (Mathworks, 2006a); GEATbx (Pohlheim, 2005)
Artificial neural networks (Gobburu and Chen, 1996; Chow et al., 1997; Hashemi and Young, 2003; Winkler, 2004)	PK parameters (Opara et al., 1999); QSPK relationships (Nestorov et al., 1999); ADME prediction (Balakin et al., 2005); *in silico* CYP 450 estimation (Yap et al., 2006)	Matlab NN toolbox (Mathworks, 2006b); NeuroSolutions (NeuroDimension, 2006)
Bayesian methods		
Bayesian MCMC (Metropolis–Hastings–Gibbs) (Gelman, 2004)	Reconstruction of intakes with PBPK (Gosselin et al., 2006); population pharmacokinetics (Dokoumetzidis and Aarons, 2005)	MCSIM (Bois, 2000); PKBUGS (Lunn et al., 2002); Matlab toolboxes
Bayesian model averaging	Benchmark dose estimation for As (Morales et al., 2006)	Matlab toolboxes
Surrogate/FEOM modeling		
SRSM (Isukapalli et al., 1998) + MCMC	Groundwater contamination (Balakrishnan et al., 2003)	MENTOR (Georgopoulos and Lioy, 2006)
Regularization (Morozov and Stessin, 1993; Engl et al., 1996; Tenorio, 2001; Farquharson and Oldenburg, 2004)	Mass spectrometry (Mohammad-Djafari et al., 2002); reconstruction of membrane potentials (Messnarz et al., 2004)	RegTools (Hansen, 1994); ORBIT (Johansen and Foss, 1998); NLCSmoothReg (Wendlandt et al., 2005)
Deconvolution	PBPK reconstruction (Sparacino et al., 2002); estimation of metabolism (Yamashita et al., 1995); transport parameters (Nair and Gratzl, 2005)	WinStoDec (Sparacino et al., 2002)
Maximum entropy (Clarke and Janday, 1989; Gamboa and Gassiat, 1997)	Biomagnetic inversion (Clarke and Janday, 1989); absorption kinetic rates (Ablonczy et al., 2003)	BMElib (Christakos et al., 2001; Serre, 2006)
Statistical pattern recognition methods		
CART (Breiman, 1984; Hastie et al., 2001; Venables and Ripley, 1999); MART (Friedman and Meulman, 2003); C4.5 method (Quinlan, 1993; Treenet (Jitnah and Nicholson, 1997; Flouris and Duffy, 2006)	Biomarkers–demographics correlation (Balakrishnan et al., 2003); oral absorption (Bai et al., 2004); biomarkers from mass spectrometric data (Liu and Li, 2005); cervical neoplasia (Friedman and Meulman, 2003)	S-Plus (Venables and Ripley, 1999); Matlab (Mathworks, 2006d); DTREG (Sherrod, 2006); ENTOOL (Merkwirth, 2006); R (CRAN, 2006); Treenet/MART (Salford Systems, 2006)

models and biomarker data. Some of the recently proposed methods for exposure reconstruction are discussed next.

Exposure Conversion Factor Approach The exposure conversion factor (ECF) method, proposed by Tan et al. (2006), assumes that the relationship between biomarker and dose can be approximated by a linear function for exposure reconstruction purposes. This approach involves three steps: (1) generating samples for forward model runs from distributions of possible exposure, physiological, and biochemical parameters, (2) running the forward model using a set of input samples from these distributions, and (3) inverting the distribution of output (i.e. simulated biomarker levels) to obtain an "ECF." Using the ECF and the distribution of observed biomarkers, the possible exposures for that particular biomarker distribution can then be estimated through a straightforward convolution (Tan et al., 2006).

In a typical application of this simple method, the PBTK model can be run using a unit dose or concentration value, and various samples from the possible distributions of parameters such as activities, physiological parameters, biochemical parameters, biomarker sample times, etc., to generate a set of biomarker levels. These levels then provide the distribution of biomarkers for a unit exposure metric, which can be inverted to obtain an ECF in units of the exposure metric divided by biomarker level units. The ECF can then be multiplied by the values of available

biomonitoring data (e.g. from biomarker databases such as NHEXAS or NHANES) to produce an estimate of dose distributions for the corresponding population. This convolution is performed by multiplying samples from the biomarker distribution with samples from the ECF distribution. The aggregate samples then provide the distribution of reconstructed exposures. Though this method is conceptually simple and straightforward to use, as it involves direct generation of samples of the corresponding statistics from these samples, the ECF can be highly sensitive to the assumptions of the prior distributions. Furthermore, the assumption of linearity can sometimes produce unreasonably large tails in the distribution of reconstructed exposure metrics, especially when exposures occur infrequently, and the sampling time relative to the last exposure is unknown. The section titled Impact of Biochemical Properties and Sampling Characteristics on Reconstruction in the following provides further details on the impact of exposure and sampling times on the outcomes of exposure reconstruction.

Discretized Bayesian Approach This approach, which was used, for example, by Sohn et al. (2004) and Tan et al. (2007), employs a simplified, discrete Bayesian scheme to estimate the posterior probability of exposures from the biomarker data and prior exposure distributions. Posterior probabilities of exposures/doses are computed using biomarker data and forward modeling results at regularly spaced samples ("bins") or random samples, spanning the range described by prior probabilities. The forward model results are then divided into a set of regular intervals or bins, and the posterior probability is estimated from the samples that agree most closely with the biomarker data (e.g. those that belong to the same bin) (Tan et al., 2007). Like the ECF method, it requires strong informative prior probability distributions of exposure-related activities to produce a realistic reconstruction. Furthermore, as the number of dimensions increases, the corresponding sampling space becomes "vastly empty" (Tarantola, 2005), in the sense that the number of points required to adequately sample from this space increases exponentially as the number of dimensions increases. This approach, therefore, necessitates trade-offs with respect to sample size, the resolution of the sampling, and the accuracy of the results.

Bayesian Markov Chain Monte Carlo The Markov Chain Monte Carlo (MCMC) approach provides a means for sampling from the "posterior probability distribution" without having to sample the entire range of the prior distribution. The method requires defining the prior distributions, available biomarker data, and a likelihood function defining the likelihood of the data given a set of forward model parameters; then the MCMC approach involves marching in the sample space based on acceptance criteria that consider the likelihood of the data given the parameters. MCMC techniques (Gilks et al., 1998; Gamerman and Lopes, 2006) have been coupled with PBTK models for forward modeling of population health risk assessment (Covington et al., 2007), and inverse modeling for parameter estimation (Bois, 2000; Yokley et al., 2006; Bois et al., 2007; Yang et al., 2008). In practice, with adequate data and prior information on exposure metrics and population toxicokinetic parameters, it is possible to directly apply MCMC techniques to estimate individual and population exposure parameters. For example, Allen et al. (2007) applied the MCMC technique to the reconstruction of long-term exposures to MeHg using steady-state approximations, basing this approach on the fact that the half-life of MeHg is very large, and assuming that time-varying exposures were unimportant. However, for many real-life exposure scenarios, the time-varying exposures cannot be assumed to be negligible, and, in general, would have to be explicitly incorporated into the reconstruction process, thus leading to intensive data requirements (including multiple biomarker measurements for each individual during each measurement period) and non-steady-state models. When sufficient information is not available for estimating individual parameters, the estimation of population parameters using MCMC and PBTK models becomes impractical, as there would be limited data for the hierarchical modeling needed to estimate these parameters. This is typically the case, as for example with studies such as NHEXAS and NHANES, which include relatively few biomarker measurements (often just one measurement) for different individual–chemical combinations.

Even though the estimation of population parameters without the estimation of the parameters for individuals may appear feasible using the MCMC approach, it is in fact impractical. The estimation of population parameters, when one distribution is assumed to be representative of "all" individuals, quite often results in an artificial fit of a single parameter and thus results in a large "error term." These errors will be substantial when there are large inter-individual variabilities with regard to exposure patterns.

Major Factors Influencing Exposure Reconstruction

Several major factors influence the feasibility and efficacy of exposure reconstruction from biomarkers. The specificity and the sensitivity of the biomarker with respect to the exposure metric of interest (e.g. concentration of the agent of concern) are two of the most important such factors. Lack of specificity can lead to the problem of identifiability, whereas the lack of sensitivity may result in large uncertainties in the reconstruction results. Biochemical properties of absorption, distribution, metabolism, and elimination (ADME) impact the types of exposures that can be estimated from the biomarker data (Hays et al., 2007). Variability in ADME characteristics also results in a significant variation in the biological half-lives of different groups of environmental

pollutants, such as volatile organics, organophosphate pesticides, and toxic metals. Some of these properties are also highly variable for a given contaminant within a population, reflecting inter-individual physiological and biochemical variabilities, which may result in significant uncertainties in the use of biomarkers for exposure reconstruction.

Various exposure characteristics, such as the frequency, magnitude, and duration of exposures, provide supplemental information that is valuable and often necessary for reconstructing exposures from biomarkers. Because biologically relevant exposures can occur through multiple pathways and mechanisms (e.g. direct exposure to a metabolite *versus* exposure to the parent compound), supplemental exposure information such as data on mouthing behavior, activities, macro- and micro-environmental source locations and personal concentration levels monitoring, also provide realistic constraints on possible exposures from specific pathways. Additionally, the adequacy of the biomarker data can be characterized in terms of the specificity of the available supplementary information such as the time of collection of biomarkers in relation to other relevant parameters (e.g. amount of urine collected, total urinary void volume, last time of urination, etc.). Other characteristics of the data sets, such as the detection limits of contaminants of concern, can also significantly impact the reconstruction process. However, while all of the above should be considered, in practice, two other factors impact the inversion process significantly: the applicability and adequacy of the forward model (e.g. toxicokinetic/toxicodynamic model) and the efficacy of the computational inversion technique employed for the reconstruction.

Impact of Biochemical Properties and Sampling Characteristics on Reconstruction The "residence time" or "age" of an "observed" (i.e. measured) biomarker molecule can be defined as the time elapsed since it entered (or was generated in) the organ or organism studied. The observed biomarker levels (molecules of either a chemical or its metabolites, potentially involving multiple exposures across multiple timescales) represent an "integration" over molecules of different "ages," dependent on the time each molecule entered or was generated in the system, and on elimination kinetics. As an example, for chemicals and metabolites with relatively short half-lives, only the exposure history of the previous days or weeks can be estimated. For those with longer half-lives, larger timescales of exposure history have to be considered, and the influence of confounding sources creates additional uncertainties.

In general, the age or residence time distribution can be defined as $R(t) = b(t)/\int_0^\infty b(t)\mathrm{d}t$, where $b(t)$ is the concentration of a chemical in the system (i.e. a single biomarker concentration in blood, tissue, or urine) at time t. Furthermore, $R(t)\mathrm{d}t$ represents the fraction of the molecules that have spent a time between t and $t+\mathrm{d}t$ in the system. The corresponding cumulative distribution function, $F(t)$ is the fraction of the molecules that have spent time t or less in the system, and is given by $F(t) = \int_0^t R(t)\mathrm{d}t$.

In the simplest case of a steady, continuous exposure, and of toxicokinetics that can be described adequately by a single-compartment PK model, $b(t) = x'\mathrm{e}^{-kt}$, $R(t) = k\mathrm{e}^{-kt}$, and $F(t) = 1-\mathrm{e}^{-kt}$, where x' is a known exposure concentration, and the elimination rate k is related to half-life $t_{1/2}$ as $k = \log_2/t_{1/2}$. These equations for age distribution also represent the output *versus* time function from a bolus input.

In the case of discrete repeated ("cyclical") exposures of time period Δt, assuming that the biomarkers are collected at a time $\lambda\Delta t$ after the end of the last exposure ($0<\lambda\leq 1$), and that all exposure concentrations are equal to x', the relative contribution of exposures that occurred at different times can be expressed in terms of "cycles" of exposure, as follows:

$$b(n) = \sum_{i=0}^{n} x'\mathrm{e}^{-k(i+\gamma)\Delta t} \text{ and } \quad F(n) = \sum_{i=0}^{n} \mathrm{e}^{-ik\Delta t} \Big/ \sum_{j=0}^{\infty} \mathrm{e}^{-jk\Delta t} \tag{10}$$

Figure 2 shows, as a function of different half-lives, the relative contributions of different timescales of continuous, steady exposures to observed chemical biomarker levels in a single-compartment system. These calculations are an extension of the approach presented in NRC (2006). Figure 3 shows the relative contributions of prior exposures (discrete bolus doses) to observed chemical biomarker levels, as a function of different sampling times and exposure frequencies. When Δt is significantly larger than the exposure timescale of interest (β), $\lambda\Delta t$ becomes an important variable. When $\lambda\Delta t \leq \beta$, the biomarker captures exposures occurring within the past time period β. Otherwise, exposures occurring within β are not captured. It must be noted that the relative contributions are independent of the time of sampling within a "cycle," whereas the relationship between the magnitude of the observed biomarker concentration and the exposure concentration is highly dependent on the time of sampling within a cycle.

As an example, consider a scenario where exposures occur once per month ($\Delta t = 1$ month), and the biomarker contribution due to exposure within the past 1 day ($\beta = 1$ day) is known. If a sample is taken 12 h after the last exposure ($\lambda\Delta t = 12$ h), the biomarker will incorporate contributions that originate from exposure occurring within the past 24 h. If the sample is taken 30 h after the last exposure, there is zero mass contribution due to exposures within the past 24 h. Figure 3 illustrates the effect of $\lambda\Delta t$ and Δt on the half-life and exposure timescale relationship previously shown for the steady-state case in Figure 2.

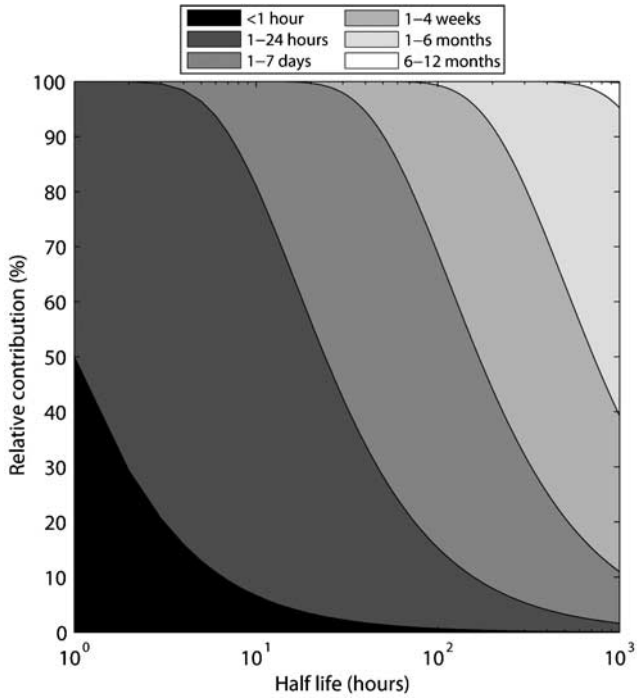

Figure 2. Impact of half-life on the relative contributions of different timescales of exposures to observed chemical biomarker levels. The example shown above is based on a one-compartment PK model (linear decay); the biomarker represents the level of chemical in the compartment.

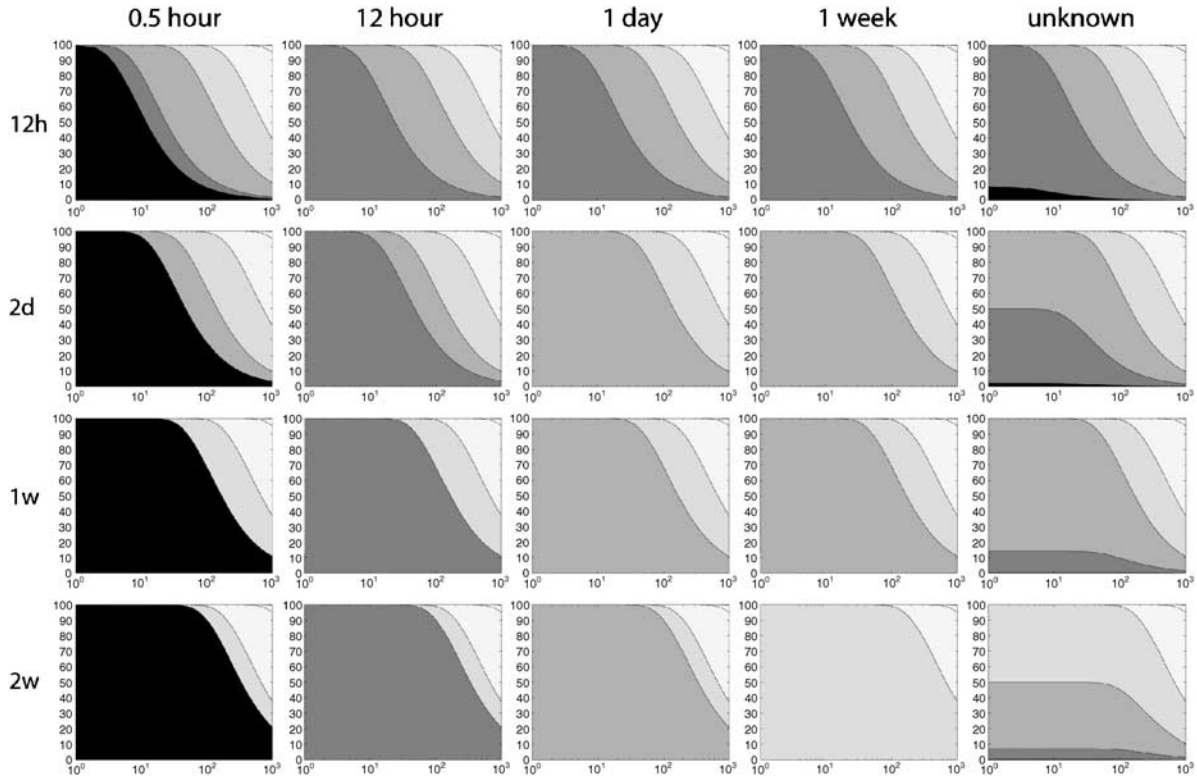

Figure 3. Contribution of prior exposures to observed biomarker levels as a function of intake frequency, sampling time, and biochemical properties. The rows represent the time period of exposure (e.g. every 12 h, every 2 days, etc), the columns represent the time of sampling after the last exposure. For cases when sampling time is unknown, the mean values of the contributions are shown, assuming a uniformly random sampling time. The legend for the scales of gray is the same as in Figure 2.

As the length of time between biomarker sampling and the last exposure increases, the contribution to biomarker levels from recent exposures diminishes. As the frequency of cyclic exposure decreases, relative contributions become more sensitive to biomarker sampling time. If exposures occur with high frequency, the system approaches a pseudo steady state and relative contributions are less sensitive to sampling time. In general, the residence time analysis approach can be readily expanded to include realistic PBTK models instead of half-life parameterizations, and estimates such as those shown in Figure 3, can be generated for different types of exposure and sampling profiles for different classes of compounds. Screening level estimates from the residence time analysis can provide help in identifying of the types of reconstruction that are feasible and realistic for a given chemical, biomonitoring characteristics, and possible exposure profiles.

A Simple Toxicokinetic Model for CPF Exposure

The case studies used for evaluation of reconstruction approaches in the following sections utilize a simple toxicokinetic model for CPF exposures. This model, formulated for demonstration purposes, is adapted from a two-compartment PK model for CPF and 3,5,6-trichloro-2-pyridinol (TCPy) used by Rigas et al. (2001). The rapid transformation of CPF to TCPy, and the fact that currently available CPF PBTK models (Timchalk et al., 2002a, b) describe TCPy using one-compartment kinetics, provide the justification for employing this assumption in this simple demonstration. This simple toxicokinetic model can be described by the following equations:

$$\begin{aligned} \frac{\mathrm{d}C_{\mathrm{a}}}{\mathrm{d}t} &= -k_{\mathrm{a}}C_{\mathrm{a}} \\ \frac{\mathrm{d}C_{\mathrm{b}}}{\mathrm{d}t} &= k_{\mathrm{a}}C_{\mathrm{a}} - k_{\mathrm{e}}C_{\mathrm{b}} \\ \frac{\mathrm{d}U_{\mathrm{TCP}}}{\mathrm{d}t} &= k_{\mathrm{e}}C_{\mathrm{b}}V_{\mathrm{d}} \\ C_{\mathrm{a}} &= C_{\mathrm{a}}(t_0) + \frac{S\,R\,F\,D_{\mathrm{CPF}}}{V_{\mathrm{d}}}\left(\frac{M_{\mathrm{TCP}}}{M_{\mathrm{CPF}}}\right) \end{aligned} \tag{11}$$

where M_{TCP} is the molecular weight of TCPy; M_{CPF}, molecular weight of CPF; D_{CPF}, bolus dose of CPF (μg; the dose here refers to the total amount instead of being scaled by the bodyweight); C_{a}, concentration of TCPy in absorption reservoir (μg/l); $C_{\mathrm{a}}(t_{\mathrm{o}})$, concentration of TCPy in absorption reservoir before bolus event at time t_{o} (μg/l); C_{b}, concentration of TCPy in body (μg/l); F, absorption factor for ingested dose; R, stoichiometric ratio of CPF to TCPy conversion; S, selectivity (which refers to the amount on molar basis of the absorbed material that can be collected as metabolite of interest; Rigas et al., 2001); V_{d}, volume of distribution (L); k_{e}, elimination rate constant (1/h); and k_{a}, absorption rate constant (1/h).

Since the focus of the case studies is on method evaluation and on highlighting data gaps, the demonstration case studies employ significant simplifications. For example, the entire absorbed CPF is assumed to be directly converted to TCPy, and variability in fractional absorption is neglected; thus, the reconstruction focuses only on the absorbed amount of CPF. These simplifications do not significantly affect the evaluation process here. In fact, more detailed PBTK models, which incorporate additional information about bioavailability, can be used directly for exposure reconstruction employing the approaches evaluated here; however, the approach adopted here aims to help the reader focus on the reconstruction process and methodology rather than on the details of comprehensive PBTK modeling that can be employed in the reconstruction process.

Case study 1: CPF dose reconstruction from available NHEXAS-MD data

The NHEXAS Maryland (NHEXAS-MD) data set is longitudinal, and contains multiple biomarker and environmental measurements for households over a period of time. With respect to CPF, the available biomarker data are urinary TCPy measurements corresponding to the first void of the day. These biomarkers represent statistically weighted probability samples, with each biomarker associated with a statistical weight. For the purposes of the evaluation of reconstruction methodologies, these weights are assumed to be equal for simplification purposes, and only the adult population (over 18 years of age) was modeled. The concentrations of CPF in food, air (at home), dust, etc., are also available in the NHEXAS-MD dataset. The corresponding TCPy concentrations in food, however, were not measured. The food intake can be estimated through the 4-day duplicate plate, but the actual amount of food consumed was not readily available. The urinary void volume, the time of earlier urination, and the last food intake time, are also not available, thus introducing significant uncertainties into the process of exposure reconstruction.

Two exposure scenarios are examined here:

(1) Steady-state, continuous exposure, which neglects potential issues in temporal variability in both exposure timing and biomarker sampling (referred to in the following as scenario 1).
(2) Time-varying dose and biomarker collection, assuming "reasonable" distributions for the frequency of intake and for the timing of the biomarker collection. An analysis of three separate time-varying exposure scenarios is performed, where each scenario employs different constraints on the exposure profile (referred to in the following as scenarios 2a–c):
    (a) Bolus dose frequency is modeled as a complex time–activity profile. An individual may consume zero, one, two, or three meals containing CPF in a given

day. An exposure may occur randomly on any, all, or none of the days in a week. More details on this scenario are given in Case Study 2.

(b) Bolus dose frequency is fixed at certain values (i.e. once per day, once per week).

(c) Modifications and combinations of scenarios 2a and 2b. These include scenarios where a background CPF inhalation dose occurs (sampled from summary distributions presented by Pang et al., 2002), and/or dietary exposure to the metabolite TCPy occurs.

The NHEXAS-MD data set contains multiple samples per individual. However, for any given individual, the sampling intervals are separated by at least a month. Since the biomarker half-life is approximately 1 day (ATSDR, 1997), it was assumed that the measurements for the same individual at different sampling intervals are independent, and the correlations between them are assumed to be negligible.

There are significant differences in the types of uncertainties associated with scenarios 1 and 2. For scenario 1, the average daily dose (ADD) is distributed evenly throughout the day, and therefore dynamic uncertainties do not exist. For scenarios 2a–c, the ADD is apportioned differently throughout time: t. The ADD is not known, the apportionment is not known, and the modeled "biomarker collection time" is randomly assigned to a time before the first meal of the day. Since the half-life of TCPy is 1 day, there is the potential that a sampled individual receives a high ADD, which occurs infrequently. Depending on the sampling time, the biomarker concentrations can vary from extremely high to non-detectable levels. The less frequent the exposure, the more likely the biomarker measurement will be a non-detect, regardless of the exposure magnitude. Some simulated individuals may receive a moderate dose once per day, and the biomarker profile approaches the steady-state condition. A more detailed discussion of exposure frequency and ADD in relation to biomonitoring can be found in Hays et al. (2007).

In this case study, uncertainties in PK parameters are assumed to be negligible and so these parameters are set to the mean values derived from an original toxicokinetics study by Nolan et al. (1984). For both the steady-state and time-variant biomarker dose reconstructions, only the total dose absorbed was considered because the absorption fraction and intake amount are coupled, and hence, individually unidentifiable. There is also a factor-of-two difference in the fraction absorbed that has been found across different studies (e.g. Nolan et al., 1984; Timchalk et al., 2002a).

The simple toxicokinetic model is used to predict cumulative amount of TCPy excreted in urine. The biomarker is estimated from the model calculations of cumulative TCPy excreted over a 6- to 10-h period (according to a uniform random distribution), resulting in an average TCPy excretion rate. NHEXAS biomarkers are assumed to represent morning void samples, with overnight bladder TCPy accumulation occurring over approximately 8 h. Converting model output (in mass TCPy/day) to biomarker measurement units (mass TCPy/volume urine, and mass TCPy/mass creatinine) introduces additional uncertainties, since urinary liquid or creatinine excretion rates must be known. The NRC biomonitoring report (NRC, 2006) notes that inter- and intra-individual variation in urinary water and creatinine content can be a source of biomarker misinterpretation (specific discussion on CPF exposure and TCPy biomarkers is also contained in the above report). The following three methods for converting the units are considered here:

- For data from the Minnesota study, Rigas et al. (2001) converted measured urinary TCPy concentrations to urinary TCPy excretion rates. TCPy concentration in urine is multiplied by urine void volume, and divided by the length of time urine accumulated in the bladder (estimated from the Minnesota study questionnaire). However, the NHEXAS-MD study does not provide urine void volume amounts, so this approach is not used here.
- An assumption of a mean daily liquid urine output of 22 ml/kg body weight is used to convert TCPy rates to mass TCPy/volume urine (Wilson et al., 2003; Morgan et al., 2005).
- Age, gender, body weight, and body height relationships are used to estimate creatinine excretion rates (Mage et al., 2004). This converts TCPy rate to mass TCPy/mass creatinine (which is reported in NHEXAS-MD). Advantages of this method over the previous two methods listed above for population exposure assessment purposes have been previously discussed in Barr et al. (2005).

Analysis of the Linked PK/Biomarker Approach Using Forward Modeling

A CPF exposure/dose model for the NHEXAS-MD population was coded based on the approach of Pang et al. (2002) and evaluated for its use as a "prior" estimate of CPF dose. It was found that by using the exposure estimates from Pang et al. (2002) as inputs to the PK model (using both steady-state and time-variant assumptions) the TCPy biomarker levels are significantly under-predicted (Figure 4). This under-prediction is likely to be even more severe when the variability in the CPF absorption, which can be as low as 18% (Timchalk et al., 2002b), is taken into account. This leads to the conclusion that the NHEXAS-MD-derived exposure estimates are inadequate for use as "prior" exposure information in this particular case. This limitation can be attributed to one or more of the following possible reasons:

- The exposure model is not appropriate, as the available environmental and biomarker measurements correspond to spot samples;

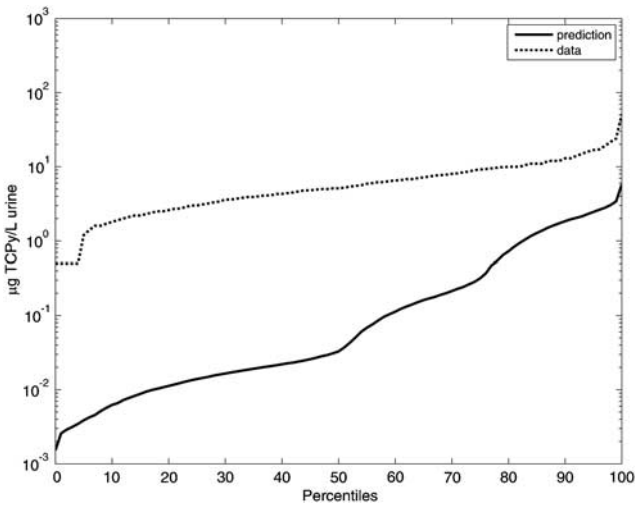

Figure 4. Comparison of urinary TCPy biomarker levels predicted using exposure estimates of Pang et al. (2002) for the NHEXAS-MD population with the toxicokinetic model, and actual NHEXAS-MD biomarker measurements.

- Food and air concentrations change with time, and more of an exposure history is needed to reduce exposure/dose uncertainty;
- The NHEXAS-MD population also experienced direct TCPy exposure, or exposure to CPF-methyl, a grain fumigant, which metabolizes to TCPy; and
- Uncertainties in both TCPy exposure and in CPF exposure/dose modeling exist.

Previous studies mentioned in the NRC biomarkers report have found that levels of TCPy in multimedia situations are generally comparable to CPF, and TCPy levels in solid food are an order of magnitude higher than CPF (Wilson et al., 2003; Morgan et al., 2005; NRC, 2006). It has also been noted that only a small percentage of dietary CPF exposure correlates with the TCPy levels seen in urine for the NHEXAS-MD data set (MacIntosh et al., 2001; Egeghy et al., 2005), further supporting the strong possibility that neglecting TCPy exposure in dose reconstruction over-predicts CPF exposure (Barr and Angerer, 2006).

In Case Study 1, it was assumed that the CPF exposure is the primary source of the TCPy biomarker. However, given the apparent inconsistency of NHEXAS exposure and biomarkers data, an additional biomarker inversion was performed to estimate population CPF exposure that considers direct TCPy exposures (see Case study 2). This inversion also used the estimates of NHEXAS-MD population CPF exposures developed by Pang et al. (2002). The PK model was modified to allow for both CPF and TCPy exposure contributions to urinary TCPy biomarkers, and was used in subsequent exposure reconstruction. However, without CPF exposure data, the decoupling of TCPy and CPF exposure contributions to biomarker levels is not possible.

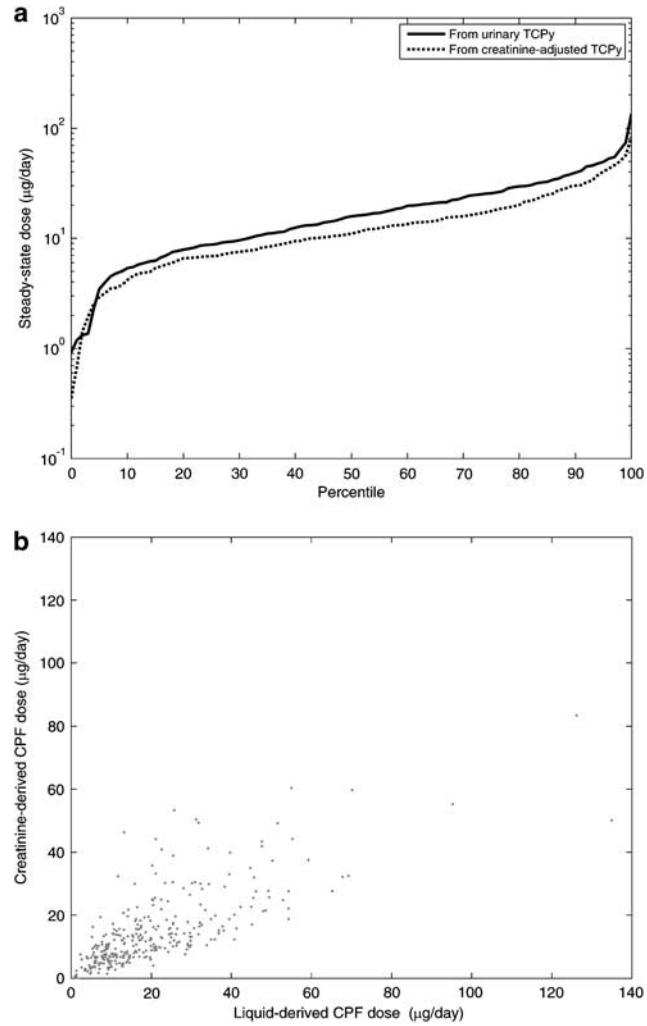

Figure 5. Comparison of predicted steady-state doses using creatinine-adjusted urinary TCPy biomarkers, and using liquid urinary concentration TCPy biomarkers: (a) cumulative distributions, and (b) scatter plot.

Computational Inversion Techniques used for Evaluation

Steady-State Approximation Assuming intake and excretion rates on a molar basis are equal, the following equations relate TCPy biomarker and CPF dose:

$$D_{\mathrm{CPF}} = C_{\mathrm{TCP\text{-}adj}} E_{\mathrm{cre}} (M_{\mathrm{CPF}} / M_{\mathrm{TCP}}) \tag{12}$$

$$D_{\mathrm{CPF}} = C_{\mathrm{TCP\text{-}liq}} E_{\mathrm{liq}} (M_{\mathrm{CPF}} / M_{\mathrm{TCP}}) \tag{13}$$

where D_{CPF} in the bolus dose of CPF (μg); E_{liq} in the urinary excretion rate (l/day); E_{cre} in the creatinine excretion rate (g/day); $C_{\mathrm{TCP\text{-}liq}}$ in the urinary TCPy biomarker (μg/l urine); $C_{\mathrm{TCP\text{-}adj}}$ in the urinary TCPy biomarker adjusted for creatinine concentration (μg/g creatinine); M_{TCP} in the molecular weight of TCPy; and M_{CPF} in the molecular weight of CPF.

Equation (12) is applicable when TCPy urinary biomarker measurements are specified in terms of mass TCPy/mass

creatinine, and Eq. (13) is applicable when biomarker measurements are specified in terms of mass TCPy/volume urine. An estimation of either liquid urinary production or urinary creatinine elimination is needed (as this is a source of potential biomarker misinterpretation stated earlier (NRC, 2006)). Both approaches have been previously used in the context of CPF exposure: Mage et al. (2004) used creatinine-adjusted biomarkers to estimate CPF exposures of the NHANES population; and Morgan et al. (2005) used absolute liquid concentrations and excretion rates in assessing CPF and TCPy exposures.

Time-Varying Approaches When incorporating temporal uncertainties in dose and biomarker collection, an algebraic solution is usually not possible. Dose and sample time uncertainties, and uncertainties in PK model parameters affecting dynamics and TCPy half-life need to be accounted for. The discretized Bayesian approach and the ECF approach were used in this study, as they have been developed relatively recently and have been applied for population dose reconstruction of short half-life chemicals (Tan et al., 2006, 2007). The Bayesian approach employed here used 40,000 model simulations, while the ECF approach used 5000 simulations. These TCPy biomarker data from NHEXAS-MD represent statistically weighted probability samples, with each biomarker associated with a statistical weight. In the case studies presented here, these weights are assumed to be equal; however, it should be pointed out that the incorporation of any statistical weights into the reconstruction process is relatively straightforward because, in the reconstruction process, random samples generated for inversion can use the weighted data instead of unweighted sampling from the available biomarker data.

Results for Case Study 1

Figure 5 compares CPF exposures estimated via a steady-state assumption, using either liquid urinary biomarkers, or creatinine-adjusted urinary biomarkers. Both methods appear to agree, with creatinine-adjusted predictions of daily CPF doses being slightly lower than the absolute liquid concentration predictions. At the tails of the distribution, however, there is a difference by about a factor of 5 in the lower tail of the distribution and about a factor of 2.5 in the higher tail, with the urinary concentration-based method predicting systematically higher intakes.

The inversion results appear to be sensitive to the dose profile assumption, as shown in Figures 6 and 7. The results also indicate that apportioning "ADD" randomly to once a

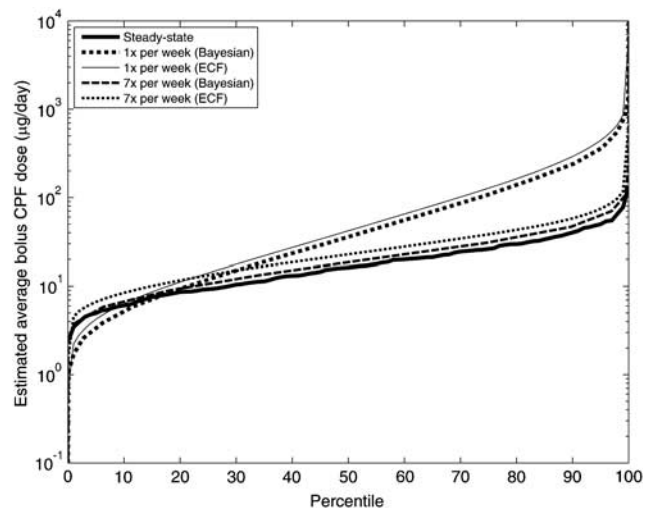

Figure 6. Cumulative distribution functions of dietary CPF uptakes for the NHEXAS-MD population estimated by the ECF and Bayesian methods, using three different dose assumptions. Results only show population for which the biomarkers were above the detection limit (which comprised 95% of the samples).

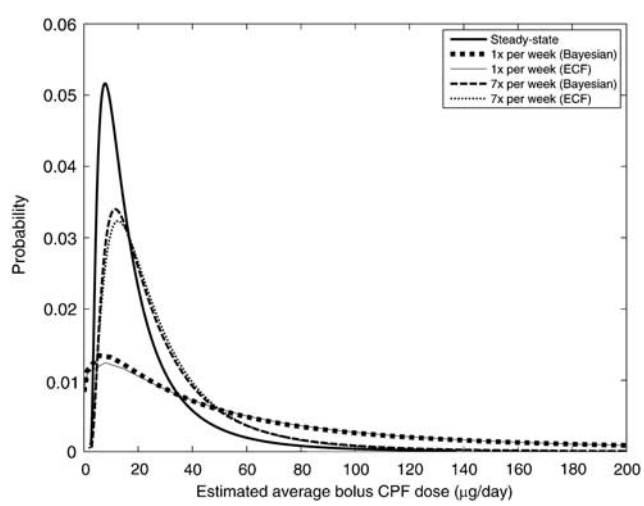

Figure 7. Probability density functions of dietary CPF uptakes for the NHEXAS-MD population estimated by the ECF and Bayesian methods, corresponding to results in Figure 6.

week resulted in greater uncertainty (evident in the upper tails), when compared to constraining doses to once per day. Incorporating the more complex time–activity assumption (scenario 2a) resulted in estimated doses similar to the once-per-day case (results not shown). This may be due to the fact that both scenarios approach steady-state conditions, due to the high frequency of exposures. As the exposures become less frequent, there are more cases of low simulated biomarker levels implying a high exposure, since the biomarkers from any magnitude exposure have likely reduced to low levels by the time a hypothetical sample is taken. As previously noted, inversion becomes more sensitive to uncertainties in biomarker sampling time for less frequent exposures.

Additionally, there was only a slight difference between the results from the ECF and the discretized Bayesian methods. Because of the uncertainties involved, none of these three methods considered can be conclusively termed superior when examining a short half-life biomarker, such as TCPy. As shown in Figure 6, for higher frequency exposures (once a day or seven per week), all the three methods considered showed similar results. For the lower frequency exposures (once a week) case, the steady-state approximation is not valid, while there was only a slight difference between the results from the ECF and the discretized Bayesian methods. Theoretically, the simple Bayesian method should produce more realistic results than the ECF method, since it does not involve the assumption of a linear relationship of biomarker to dose. The Bayesian method can be further improved by including better prior information of population doses or activities. However, this was not possible for the NHEXAS-MD data used here, due to the apparent non-specificity of the TCPy biomarker to low-dose CPF exposure, and the difficulty in adequately characterizing exposure time profile and biomarker collection times.

Value of Additional Information From the preceding, it is clear that more detailed information on exposure parameters reduces uncertainty in back-calculating doses from biomarker information. Many "data-intensive" methods to collect more data are relatively impractical for large-scale studies (e.g. urine collection through an entire day, blood collection, and longitudinal dietary and multimedia concentration measure ments). While more detailed dietary recall and exposure information has previously been shown to aid in exposure assessment and biomarker interpretation for chemicals such as CPF (Wilson et al., 2003; Meeker et al., 2005; Morgan et al., 2005), it is likely that the short biomarker half-life prohibits practical reconstruction of exposures via toxicokinetic modeling inversion for large population studies. However, sub-population studies that collect detailed measurements can produce complementary data needed to reduce uncertainties in exposure reconstruction.

For example, in the case of a long half-life chemical such as methylmercury, it has been illustrated that by merging fish consumption surveys, dietary intake models, and national fish methylmercury data, reasonable agreement between biomarkers and data can be reached (Carrington and Bolger, 2002).

Case study 2: CPF dose reconstruction from "synthetically augmented" NHEXAS-MD data

This case study presents an evaluation of inversion methods using "augmented" biomarker data, so as to assess the uncertainties associated with gaps in actual data. The data set used is thus a "synthetically augmented" version of the NHEXAS-MD database that was developed by filling in missing information via randomly sampling distributions derived from estimates available in the literature (e.g. Pang et al., 2002). This data set, referred here also as "synthetic data," was then used to evaluate approaches for estimating population parameters. The advantage of this approach is that since all the relevant parameter values were known for the new data set, the performance of various inversion methods could be directly assessed. Table 4 presents the distributions used for filling in missing information and to synthetically augment the biomarker data.

It was assumed there are only two routes of CPF exposure: (a) continuous background inhalation exposure, and (b) bolus doses due to dietary ingestion. It was also assumed that dietary CPF exposure was directly correlated with dietary TCPy exposure, based on studies showing that TCPy may exist in food at levels higher than CPF (Wilson et al., 2003; Morgan et al., 2005; NRC, 2006). For simplicity, a 10:1 ratio of TCPy:CPF dietary exposure was initially assumed for the generation of synthetic data.

While the synthetic dietary and background exposures were based on the summary exposure distributions from Pang et al. (2002), they were not reassigned to specific individuals as the corresponding data on intake amounts are not available. For the purpose of this case study, the impact of this assignment is not significant, as the aim is simply to obtain augmented data that are similar to NHEXAS-MD, but with the ability to specify all exposure-relevant parameters for each individual. Only the total doses absorbed are considered because the absorption fraction and intake amount are coupled, and, hence, individually unidentifiable.

The assignment of synthetic data for each of the 339 entries in the NHEXAS-MD database was performed through the following procedure:

- Individuals were randomly drawn from the NHEXAS-MD population (relevant physiological parameters were age, gender, body weight, and body height).
- Parameters, such as daily liquid urination rate, daily creatinine elimination rate, and inhalation rate, were calculated based on physiological parameter distributions for populations from available sources such as

Table 4. Ranges of supporting exposure parameters used in augmenting the NHEXAS-MD data.

Parameter (and distribution)	Values/ranges	Description
Meal time ranges (U)	7–9 (breakfast); 11–14 (lunch); 17.5–20.5 (dinner)	Hours of the day a meal may be consumed
Sample time ranges (U)	6.5–8	Hour of the day a sample is collected
Prior urination time (U)	21–24	Hour of previous day the bladder is emptied, before biomarker collection
Meal probabilities (C)	0.3 (breakfast); 0.3 (lunch); 0.4 (dinner)	Probability of CPF being in a given meal
Background CPF concentrations (CDF)	From Pang et al. (2002)	Percentiles of background CPF concentrations
Daily ingested CPF (dietary) (CDF)	From Pang et al. (2002)	Percentiles of daily CPF intake dose
Ratio of TCPy to CPF in diet (CDF)	10:1 assumed for simplicity	Amount of TCPy ingested (synthetic)
Biomarker	PBTK model output	PBTK model run with corresponding sampled inputs

NHANES, the data of the International Commission on Radiological Protection (ICRP — ICRP, 2003), and the Physiological Parameters for PBPK Modeling (P3M — Lifeline Group, 2004) database/model.

- An average daily background CPF dose, average daily dietary CPF dose, and average daily TCPy dose (10 times the CPF dose) were drawn randomly from distributions available in Pang et al. (2002), and randomly assigned to individuals of the population.
- A week-long exposure profile was generated, randomly assigning bolus doses to three meals per day. Each day was a separate random assignment of both meal times, and distribution of CPF dose among the three meals (consumption of CPF during one, two, or three meals was possible, with uneven distribution). The problem was constrained so that the total bolus weekly dose was consistent with the individual's assigned ADD.
- The week-long exposure profile was used as an input to the CPF/TCPy PK model, and repeated until a quasi-steady state was reached (a little over a month).
- A biomarker collection day was randomly chosen within the quasi-steady-state region. A biomarker collection time was simulated to occur some time in the morning before the first meal of the day.
- The biomarker simulated was the total amount of TCPy which had accumulated in the bladder since the previous urinary excretion time (dependent on the sample time and prior urination time). This is assumed to be the total urinary void, even though a residual volume of urine remains from the previous void that mixes in with the next quantity of urine that the kidney sends to the bladder.

Results for Case Study 2

Different inversion approaches were applied to the "synthetically augmented" data, with varying degrees of augmentation being introduced. It was assumed that the population-level exposure parameters are known to the same degree as the initial biomarker generation (i.e. the inversion method utilized the same probability distributions for meal times and sample times for the population). The physiological parameters were drawn from the adult NHANES population. The inversion was carried out employing different assumptions: (a) assuming no TCPy exposure; (b) assuming a 10:1 TCPy/CPF bolus dose exposure ratio (consistent with the approach used in the generation scenario), (c) assuming no background CPF exposure, and (d) assuming increased randomness/uncertainty in PK parameters (i.e. PK parameters and meal time parameters). As in case study 1, the Bayesian approach used 40,000 model simulations, while the ECF used 5000 simulations. Neglecting background CPF exposure, or increasing PK parameter or exposure uncertainty resulted in only minor differences in predicted doses (not shown). This is likely due to an overall high frequency of bolus doses, and a quasi-steady state is reached in biomarker measurement.

Analyzing the synthetic biomarker data with different constraints on bolus dose affected the results in a manner similar to Case Study 1. Constraining doses to once per day gave results nearly identical to the steady-state and "base-case" assumptions, and constraining doses to once per week gave a wider distribution (results are not shown).

As shown in Figures 8 and 9, the ECF and the simplified Bayesian techniques do not adequately reconstruct the exposures, despite using the same sampling time assumptions. This limitation arises from the fact that these methods, as used currently, cannot utilize the correlations between the exposure and supporting biomarker data at the "actual" individual level. Instead, in the current usage of these methods, information on the exposure–biomarker relationship for specific individuals is "lumped" to obtain statistical distributions for the population, thus "losing information" in the process. While population-level probability distributions in exposure, sampling, and TCPy exposure remained the same for both forward and inverse modeling, the inability to match individual-level

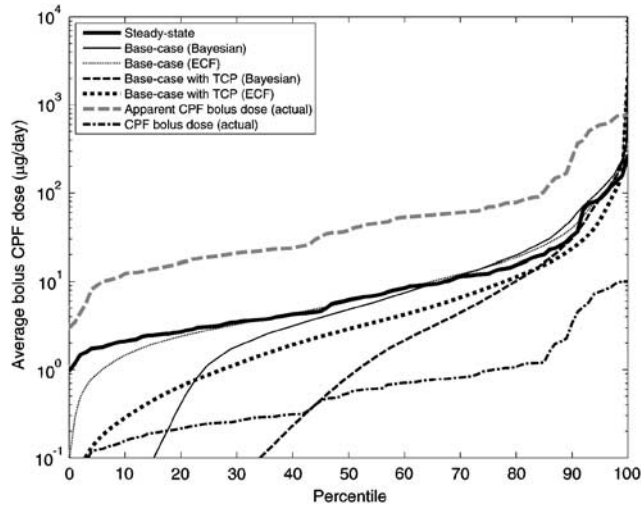

Figure 8. Cumulative distribution functions of dietary CPF uptakes for the synthetic population estimated by the ECF and Bayesian methods, using different dose assumptions. Results only show the synthetic individuals with detectible biomarker levels (which comprised 42% of the samples). ''Apparent'' CPF dose denotes the sum of the actual CPF and TCPy bolus doses (corrected for molecular weight), which may be misinterpreted as a CPF dose if TCPy exposure is neglected. ''Base-case'' results denote those results obtained using the random time–activity assumption with three potential CPF doses per day, 0–7 days per week.

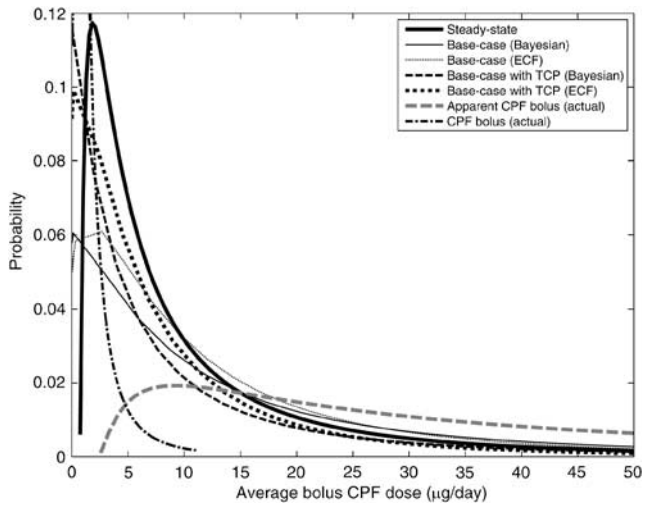

Figure 9. Probability density functions of dietary CPF uptakes for the synthetic population estimated by the ECF and Bayesian methods, using different dose assumptions, corresponding to results in Figure 8.

exposure information prevents a more accurate analysis of the biomarker data. This is evident in Figures 8 and 9 where, despite neglecting or incorporating TCPy exposure, predicted dose levels remain between the levels of apparent and actual CPF doses.

Discussion

A comparison of general characteristics of the NHEXAS and NHANES databases *vis a vis* exposure and dose reconstruction is shown in Table 5. Although the NHEXAS data set provides a significant amount of more detailed supporting or complementary exposure-related information, when compared to NHANES, this information is still not adequate for detailed reconstruction of exposures under the various types of conditions considered in the case studies. The analysis presented here provides a starting point for developing improved designs for future biomonitoring studies from the perspective of exposure reconstruction. Furthermore, the analysis here identifies specific limitations in existing exposure reconstruction methods that can be applied to population biomarker data, and suggests potential approaches for addressing exposure reconstruction from population biomarker data based on the supporting or complementary exposure data available. Such information needs to be incorporated in the development of future biomonitoring study designs. For example, simply recording the last urinary void time along

Table 5. Comparison of NHEXAS and NHANES data sets *vis a vis* needs for detailed exposure reconstruction. "X" represents available, "–" represents not available, and "o" represents available for some chemicals.

Attribute	NHANES	NHEXAS	Requirement*
Baseline parameters (individual characteristics)			
Smoking and tobacco use	X	X	Chemical dependent
Dietary recall	X	X	Required
Pesticide use	X	X	Chemical dependent
Demographic background/occupation	X	X	Dependent on model complexity and biomarker
Recent activity diary	—	X	Dependent on model complexity
Housing characteristics	X	X	Dependent on model complexity
Physiological characteristics (at individual level)			
Age	X	X	Required
Gender	X	X	Required
Body weight	X	X	Required
Body height	X	X	Required
Cardiac output	—	—	Required (population distributions)
Urinary excretion rate	—	—	Required for urinary biomarkers
Urinary creatinine excretion rate	—	—	Required
Environmental concentrations			
Food residues	—	X	Required
Personal air	o (VOCs)	X	Required
Indoor/outdoor air	—	X	Required
Dust	o (lead)	X	Required
Soil	—	X	Required
Tap water	—	X	Required
Behavioral/health			
Diet behavior and nutrition	X	X	Dependent on the chemical of interest
Baseline health condition	X	—	Dependent on the biomarker
Biochemical parameters	—	—	Required (from other studies)
Biomarker data characteristics			
Actual sampling time	—	—	For short half-life chemicals
Time of last void	—	—	Required for urinary biomarkers
Time of last meal before void	—	—	Urinary biomarkers
Multiple biomarkers	o	o	Chemical and pathway dependent

with urinary biomarker samples can significantly aid in the interpretation of the urinary biomarker data.

One component of the problem of reconstructing exposures from biomarkers can be addressed through the development or use of more sophisticated numerical inversion techniques. The techniques used in the evaluation presented in this study (i.e. the ECF and the discrete Bayesian approach) are not fully capable of utilizing all available biomarker and supporting exposure-related data (e.g. data for specific individuals). For example, when partial individual-level exposure information is available, the data can be potentially used in reconstructing population exposures. However, such data cannot be used in reconstructing exposures at an individual level. This implies that when sparse, individual level exposure data are available, the traditional techniques such as MLE cannot be applied (as the data are too sparse), whereas, on the other hand, methods such as ECF and the discrete Bayesian approaches cannot utilize all the available information. Clearly there is a need to improve inversion approaches for the purpose of characterizing real world exposures. One major goal for a useful exposure reconstruction framework should be the development or identification of methods that can fully utilize "incomplete" exposure-related data in exposure reconstruction.

Two other areas for potential improvement in exposure reconstruction involve reconstruction of route- or pathway-specific exposures (i.e. simultaneous reconstruction/estimation of multiple inputs/parameters), and the ability to provide a gradual convergence process. The simplified approaches evaluated here, ECF and discretized Bayesian, require performing a number of model simulations at a time, and then estimating the exposures. Successive refinement of exposure estimates, therefore, requires running a pre-set number of simulations. Several time-marching techniques such as the Bayesian MCMC can be used to achieve a more gradual iterative process. The methods evaluated here are applicable in their current form to the reconstruction/

estimation of a single input/parameter. Furthermore, the discretized Bayesian method requires a large set of simulations at a fixed discretization to assure that all the discretized ''bins'' contain adequate number of samples for reconstruction. Refinement of resolution, therefore, requires a substantially large number of additional simulations. So, development or use of new methods that utilize a ''time-stepping'' type of algorithm (e.g. similar to the Bayesian MCMC approach) will clearly aid the exposure reconstruction process.

There is also a need for ''optimal'' exposure reconstruction using PBTK models, based on the rationale that incorporating optimization approaches in the inverse modeling process can result in faster convergence and more robust solutions. Though typical PBTK/PK models can be run quickly on modern computers (of the order of seconds to tens of seconds per simulation), the computational demands can become challenging when hundreds of thousands of simulations are used in the inverse modeling on desktop computers. Likewise, complex PBTK modeling scenarios (e.g. mixtures of metals or pesticides with large half-lives) may need significantly more computational time for a single simulation. Thus, use of fast equivalent operational models (FEOMs) (e.g. see Li et al., 2002; Balakrishnan et al., 2003; Wang et al., 2005; and Table 3) may be necessary to achieve reasonable computational performance. Figure 10 presents the conceptual framework depicting the steps involved in using optimization techniques in conjunction with inverse modeling for faster convergence that incorporates the use of FEOMs for faster simulation times. The framework shown in Figures 1 and 10 is not limited to exposure reconstruction. It can also be used for estimating distributions of physiological and biochemical PBTK model parameters for individuals and populations that are consistent with available biomarker data (typically study-specific data where exposures are adequately characterized) by combining the data with prior estimates of the parameters. Furthermore, the framework can be used to select appropriate PBTK model structures when alternative formulations are available.

One of the issues not considered in this analysis was the uncertainty associated with genetic polymorphisms. A genetic polymorphism affecting the metabolism of an intermediate metabolite (CPF-oxon) exists in the population, and has been previously incorporated into a CPF PBTK model (Timchalk et al., 2002b). However, the metabolism of CPF-oxon to TCPy is rapid in relation to overall TCPy formation and elimination. The polymorphism alters toxicodynamics, but does not significantly alter the TCPy biomarker. For chemicals besides CPF, genetic polymorphisms may, however, affect the biomarker level distributions within the populations considered (e.g. arsenic metabolites and the AS3MT polymorphism; Thomas et al., 2007). For short half-life chemicals, the uncertainties in exposure and biomarker assumptions overshadow any effect that polymorphisms may have. For long half-life chemicals, where the biomarker levels are at a quasi-steady state, it may be more practical to analyze the effect of polymorphisms in the context of dose reconstruction. Table 6 lists examples of major underlying biochemical and genetic factors that can contribute to the variability in the metabolism and biological transport of organophosphates, VOCs, and metals in humans and animals. When sufficient biomarker and exposure data are available, numerical inversion exposure methods can also be applied in a hierarchical manner to

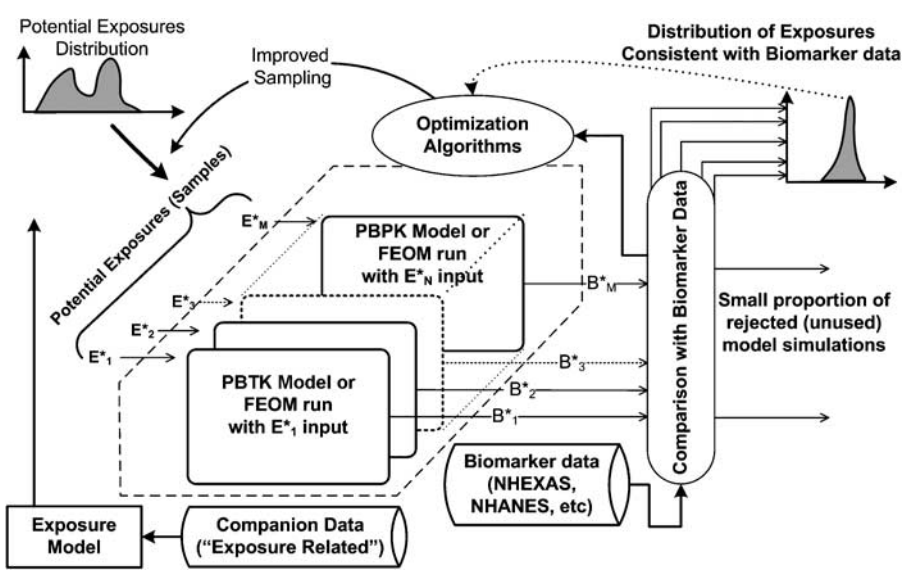

Figure 10. Exposure reconstruction process using optimization-aided approach with the original PBTK model or fast equivalent operational models (FEOMs). The coupling with optimization techniques reduces the number of simulations significantly, and the use of FEOMs reduces the time required for each run.

Table 6. Selected information available on underlying mechanisms for inter-individual variability due to genetic factors that affect the transport and transformation of organophosphates, VOCs, and metals in humans and animals.

Species class	Metabolism and transport	Notes
Organophosphates	Liver microsomes, CYP-1A2, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4, and 3A5 (Hodgson, 2003)	Organophosphates share P450s for metabolism (Tang et al., 2001; Rose and Hodgson, 2005). Chlorpyrifos-oxonase (PON1) polymorphism is incorporated into chlorpyrifos PBPK model (Timchalk et al., 2002a)
Metals/metalloids	Glutathione, cystein (Quig, 1998; Patrick, 2003), arsenic methyltransferase (arsenic) (Drobna et al., 2004; Meza et al., 2005; Wood et al., 2006; Thomas et al., 2007), metallothionein (Nordberg and Nordberg, 2000), divalent metal transporter-1 (DMT1) (Garrick et al., 2003; Bressler et al., 2004), bone mineral exchange (lead)	Interactions with essential elements (Goyer, 1997; Ballatori, 2002). Arsenic metabolism affected by CYT19, hNP, and hGSTO1-1 genotypes (Drobna et al., 2004; Meza et al., 2005). Cadmium and lead can alter P450 levels and activities (Moore, 2004; Baker et al., 2005). Metal speciation significantly alters toxicity (Yokel et al., 2006)
VOCs	Glutathione, glutathione-*S*-transferase, CYP-2E1, 2C6, 2C11, 1A1, and 1A2 (Reddy et al., 2005)	GST genotypes effect metabolism and toxicity (Haber et al., 2002; Qu et al., 2003; Dirksen et al., 2004; Silva Mdo et al., 2004; Reddy et al., 2005)

estimate parameters for populations of interest. The results of such analysis would allow one to elicit differences in the toxicokinetic characteristics due to genetic polymorphisms. However, this is a significantly data-intensive exercise, and the availability of databases from population studies that explicitly incorporate genetic variability is presently very limited. Likewise, when the impacts of specific genetic polymorphisms are known, data on these impacts could provide relevant information for parameterizing the underlying PBTK models to improve the exposure reconstruction process.

Conclusion

The analyses and discussion presented in this study provide a foundation for developing new and improved approaches for exposure reconstruction. Clearly, there is a strong incentive to implement more biomonitoring tools into the field of exposure science. However, it must be recognized that exposure reconstruction requires not only a sound set of information from monitoring tools but also a well-defined set of modeling tools that can be used within an integrative analytical framework to reduce uncertainties about the origins of exposure. Thus, exposure measurement and modeling professionals must work together to improve the accuracy of reconstruction methods and applications. This is essential for risk assessment and, more importantly, risk management.

The case studies presented here highlighted the gaps in existing biomonitoring studies with respect to supplemental or complementary data needed on exposure, contaminant sources, and human activities. These gaps need to be seriously considered when developing improved designs for future biomonitoring studies if there is a desire to complete realistic exposure modeling simulations. The analyses here identified specific limitations in existing exposure reconstruction methods, which have been applied to population biomarker data, and suggested potential approaches for addressing exposure reconstruction from population biomarker data based on the availability of supporting exposure data. Included are state-of-the-art numerical inversion techniques, customized approaches for reconstruction of route- or pathway-specific exposures, and optimization tools for more effective use of PBTK models.

Acknowledgements

The United States Environmental Protection Agency (USEPA), through its Office of Research and Development (ORD), partially funded and collaborated in the research described here under University Partnership Agreement CR 83162501 to the Center for Exposure and Risk Modeling (CERM) of the Environmental and Occupational Health Sciences Institute. The research and this manuscript have been subjected to Agency review and approved for publication. USEPA has also supported this work through the Environmental Bioinformatics and Computational Toxicology Center (ebCTC – GAD R 832721-010). Additional support has been provided by the NIEHS sponsored UMDNJ Center for Environmental Exposures and Disease, Grant#: NIEHS P30ES005022. We acknowledge feedback and suggestions of numerous USEPA collaborators including C. Dary, R. Tornero-Velez, M. Morgan, M. Dellarco, F. Power, J.N. Blancato, and L. Sheldon.

References

Ablonczy Z., Lukacs A., and Papp E. Application of the maximum entropy method to absorption kinetic rate processes. *Biophys Chem* 2003: 104(1): 249–258.

Abramov S., Adamovich A., Inyukhin A., Moskovsky A., Roganov V., Shevchuk E., Shevchuk Y., and Vodomerov A. OpenTS: an outline of dynamic parallelization approach. *Parallel Comput Technol* 2005: 3606: 303–312.

ACM. *Software from the Transactions on Mathematical Software (TOMS)* 2006: http://www.netlib.org/toms/.

Albertini R., Bird M., Doerrer N., Needham L., Robison S., Sheldon L., and Zenick H. The use of biomonitoring data in exposure and human health risk assessments. *Environ Health Perspect* 2006: 114(11): 1755–1762.

Allen B.C., Hack C.E., and Clewell H.J. Use of Markov chain Monte Carlo analysis with a physiologically-based pharmacokinetic model of methylmercury to estimate exposures in US women of childbearing age. *Risk Anal* 2007: 27(4): 947–959.

Anderson E. *LAPACK User's Guide. Software, Environments, Tools*. Society for Industrial and Applied Mathematics, Philadelphia, 1999: xxi: 407 p.

Armbruster B., Hamilton R.J., and Kuehl A.K. Spectrum reconstruction from dose measurements as a linear inverse problem. *Phys Med Biol* 2004: 49(22): 5087–5099.

Aster R.C., Thurber C.H., and Borchers B. *Parameter Estimation and Inverse Problems. International Geophysics Series. v. 90*. Elsevier Academic Press, Amsterdam, Boston, 2005: xii: 301 p.

ATSDR. *Toxicological Profile for Chlorpyrifos*. Agency for Toxic Substances and Disease Registry, Division of Toxicology, Atlanta, GA, 1997.

Bai J.P.F., Utis A., Crippen G., He H.D., Fischer V., Tullman R., Yin H.Q., Hsu C.P., Jiang L., and Hwang K.K. Use of classification regression tree in predicting oral absorption in humans. *J Chem Inf Comput Sci* 2004: 44(6): 2061–2069.

Baker J.R., Edwards R.J., Lasker J.M., Moore M.R., and Satarug S. Renal and hepatic accumulation of cadmium and lead in the expression of CYP4F2 and CYP2E1. *Toxicol Lett* 2005: 159(2): 182–191.

Balakin K.V., Ivanenkov Y.A., Savchuk N.P., Ivashchenko A.A., and Ekins S. Comprehensive computational assessment of ADME properties using mapping techniques. *Curr Drug Discov Technol* 2005: 2(2): 99–113.

Balakrishnan S., Roy A., Ierapetritou M.G., Flach G.P., and Georgopoulos P.G. Uncertainty reduction and characterization for complex environmental fate and transport models: an empirical Bayesian framework incorporating the stochastic response surface method. *Water Resour Res* 2003: 39(12): 1350.

Ballatori N. Transport of toxic metals by molecular mimicry. *Environ Health Perspect* 2002: 110(Suppl. 5): 689–694.

Barr D.B., and Angerer J. Potential uses of biomonitoring data: a case study using the organophosphorus pesticides chlorpyrifos and malathion. *Environ Health Perspect* 2006: 114(11): 1763–1769.

Barr D.B., Wilder L.C., Caudill S.P., Gonzalez A.J., Needham L.L., and Pirkle J.L. Urinary creatinine concentrations in the U.S. population: implications for urinary biologic monitoring measurements.. *Environ Health Perspect* 2005: 113(2): 192–200.

Beal S.L. Commentary on significance levels for covariate effects in NONMEM. *J Pharmacokinet Pharmacodyn* 2002: 29(4): 403–410.

Bender R., and Heinemann L. Fitting nonlinear-regression models with correlated errors to individual pharmacodynamic data using SAS software. *J Pharmacokinet Biopharm* 1995: 23(1): 87–100.

Bischof C., Khademi P., Mauer A., and Carle A. ADIFOR 2.0: automatic differentiation of Fortran 77 programs. *IEEE Comput Sci Eng* 1996: 3(3): 18–32.

Bischof C.H., Roh L., and Mauer-oats A.J. ADIC: an extensible automatic differentiation tool for ANSI-C. *Software Pract Ex* 1997: 27(12): 1427–1456.

Blazewicz J., Lukasiak P., and Milostan M. Application of tabu search strategy for finding low energy structure of protein. *Artif Intell Med* 2005: 35(1–2): 135–145.

Bois F., Maszle D., Revzan K., Tillier S., and Yuan Z. *MCSim User Manual* 2007: http://fredomatic.free.fr/mcsim.html.

Bois F.Y. Statistical analysis of Fisher et al. PBPK model of trichloroethylene kinetics. *Environ Health Perspect* 2000: 108: 275–282.

Bourne D.W. BOOMER, a simulation and modeling program for pharmacokinetic and pharmacodynamic data analysis. *Comput Methods Programs Biomed* 1989: 29(3): 191–195.

Breiman L. *Classification and Regression Trees. The Wadsworth Statistics/Probability Series*. Wadsworth International Group, Belmont, CA, 1984: x: 358 p.

Bressler J.P., Olivi L., Cheong J.H., Kim Y., and Bannona D. Divalent metal transporter 1 in lead and cadmium transport. *Ann NY Acad Sci* 2004: 1012: 142–152.

Carrington C.D., and Bolger M.P. An exposure assessment for methylmercury from seafood for consumers in the United States. *Risk Anal* 2002: 22(4): 689–699.

Castorina R., Bradman A., McKone T.E., Barr D.B., Harnly M.E., and Eskenazi B. Cumulative organophosphate pesticide exposure and risk assessment among pregnant women living in an agricultural community: a case study from the CHAMACOS cohort. *Environ Health Perspect* 2003: 111(13): 1640–1648.

CDC. *Third National Report on Human Exposure to Environmental Chemicals*. NCEH Pub. no. 05-0570, Centers for Disease Control and Prevention, Atlanta, GA, 2005: http://www.cdc.gov/exposurereport/.

Cherkassky V., and Gehring D. Comparison of adaptive methods for function estimation from samples. *IEEE Trans Neural Netw* 1996: 7(4): 969–984.

Chow H.H., Tolle K.M., Roe D.J., Elsberry V., and Chen H.C. Application of neural networks to population pharmacokinetic data analysis. *J Pharm Sci* 1997: 86(7): 840–845.

Christakos G., Bogaert P., and Serre M.L. *Temporal GIS: Advanced Functions for Field-Based Applications*. Springer, Berlin, New York, 2001: xii: 217 p.

Clarke C.J.S., and Janday B.S. The solution of the biomagnetic inverse problem by maximum statistical entropy. *Inverse Probl* 1989: 5(4): 483.

Coello C.A.C., and Pulido G.T. Multiobjective structural optimization using a microgenetic algorithm. *Structural and Multidisciplinary Optimization* 2005: 30(5): 388–403.

Conway G.R., Glass N.R., and Wilcox J.C. Fitting nonlinear models to biological data by Marquardt's algorithm. *Ecology* 1970: 51(3): 503.

Covington T.R., Gentry P.R., Van Landingham C.B., Andersen M.E., Kester J.E., and Clewell H.J. The use of Markov chain Monte Carlo uncertainty analysis to support a public health goal for perchloroethylene. *Regul Toxicol Pharmacol* 2007: 47(1): 1–18.

CRAN. *CRAN — The Comprehensive R Archive Network* 2006: (http://cran.r-project.org/).

Deconinck E., Xu Q.S., Put R., Coomans D., Massart D.L., and Vander Heyden Y. Prediction of gastro-intestinal absorption using multivariate adaptive regression splines. *J Pharm Biomed Anal* 2005: 39(5): 1021–1030.

Denison D.G.T., Mallick B.K., and Smith A.F.M. Bayesian MARS. *Stat Comput* 1998: 8(4): 337–346.

Dennis J., Gay D., and Walsh R. An adaptive nonlinear least-squares algorithm. *ACM Trans Math Softw* 1981: 7(3): 348–368.

Dirksen U., Moghadam K.A., Mambetova C., Esser C., Fuhrer M., and Burdach S. Glutathione *S* transferase theta 1 gene (GSTT1) null genotype is associated with an increased risk for acquired aplastic anemia in children. *Pediatr Res* 2004: 55(3): 466–471.

Dokoumetzidis A., and Aarons L. Propagation of population pharmacokinetic information using a Bayesian approach: comparison with meta-analysis. *J Pharmacokinet Pharmacodyn* 2005: 32(3–4): 401–418.

Drobna Z., Waters S.B., Walton F.S., LeCluyse E.L., Thomas D.J., and Styblo M. Interindividual variation in the metabolism of arsenic in cultured primary human hepatocytes. *Toxicol Appl Pharmacol* 2004: 201(2): 166–177.

Egeghy P.P., Quackenboss J.J., Catlin S., and Ryan P.B. Determinants of temporal variability in NHEXAS-Maryland environmental concentrations, exposures, and biomarkers. *J Expo Anal Environ Epidemiol* 2005: 15(5): 388–397.

Elzo M.A. Restricted maximum likelihood procedures for the estimation of additive and nonadditive genetic variances and covariances in multibreed populations. *J Anim Sci* 1994: 72(12): 3055–3065.

Engl H.W., Hanke M., and Neubauer A. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, Boston, 1996: viii: 321 p.

Farquharson C.G., and Oldenburg D.W. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophys J Int* 2004: 156(3): 411–425.

Flouris A., and Duffy J. Applications of artificial intelligence systems in the analysis of epidemiological data. *Eur J Epidemiol* 2006: 21(3): 167.

Friedman J.H., and Meulman J.J. Multiple additive regression trees with application in epidemiology. *Stat Med* 2003: 22(9): 1365–1381.

Friedman J.H., and Roosen C.B. An introduction to multivariate adaptive regression splines. *Stat Methods Med Res* 1995: 4(3): 197–217.

Friedman J.H., and Stuetzle W. Projection pursuit regression. *J Am Stat Assoc* 1981: 76(376): 817.

Gamboa F., and Gassiat E. Bayesian methods and maximum entropy for ill-posed inverse problems. *Ann Stat* 1997: 25(1): 328–350.

Gamerman D., and Lopes H.F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Taylor & Francis, Boca Raton, 2006: xvii: 323 p.

Garrick M.D., Dolan K.G., Horbinski C., Ghio A.J., Higgins D., Porubcin M., Moore E.G., Hainsworth L.N., Umbreit J.N., Conrad M.E., Feng L., Lis A., Roth J.A., Singleton S., and Garrick L.M. DMT1: a mammalian transporter for multiple metals. *Biometals* 2003: 16(1): 41–54.

Gelman A. *Bayesian Data Analysis. Texts in Statistical Science*. Chapman & Hall/CRC, Boca Raton, FL, 2004: xxv: 668 p.

Georgopoulos P.G., Balakrishnan S., Roy A., Isukapalli S., Sasso A., Chien Y.C., and Weisel C.P. *A Comparison of Maximum Likelihood Estimation Methods for Inverse Problem Solutions Employing PBPK Modeling with Biomarker Data: Application to Tetrachloroethylene*. CCL:2008.01, Computational Chemodynamics Laboratory, EOHSI, Piscataway, NJ, 2008: http://ccl.rutgers.edu/reports/MENTOR3P/CCL_2008-report_Bayesian.pdf.

Georgopoulos P.G., and Lioy P.J. From theoretical aspects of human exposure and dose assessment to computational model implementation: The MOdeling ENvironment for TOtal Risk studies (MENTOR). *J Toxicol Environ Health B Crit Rev* 2006: 9(6): 457–483.

Georgopoulos P.G., Roy A., and Gallo M.A. Reconstruction of short-term multiroute exposure to volatile organic-compounds using physiologically-based pharmacokinetic models. *J Expo Anal Environ Epidemiol* 1994: 4(3): 309–328.

Gilks W.R., Richardson S., and Spiegelhalter D.J. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, Boca Raton, FL, 1998: xvii: 486 p.

Glover F. Tabu search for nonlinear and parametric optimization (with links to genetic algorithms). *Discrete Appl Math* 1994: 49(1–3): 231–255.

Glover F., Kelly J.P., and Laguna M. Genetic algorithms and tabu search — hybrids for optimization. *Comput Oper Res* 1995: 22(1): 111–134.

Gobburu J.V.S., and Chen E.P. Artificial neural networks as a novel approach to integrated pharmacokinetic–pharmacodynamic analysis. *J Pharm Sci* 1996: 85(5): 505–510.

Gosselin N.H., Brunet R.C., Carrier G., Bouchard M., and Feeley M. Reconstruction of methylmercury intakes in indigenous populations from biomarker data. *J Expo Sci Environ Epidemiol* 2006: 16(1): 19–29.

Goyer R.A. Toxic and essential metal interactions. *Annu Rev Nutr* 1997: 17: 37–50.

Haber L.T., Maier A., Gentry P.R., Clewell H.J., and Dourson M.L. Genetic polymorphisms in assessing interindividual variability in delivered dose. *Regul Toxicol Pharmacol* 2002: 35(2 Pt 1): 177–197.

Hansen P.C. Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems. *Numer Algorithms* 1994: 6(1–2): 1–35.

Hashemi R.R., and Young J.F. The prediction of methylmercury elimination half-life in humans using animal data: a neural network/rough sets analysis. *J Toxicol Environ Health A* 2003: 66(23): 2227–2252.

Hastie T., Tibshirani R., and Friedman J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction: with 200 Full-Color Illustrations*. Springer series in statistics. Springer, New York, 2001: xvi: 533 p.

Hays S.M., Becker R.A., Leung H.W., Aylward L.L., and Pyatt D.W. Biomonitoring equivalents: a screening approach for interpreting biomonitoring results from a public health risk perspective. *Regul Toxicol Pharmacol* 2007: 47(1): 96–109.

Hill Jr R.H., Head S.L., Baker S., Gregg M., Shealy D.B., Bailey S.L., Williams C.C., Sampson E.J., and Needham L.L. Pesticide residues in urine of adults living in the United States: reference range concentrations. *Environ Res* 1995: 71(2): 99–108.

Hodgson E. *In vitro* human phase I metabolism of xenobiotics I: pesticides and related compounds used in agriculture and public health, May 2003. *J Biochem Mol Toxicol* 2003: 17(4): 201–206.

Holmes C.C., and Denison D.G.T. Classification with Bayesian MARS. *Machine Learn* 2003: 50(1–2): 159–173.

Holmes S.L., Ward R.C., Galambos J.D., and Strickler D.J. A method for optimization of pharmacokinetic models. *Toxicol Methods* 2000: 10(1): 41–53.

Hovland P., Norris B., and Smith B. Making automatic differentiation truly automatic: coupling PETSc with ADIC. *Future Generat Comput Syst* 2005: 21(8): 1426–1438.

Hristov D., Stavrev P., Sham E., and Fallone B.G. On the implementation of dose-volume objectives in gradient algorithms for inverse treatment planning. *Med Phys* 2002: 29(5): 848–856.

ICRP. International Commission on Radiological Protection, 2003. http://www.icrp.org.

Isukapalli S.S., Lioy P.J., and Georgopoulos P.G. Mechanistic modeling of emergency events: assessing impact of hypothetical releases of anthrax. *Risk Anal* 2008 (in press).

Isukapalli S.S., Roy A., and Georgopoulos P.G. Efficient sensitivity/uncertainty analysis using the combined stochastic response surface method and automated differentiation: application to environmental and biological systems. *Risk Anal* 2000: 20(5): 591–602.

Isukapalli S.S., Roy A., and Georgopoulos P.G. Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems. *Risk Anal* 1998: 18(3): 351–363.

Jamshidian M. On algorithms for restricted maximum likelihood estimation. *Comput Stat Data Anal* 2004: 45(2): 137–157.

Jitnah N., and Nicholson A. treeNets: a framework for anytime evaluation of belief networks. *Qual Quant Pract Reason* 1997: 1244: 350–364.

Johansen T.A., and Foss B.A. ORBIT — operating-regime-based modeling and identification toolkit. *Control Eng Pract* 1998: 6(10): 1277–1286.

Kaltenbach M.L., and Vistelle R. KINFIT — a non-linear least-squares computer-program for the estimation of pharmacokinetic parameters after intravenous administration. *Anticancer Res* 1994: 14(6A): 2375–2377.

Kazarlis S.A., Papadakis S.E., Theocharis J.B., and Petridis V. Microgenetic algorithms as generalized hill-climbing operators for GA optimization. *IEEE Trans Evol Comput* 2001: 5(3): 204–217.

Kharche R.V., and Forth S.A. Source transformation for MATLAB automatic differentiation. *Computational Science — ICCS 2006, Pt 4, Proceedings* 2006, Springer, Berlin, Vol. 3994, pp. 558–565.

LeBlanc M., Moon J., and Kooperberg C. Extreme regression. *Biostatistics* 2006: 7(1): 71–84.

Li G.Y., Wang S.W., Rabitz H., Wang S.Y., and Jaffe P. Global uncertainty assessments by high dimensional model representations (HDMR). *Chem Eng Sci* 2002: 57(21): 4445–4460.

Lifeline Group. Physiological Parameters for PBPK Modeling Version 1.2 (P3M), 2004. http://www.thelifelinegroup.org/P3M.

Lingjaerde O.C., and Liestol K. Generalized projection pursuit regression. *SIAM J Sci Comput* 1998: 20(3): 844–857.

Liu J., and Li M. Finding cancer biomarkers from mass spectrometry data by decision lists. *J Comput Biol* 2005: 12(7): 971–979.

Lunn D.J., Best N., Thomas A., Wakefield J., and Spiegelhalter D. Bayesian analysis of population PK/PD models: general concepts and software. *J Pharmacokinet Pharmacodyn* 2002: 29(3): 271–307.

MacIntosh D.L., Kabiru C., Echols S.L., and Ryan P.B. Dietary exposure to chlorpyrifos and levels of 3,5,6-trichloro-2-pyridinol in urine. *J Expo Anal Environ Epidemiol* 2001: 11(4): 279–285.

Mage D.T., Allen R.H., Gondy G., Smith W., Barr D.B., and Needham L.L. Estimating pesticide dose from urinary pesticide concentration data by creatinine correction in the Third National Health and Nutrition Examination Survey (NHANES-III). *J Expo Anal Environ Epidemiol* 2004: 14(6): 457–465.

Marquardt D.W. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Indt Appl Math* 1963: 11(2): 431.

MathWorks Inc.. *Matlab, Version 7*. Prentice Hall, Englewood Cliffs, NJ, 2006: http://www.mathworks.com.

Mathworks. *User's Guide to Genetic Algorithms and Direct Search Toolbox for Matlab, Version 2.0* 2006a: http://www.mathworks.com.

Mathworks. *User's Guide to Neural Networks Toolbox for Matlab, Version 2.0* 2006b: http://www.mathworks.com.

Mathworks. *User's Guide to Optimization Toolbox for Matlab, Version 3.0* 2006c: http://www.mathworks.com.

Mathworks. *User's Guide to Statistics Toolbox for Matlab, Version 5.0* 2006d: http://www.mathworks.com.

Meeker J.D., Barr D.B., Ryan L., Herrick R.F., Bennett D.H., Bravo R., and Hauser R. Temporal variability of urinary levels of nonpersistent insecticides in adult men. *J Expo Anal Environ Epidemiol* 2005: 15(3): 271–281.

Merkwirth C. *ENTOOL — Ensemble Toolbox for Statistical Learning (Matlab Toolbox)* 2006: http://zti.if.uj.edu.pl/~merkwirth/entool.htm.

Messnarz B., Tilg B., Modre R., Fischer G., and Hanser F. A new spatiotemporal regularization approach for reconstruction of cardiac transmembrane potential patterns. *IEEE Trans Biomed Eng* 2004: 51(2): 273–281.

Meza M.M., Yu L., Rodriguez Y.Y., Guild M., Thompson D., Gandolfi A.J., and Klimecki W.T. Developmentally restricted genetic determinants of human arsenic metabolism: association between urinary methylated arsenic and

CYT19 polymorphisms in children. *Environ Health Perspect* 2005: 113(6): 775–781.

Mohammad-Djafari A., Giovannelli J.F., Demoment G., and Idier J. Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems. *Int J Mass Spectrom* 2002: 215 (1–3): 175–193.

Moles C.G., Mendes P., and Banga J.R. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res* 2003: 13(11): 2467–2474.

Moore M.R. A commentary on the impacts of metals and metalloids in the environment upon the metabolism of drugs and chemicals. *Toxicol Lett* 2004: 148(3): 153–158.

Morales K.H., Ibrahim J.G., Chen C.J., and Ryan L.M. Bayesian model averaging with applications to benchmark dose estimation for arsenic in drinking water. *J Am Stat Assoc* 2006: 101(473): 9–17.

Morgan M.K., Sheldon L.S., Croghan C.W., Jones P.A., Robertson G.L., Chuang J.C., Wilson N.K., and Lyu C.W. Exposures of preschool children to chlorpyrifos and its degradation product 3,5,6-trichloro-2-pyridinol in their everyday environments. *J Expo Anal Environ Epidemiol* 2005: 15(4): 297–309.

Morozov V.A., and Stessin M. *Regularization Methods for Ill-Posed Problems*. CRC Press, Boca Raton, FL, 1993, 257 p.

Moscato P., Mendes A., and Berretta R. Benchmarking a memetic algorithm for ordering microarray data. *Biosystems* 2007: 88(1–2): 56–75.

Muzic R.F., and Christian B.T. Evaluation of objective functions for estimation of kinetic parameters. *Med Phys* 2006: 33(2): 342–353.

Nair S., and Gratzl M. Deconvolution of concentration recordings at live cell preparations via shape error optimization. *Anal Chem* 2005: 77(9): 2875–2881.

Needham L.L., Calafat A.M., and Barr D.B. Uses and issues of biomonitoring. *Int J Hyg Environ Health* 2007: 210(3–4): 229–238.

Nestorov I.S., Hadjitodorov S.T., Petrov I., and Rowland M. Empirical *versus* mechanistic modelling: comparison of an artificial neural network to a mechanistically based model for quantitative structure pharmacokinetic relationships of a homologous series of barbiturates. *AAPS Pharmsci* 1999: 1(4): E17.

Netlib. *Miscellaneous Approximation Algorithms from Netlib* 2006: http://gams.nist.gov/serve.cgi/Package/A/.

NeuroDimension I. *Documentation for NeuroSolutions by NeuroDimension Inc.* 2006.

NIST. *NIST Guide to Available Software: Unconstrained Optimization of a Smooth Multivariate Function, User Provides No Derivatives* 2006: http://gams.nist.gov/serve.cgi/Class/G1b1a/.

Nolan R.J., Rick D.L., Freshour N.L., and Saunders J.H. Chlorpyrifos: pharmacokinetics in human volunteers. *Toxicol Appl Pharmacol* 1984: 73(1): 8–15.

Nordberg M., and Nordberg G.F. Toxicological aspects of metallothionein. *Cell Mol Biol (Noisy-le-grand)* 2000: 46(2): 451–463.

Nott D.J., Kuk A.Y.C., and Duc H. Efficient sampling schemes for Bayesian MARS models with many predictors. *Stat Comput* 2005: 15(2): 93–101.

NRC. *Human Biomonitoring for Environmental Chemicals, Committee on Human Biomonitoring for Environmental Toxicants, Board on Environmental Studies and Toxicology, Division on Earth and Life Studies*. National Research Council, Washington DC, 2006: http://newton.nap.edu/catalog/11700.html.

Opara J., Primozic S., and Cvelbar P. Prediction of pharmacokinetic parameters and the assessment of their variability in bioequivalence studies by artificial neural networks. *Pharm Res* 1999: 16(6): 944–948.

OptTek Systems Inc. OptQuest Engine, 2006. http://www.opttek.com/document.html.

Overgaard R.V., Jonsson N., Tornoe C.W., and Madsen H. Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *J Pharmacokinet Pharmacodyn* 2005: 32(1): 85–107.

Paige C.C. Computing the generalized singular value decomposition. *SIAM J Sci Stat Comput* 1986: 7(4): 1126–1146.

Pang Y.H., MacIntosh D.L., Camann D.E., and Ryan B. Analysis of aggregate exposure to chlorpyrifos in the NHEXAS-Maryland investigation. *Environ Health Perspect* 2002: 110(3): 235–240.

Patrick L. Toxic metals and antioxidants: Part II. The role of antioxidants in arsenic and cadmium toxicity. *Altern Med Rev* 2003: 8(2): 106–128.

Pirkle J.L., Sampson E.J., Needham L.L., Patterson D.G., and Ashley D.L. Using biological monitoring to assess human exposure to priority toxicants. *Environ Health Perspect* 1995: 103(Suppl. 3): 45–48.

Pohlheim H. GEATbx: Genetic and Evolutionary Algorithm Toolbox for Use with MATLAB Documentation, 2005. http://www.geatbx.com/docu.

Press W.H. *Numerical Recipes Code CD-ROM*. Cambridge University Press, Cambridge, England, 2002; 1 CD-ROM. http://www.nr.com/com/storefront.html.

Qu Q., Shore R., Li G., Jin X., Chen L.C., Cohen B., Melikian A.A., Eastmond D., Rappaport S.M., Li H., Rupa D., Waidyanatha S., Yin S., Yan H., Meng M., Winnik W., Kwok E.S., Li Y., Mu R., Xu B., Zhang X., and Li K. Validation and evaluation of biomarkers in workers exposed to benzene in China. *Res Rep Health Eff Inst* 2003: 115: 1–72; discussion 73–87.

Quackenboss J.J., Pellizzari E.D., Shubat P., Whitmore R.W., Adgate J.L, Thomas K.W., Freeman N.C., Stroebel C., Lioy P.J., Clayton A.C., and Sexton K. Design strategy for assessing multi-pathway exposure for children: the Minnesota Children's Pesticide Exposure Study (MNCPES). *J Expo Anal Environ Epidemiol* 2000: 10(2): 145–158.

Quig D. Cysteine metabolism and metal toxicity. *Altern Med Rev* 1998: 3(4): 262–270.

Quinlan J.R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

Reddy M.B., Yang R.S.H., Clewell III H.J., and Andersen M.E. *Physiologically Based Pharmacokinetic Modeling: Science and Applications*. John Wiley & Sons, Hoboken, NJ, 2005.

Ren S.J., and Kim H. Comparative assessment of multiresponse regression methods for predicting the mechanisms of toxic action of phenols. *J Chem Inform Comput Sci* 2003: 43(6): 2106–2110.

RFSC. *R: a Language and Environment for Statistical Computing*. R Foundation for Statistical Computing (RFSC), Vienna, Austria, 2006: http://www.R-project.org.

Rigas M.L., Okino M.S., and Quackenboss J.J. Use of a pharmacokinetic model to assess chlorpyrifos exposure and dose in children, based on urinary biomarker measurements. *Toxicol Sci* 2001: 61(2): 374–381.

Ringrose T.J., and Forth S.A. Simplifying multivariate second-order response surfaces by fitting constrained models using automatic differentiation. *Technometrics* 2005: 47(3): 249–259.

Robertson G.L., Lebowitz M.D., O'Rourke M.K., Gordon S., and Moschandreas D. The National Human Exposure Assessment Survey (NHEXAS) study in Arizona — introduction and preliminary results. *J Expo Anal Environ Epidemiol* 1999: 9(5): 427–434.

Rose R.L., and Hodgson E. Pesticide metabolism and potential for metabolic interactions. *J Biochem Mol Toxicol* 2005: 19(4): 276–277.

Roy A., and Georgopoulos P.G. Reconstructing week-long exposures to volatile organic compounds using physiologically based pharmacokinetic models. *J Expo Anal Environ Epidemiol* 1998: 8(3): 407–422.

Roy A., Weisel C.P., Ballo M.A., and Georgopoulos P.G. Studies of multiroute exposure/dose reconstruction using physiologically based pharmacokinetic models. *Toxicol Ind Health* 1996: 12(2): 153–163.

Salford Systems. *Treenet/MART* 2006: http://www.salford-systems.com/treenet.php.

Serre M. *BMElib, The Bayesian Maximum Entropy software* 2006: http://www.unc.edu/depts/case/BMELIB/.

Sherrod P.H. *DTREG User's Manual* 2006: http://www.dtreg.com.

Silva Mdo C., Gaspar J., Duarte Silva I., Faber A., and Rueff J. GSTM1, GSTT1, and GSTP1 genotypes and the genotoxicity of hydroquinone in human lymphocytes. *Environ Mol Mutagen* 2004: 43(4): 258–264.

Sohn M.D., McKone T.E., and Blancato J.N. Reconstructing population exposures from dose biomarkers: inhalation of trichloroethylene (TCE) as a case study. *J Expo Anal Environ Epidemiol* 2004: 14(3): 204–213.

Sparacino G., Pillonetto G., Capello M., De Nicolao G., and Cobelli C. WINSTODEC: a stochastic deconvolution interactive program for physiological and pharmacokinetic systems. *Comput Methods Programs Biomed* 2002: 67(1): 67–77.

Tahir M.A., Bouridane A., Kurugollu F., and Amira A. A novel prostate cancer classification technique using intermediate memory tabu search. *EURASIP J Appl Signal Processing* 2005: 2005(1): 2241–2249.

Tan Y.M., Liao K.H., and Clewell III H.J. Reverse dosimetry: interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. *J Expo Sci Environ Epidemiol* 2007: 17(7): 591–603.

Tan Y.M., Liao K.H., Conolly R.B., Blount B.C., Mason A.M., and Clewell H.J. Use of a physiologically based pharmacokinetic model to identify exposures consistent with human biomonitoring data for chloroform. *J Toxicol Environ Health A* 2006: 69(18): 1727–1756.

Tang J., Cao Y., Rose R.L., Brimfield A.A., Dai D., Goldstein J.A., and Hodgson E. Metabolism of chlorpyrifos by human cytochrome P450 isoforms and human, mouse, and rat liver microsomes. *Drug Metab Dispos* 2001: 29(9): 1201–1204.

Tarantola A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2005: xii: 342 p.

Tenorio L. Statistical regularization of inverse problems. *SIAM Rev* 2001: 43(2): 347–366.

Thomas D.J., Li J., Waters SB., Xing W., Adair B.M., Drobna Z., Devesa V., and Styblo M. Arsenic (+3 oxidation state) methyltransferase and the methylation of arsenicals. *Exp Biol Med (Maywood)* 2007: 232(1): 3–13.

Timchalk C., Kousba A., and Poet T.S. Monte Carlo analysis of the human chlorpyrifos-oxonase (PON1) polymorphism using a physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model. *Toxicol Lett* 2002a: 135(1–2): 51–59.

Timchalk C., Nolan R.J., Mendrala A.L., Dittenber D.A., Brzak K.A., and Mattsson J.L. A physiologically based pharmacokinetic and pharmacodynamic (PBPK/PD) model for the organophosphate insecticide chlorpyrifos in rats and humans. *Toxicol Sci* 2002b: 66(1): 34–53.

Tomfohr J., Lu J., and Kepler T.B. Pathway level analysis of gene expression using singular value decomposition. *BMC Bioinformatics* 2005: 6: 225.

Tornoe C.W., Overgaard R.V., Agerso H., Nielsen H.A., Madsen H., and Jonsson E.N. Stochastic differential equations in NONMEM (R): implementation, application, and comparison with ordinary differential equations. *Pharm Res* 2005: 22(8): 1247–1258.

Vankeerberghen P., Smeyersverbeke J., Leardi R., Karr C.L., and Massart D.L. Robust regression and outlier detection for nonlinear models using genetic algorithms. *Chemometr Intell Lab Syst* 1995: 28(1): 73–87.

Venables W.N., and Ripley B.D. *Modern Applied Statistics with S-PLUS. Statistics and Computing*. Springer, New York, 1999: xi: 501 p.

Vogel C.R. *Computational Methods for Inverse Problems. Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 2002: xvi: 183 p.

Wang S.W., Georgopoulos P.G., Li G., and Rabitz H. Characterizing uncertainties in human exposure modeling through the random sampling-high dimensional model representation (RS-HDMR) methodology. *Int J Risk Assess Manag* 2005: 5: 387–406.

Wendlandt M., van Beek J.D., Suter U.W., and Meier B.H. Determination of orientational order in deformed glassy PMMA from solid-state NMR data. *Macromolecules* 2005: 38(20): 8372–8380.

Whitmore R.W., Byron M.Z., Clayton C.A., Thomas K.W., Zelon H.S., Pellizzari E.D., Lioy P.J., and Quackenboss J.J. Sampling design, response rates, and analysis weights for the National Human Exposure Assessment Survey (NHEXAS) in EPA region 5. *J Expo Anal Environ Epidemiol* 1999: 9(5): 369–380.

Wilson N.K., Chuang J.C., Iachan R., Lyu C., Gordon S.M., Morgan M.K., Ozkaynak H., and Sheldon L.S. Design and sampling methodology for a large study of preschool children's aggregate exposures to persistent organic pollutants in their everyday environments. *J Expo Anal Environ Epidemiol* 2004: 14(3): 260–274.

Wilson N.K., Chuang J.C., Lyu C., Menton R., and Morgan M.K. Aggregate exposures of nine preschool children to persistent organic pollutants at day care and at home. *J Expo Anal Environ Epidemiol* 2003: 13(3): 187–202.

Winkler D.A. Neural networks in ADME and toxicity prediction. *Drugs Future* 2004: 29(10): 1043–1057.

Wood T.C., Salavagionne O.E., Mukherjee B., Wang L.W., Klumpp A.F., Thomae B.A., Eckloff B.W., Schaid D.J., Wieben E.D., and Weinshilboum R.M. Human arsenic methyltransferase (AS3MT) pharmacogenetics — gene resequencing and functional genomics studies. *J Biol Chem* 2006: 281(11): 7364–7373.

Yamashita F., Bando H., Takakura Y., and Hashida M. A deconvolution method for estimating the first-pass metabolism of orally administered drugs. *Biol Pharm Bull* 1995: 18(12): 1787–1789.

Yang Y.C., Xu X., and Georgopoulos P.G. *Evaluation of Stochastic Population and Bayesian Toxicokinetic Modeling Methods for Exposure and Dose Assessment Applications*. CCL:2008.02, Computational Chemodynamics Laboratory, EOHSI, Piscataway, NJ, 2008: http://ccl.rutgers.edu/reports/MENTOR3P/CCL_2008-report_chloroform.pdf.

Yano Y., Yamaoka K., and Tanaka H. A nonlinear least squares program, MULTI(FILT), based on fast inverse Laplace transform for microcomputers. *Chem Pharm Bull (Tokyo)* 1989: 37(4): 1035–1038.

Yap C.W., Xue Y., Li Z.R., and Chen Y.Z. Application of support vector machines to *in silico* prediction of cytochrome p450 enzyme substrates and inhibitors. *Curr Top Med Chem* 2006: 6(15): 1593–1607.

Yjokley K., Tran H.T., Pekari K., Rappaport S., Riihimaki V., Rothman N., Waidyanatha S., and Schlosser P.M. Physiologically-based pharmacokinetic modeling of benzene in humans: a Bayesian approach. *Risk Anal* 2006: 26(4): 925–943.

Yokel R.A., Lasley S.M., and Dorman D.C. The speciation of metals in mammals influences their toxicokinetics and toxicodynamics and therefore human health risk assessment. *J Toxicol Environ Health B Crit Rev* 2006: 9(1): 63–85.

York T.P., Eaves L.J., and van den Oord E.J. Multivariate adaptive regression splines: a powerful method for detecting disease-risk relationship differences among subgroups. *Stat Med* 2006: 25(8): 1355–1367.

Youssef H., Sait S.M., and Adiche H. Evolutionary algorithms, simulated annealing and tabu search: a comparative study. *Eng Appl Artif Intell* 2001: 14(2): 167–181.