

Environment and Health: A Challenge for Computer Science

Kristina Voigt, Hagen Scherb

Institute of Computational Biology
Helmholtz Zentrum Muenchen
Ingolstaedter Landstr. 1
85764 Neuherberg

kvoigt@helmholtz-muenchen.de, scherb@helmholtz-muenchen.de

Abstract: The importance of modeling and software development in the field of environment and health is briefly outlined in this paper. In this context we introduce the term "exposomics". We give two short modeling examples: the first one is on the evaluation of chemicals in breast milk with the experimental software PyHasse, and the second one on the evaluation of sex odds near the storage cask site Gorleben applying linear logistic regression technique. In both cases we detected an effect of the pollution of the environment and human health. More research and precautionary activities should be initiated in the field of environment and health.

1 Environment and Health

There is no denying the fact that the environment has an enormous impact on human health. It is also consequential that models, information systems, software play an important role in the disclosure of the interactions. Many organizations worldwide are working on the complex scientific subject "environment and health". The World Health Organization (WHO) for example cites on its homepage: The environment is a major determinant of health, estimated to account for almost 20% of all deaths in the WHO European Region [WHO12]. It is striking and scientifically not understandable that an important influence factor on human health, namely ionizing radiation is not mentioned in these definitions. The WHO provides a separate website on radiation stating that "Ionizing radiation has always been a part of the human environment. Along with natural radioactive sources present in the Earth's crust and cosmic radiation, man-made sources also contribute to our continuous exposure to ionizing radiation"[WHO12].

For decades mankind knows that dangerous chemicals are found in human beings, e.g. in breast milk, blood etc. Research has shown how chemicals in our environment can profoundly affect development, growth, maturation, and reproduction by mimicking hormones or interacting with hormone receptors. Exposure to POPs can lead to serious health effects including certain cancers, birth defects (like cryptorchidism), dysfunctional immune and reproductive systems, greater susceptibility to disease, and

even diminished intelligence. Within the last few years, questions and concerns have also focused on the hypothesis that endocrine disrupting chemicals (EDC) may be involved in the dramatic rise in the incidence of metabolic disorders such as obesity and diabetes observed worldwide in the last 40 years [LNL11]. This has been supported by other authors [AQN11] who state that evidence already exists to consider exposure to EDCs as a risk factor in the etiology of type 2 diabetes mellitus and other diseases related to insulin resistance.

Christopher Wild published a definition of the “exposome” as the environmental complement to the genome [WIL05]. That is the reason why within the environmental health discussion, the term exposome and exposomics evolved in recent years. In fact, toxic chemicals enter the body not only from exogenous sources (air, water, diet, drugs, and radiation) but also from endogenous processes, including inflammation, lipid peroxidation, oxidative stress, existing diseases, infections, and gut flora. Thus, even though current evidence suggests that non-genetic factors contribute about 90% of the risks of chronic diseases, we have not explored the vast majority of human exposures that might initiate disease processes. The concept of the exposome, representing the totality of exposures received by a person during life, encompasses all sources of toxicants and, therefore, offers scientists an approach for investigating the environmental causes of chronic diseases [RAP11]. In a recently published paper in *Science* [RSm10], the graphical characterization of the exposome is outlined. The exposome represents the combined exposures from all sources that reach the internal chemical environment.

Figure 1 characterizes the exposome theory with the emphasis on environmental pollution with chemicals as well as with ionizing radiation. The figure has already been published by us [VSB13a].

The linkage between environment and health with respect to computer science has also been outlined recently by Pillmann [PIL12].

In the following sections we will briefly introduce two computer-scientific examples of modeling important environmental health issues. An example of the relationship of the environmental pollution with pesticides and the detection of these pollutants in human breast milk samples in the Taurus Mountains in Turkey is given. The software used is the PyHasse software. The second example comprises the relationship of ionizing radiation with the shift in human sex ratio around the radioactive waste disposal facility Gorleben, Lower Saxony in Germany modeled by ordinary logistic regression [SVo12a]. These are current examples of the research work of both authors. It goes without saying that there exist other approaches in the field of environment and health. A listing of modeling approaches is given by Reis et al., [ROV13].

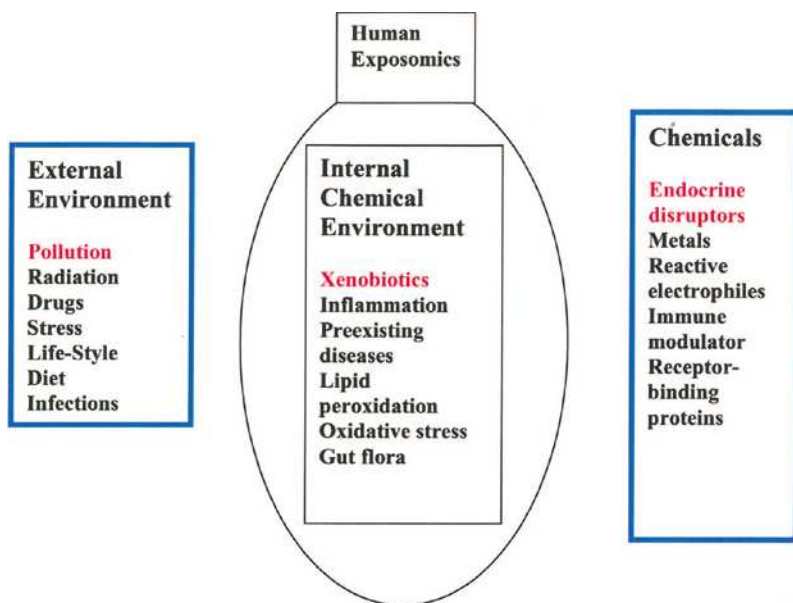


Figure 1: Characterizing the exposome with respect to chemicals (after Rappaport and Smith, 2010) [RSm10]

2 Pesticides in Human Breast Milk Samples

2.1 Data set to be analyzed: 18 pesticides in 44 breast milk samples

The occurrence of POPs/OCPs in the Taurus Mountains in Turkey was studied recently [TAM11], [VBS12a]. Taurus Mountains were suggested for this study because of their potential to act as a sink for organic pollutants by cold condensation and can reflect the atmospheric pollution in Turkey as well as neighboring countries e.g. Arabia, Africa and Russia. In this approach we want to examine the occurrence of 18 pesticides listed in Table 1 in a human medium, namely breast milk.

Table 1: 18 pesticides detected in human and environmental samples in Turkey

Nr.	Acronym	Name	CAS-Number
01	AHCH	alpha-Hexachlorcyclohexane	319-84-6
02	BHCH	beta-Hexachlorcyclohexane	319-85-7
03	GHCH	gamma-Hexachlorcyclohexane	58-89-9
04	PECB	Pentachlorobenzene	608-93-5
05	HCBE	Hexachlorobenzene	118-74-1
06	PPDT	p, p'-Dichlordiphenyltrichlorethane	50-29-3
07	OPDT	o, p'-Dichlordiphenyltrichlorethane	789-02-6
08	PPDD	p, p'-Dichlordiphenyldichlorethane	72-54-8
09	OPDD	o, p'-Dichlordiphenyldichlorethane	53-19-0
10	PPDE	p, p'-Dichlordiphenyldichlorethene	72-55-9

11	OPDE	o, p'-Dichlordiphenyldichlorethene	3424-82-6
12	OXYC	Oxychlordan	27304-13-8
13	CHCE	cis-Heptachloroepoxide	1024-57-3
14	DIEL	Dieldrin	60-57-1
15	END1	Endosulfan-1	959-98-8
16	END2	Endosulfan-2	33213-65-9
17	MECH	Methoxychlor	72-43-5
18	MIRE	Mirex	2385-85-5

2.2 Partial order ranking and software PyHasse

The theory of partial order ranking is presented as an overview in many publications. We only mention two recent ones here [BP10], [BP11].

The analysis of partial order relations as a discipline of Discrete Mathematics will be briefly explained in this paper in order to follow the data-analysis steps.

First step: We need a set of objects. We call this set of objects the ground set, and denote it as G . When speaking in a mathematical context, objects are often denoted as “elements”.

Second step: We need an operation between any two objects. As our aim is an evaluation, we must compare the objects. Is object “a” better than object “b”? If objects a and b are comparable we write $a \perp b$.

Third step: We do not only want that two objects are comparable, but we also would like to know the orientation: Is “a” better or worse than “b”? Therefore the signs $\leq$ and $\geq$ are introduced: $a \leq b$ “may” denote that a is better than b, $a \geq b$ “may” indicate that a is worse b.

Fourth step: The essential point is that we have to define, when we will consider object a as better than b, i.e. the signs “ $\leq$ ” and “ $\geq$ ” alone do not help in an evaluation procedure, we must give them an appropriate orientation.

Fifth step: The indicator values are the basis for comparisons by which objects are characterized. We call the indicators q_i and the value of an object “a” of the i th indicator, $q_i(a)$. After performing the appropriate orientation we define:

$a \geq b$ if and only if $q_i(a) \geq q_i(b)$ for all indicators, $i=1, \dots, m$. (1)

If (1) is not fulfilled, then objects a and b are called “incomparable”, denoted as $a \parallel b$.

By this definition, objects can be not only positioned relatively to each other, but due to the underlying indicator values we know, why and to which extent $a \geq b$. Note that other definitions are possible. Note the inherent consequence of (1).

Let object a be characterized by $m-1$ “good” and 1 “bad” value and b by $m-1$ “bad”, but one “good” value, then nevertheless the definition above declares objects a and b as incomparable. Many algorithms of decision analysis are aggregating the values of indicators whence the effect of the one indicator leading to “bad” for a and “good” for b is smoothed down.

Sixth step: Independent of how we define $\leq$ and $\geq$, the ground set equipped with e.g. “ $\leq$ ” must obey three axioms, if we want to speak of a partially ordered set (poset):

1. Reflexivity: An object a can be compared with itself: $a \leq a$
2. Antisymmetry: If a is better than b, and at the same time b is better than a, then $a = b$. We write:
 $a \leq b$ and $b \leq a \Rightarrow a = b$. Later we will relax this axiom.

3. Transitivity: If a is better than b and at the same time b is better than c, then a is better than c. $a \leq b, b \leq c \Rightarrow a \leq c$.

The ground set G equipped with $\leq$ is a partially ordered set. A widespread notation is: $(G, \leq)$.

Seventh step: A partially ordered set can be represented by a directed graph based on the cover relation: This acyclic, directed graph is called a Hasse diagram.

Very often it is of interest to identify the maximal, minimal and isolated elements:

- Maximal elements have no upper neighbour in the Hasse diagram.
- Minimal elements have no lower neighbour in the Hasse diagram.
- Isolated elements have neither upper nor lower neighbours.

When the indicators are oriented in such a way that large values indicate a risky situation, then the maximal elements are those which are of special concern. When there are several maximal elements, the reason of “being risky” is to be traced back to different indicators. Minimal elements are relatively harmless, isolated elements are of concern because their indicator values are very particular.

The software package PyHasse can be obtained from Dr. Rainer Bruggemann (brg_home@web.de). PyHasse is based on the interpreter programming language Python. Python is used as ‘rapid prototyping’ programming language. PYTHON, a ‘Very High Level Language’ (VHLL) allows a high level of abstraction [Msh07]. PyHasse is in a steady development and has currently more than 90 different modules, together with four libraries.

2.3 Application of PyHasse to data set: 18 chemicals x 44 breast milk samples

The same PyHasse module MHD is now applied on the sample of 18 chemicals detected in 44 breast milk samples in the Taurus Mountains, Turkey. The graphical result, the Hasse diagram is displayed in Figure 2. The 2 maximal objects are: PPDE, BHCH, the 10 minimal objects are END1, END2, OPDT, AHCH, PPDD, PECB, MECH, MIRE, OPDE, OPDD. The chemicals DDE, the first degradation product of DDT and beta-Hexachlorocyclohexane are the worst pollutants in the breast milk samples. They can be compared with those chemicals with which they are connected in the downward direction. The diagram comprises only 4 levels. The number of comparabilities is 52, the number of incomparabilities is 101. This ranking evaluation approach gives an insight into the dependencies of the chemicals. A lot of further information can be drawn out of the PyHasse program, e.g. the similarity of two data matrices, recently described by Voigt et al., [VBS13b].

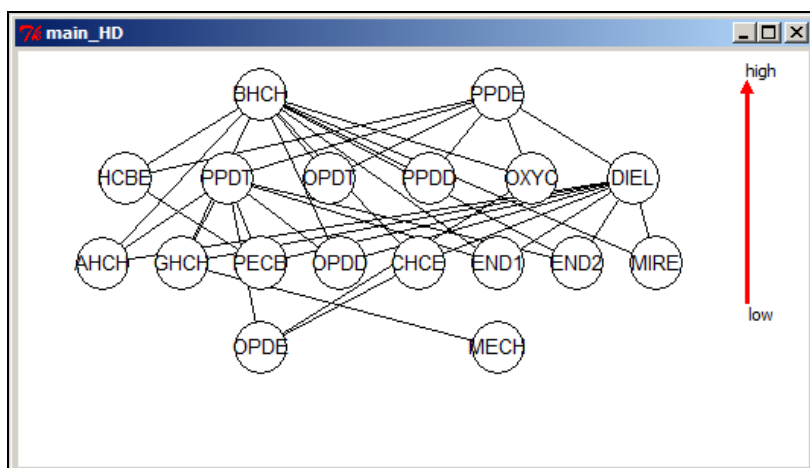


Figure 2: Main Hasse Diagram Technique Module applied on 18x44 Breast Milk Data Matrix

3 Ionizing Radiation Influencing the Human Sex Odds at Birth

3.1 Sex odds at birth: An important indicator

An important indicator for modelling the environmental impact of pollution is the human sex ratio at birth. Sex ratio is the ratio of males to females in a population. The primary sex ratio is the ratio at the time of conception, secondary sex ratio is the ratio at time of birth, and tertiary sex ratio is the ratio of mature organisms. According to Neel and Schull [NSH91], the sex odds is unique among the genetic indicators. Its uniqueness arises from the fact that maternal exposure would be expected to produce an effect different from paternal exposure. For methodological reasons, we prefer "sex odds" over "sex ratio" to not confuse it with the statistical term ratio (see also statistical methods). The ratio of male to female offspring at birth may be a simple and non-invasive way to monitor the reproductive health of a population. Except in societies where selective abortion skews the sex ratio (SR), approximately 105 boys are born for every 100 girls. The authors concluded from a large retrospective cohort study that the sex ratio at birth is remarkably constant [EMH10].

3.2 Geo-spatial background and requirements

Kusmierz et al. [KSV12] gave an overview on the data sources for modelling epidemiological effects of environmental pollution. They explained the special situation and challenges in collecting, handling and analysing the human sex odds data in Germany. The geographic coordinates given in the Gauss–Krüger coordinate system are used for geo-coding municipalities. The Gauss–Krüger coordinate system is a special

transverse Mercator map projection used in Germany, Austria and Finland rather than the UTM-system but similar to this. The central meridians of the Gauss–Krüger zones are only 3° apart, as opposed to 6° in UTM. A transverse Mercator map projection approximates the reference ellipsoid by a cylinder sector, which perimeter soothes the central meridian of the mapped zone some depth below the reference surface, so the elliptical cylinder intersects the ellipsoid. The transverse Mercator map projection provides a nearly conformal mapping of earth's surface in smaller regions, so distances can simply be computed by using the Euclidean distance from the numerical differences of the coordinate components with very small errors.

A lot of work had to be invested in collecting the data due to two major difficulties: first the federal structure of Germany and then the reunification. From 2008, birth data can be downloaded free of charge via the Internet from the "regional database" at the German Federal Statistical Office DESTATIS for all communities under the URL <https://www.regionalstatistik.de>.

3.3 Statistical Methods: Ordinary Linear Logistic Regression

To assess time trends in the occurrence of boys among all live births, and to investigate whether there have been changes in the trend functions after distinct events, we applied ordinary linear logistic regression. This involves considering the male proportion among all male (m) and female (f) births: $p_m = m/(m+f)$. Important and useful parameters in this context are the sex odds: $SO = p_m/(1 - p_m) = m/f$, and the sex odds ratio (SOR), which is the ratio of two interesting sex odds if those two sex odds have to be compared, e.g. in exposed versus non-exposed populations. We used dummy coding for single points in time and for time periods as well. For example, the dummy variable for the time window from 1971 on is defined as $d71(t) = 0$ for $t < 1971$ and $d71(t) = 1$ for $t \geq 1971$. The simple and parsimonious logistic model for a trend and a jump in 1971 has the following form (LB = live births):

$\text{Boys}_t \sim \text{Binomial}(\text{LB}_t, \pi_t)$

$$\log \text{ odds}(\pi_t) = \text{intercept} + \alpha * t + \beta * d_{1971}(t)$$

To allow for changing sex odds trend slopes (broken sticks) after chemical or radiological events, we used dummy coding of time windows and interactions of those time windows with time. The data in this study were processed with Microsoft Excel 2003. For statistical analyses, we used R 2.11.1, MATHEMATICA 8.2, and mostly SAS 9.2 (SAS Institute Inc: SAS/STAT User's Guide, Version 9.2. Cary NC: SAS Institute Inc; 2003).

3.3 Application on change in sex odds near Gorleben/Germany

We have been investigating the influence of ionizing radiation on the human birth sex odds for several years, since we had found increased stillbirths and birth defects after

Chernobyl [e.g. SWe03]. By an initial pilot study, we assessed the trends in the human sex odds of live births and stillbirths combined in several selected European countries with emphasis on the Chernobyl Nuclear Power Plant accident [SVo07]. As this study yielded positive results, we investigated the behavior of the human sex odds of live births after the atmospheric atomic bomb tests and after Chernobyl more thoroughly on a global scale. This comprehensive investigation [SVo11] fully confirmed our opening study [SVo07]. During our research, we also detected increased sex odds near NF in Germany and Switzerland [KVS10]. The continuous discussion about the nuclear waste shipping casks storage (in German: Transportbehälterlager, TBL) in Gorleben, Lower Saxony in Germany, increased our interest in taking a closer look at this location. Gorleben, Lower Saxony in Germany, is known for a radioactive waste disposal facility (TBL – Transportbehälterlager), used for intermediate storage of highly radioactive waste (HAW). Previously, we have found increased human birth sex odds within 30 to 40 km distance from the TBL Gorleben right after the first castor had arrived at Gorleben in Spring 1995 [KVS12, KVS10, SVo11, SVo12b]. This result has been confirmed by an official investigation by the “Niedersächsisches Landesgesundheitsamt” in 2011, and has been discussed in a subsequent workshop in 2012 [NLG12]. See Figure 3 for a possible alternative change-point model for Gorleben with a significant jump in 1998 instead of 1996.

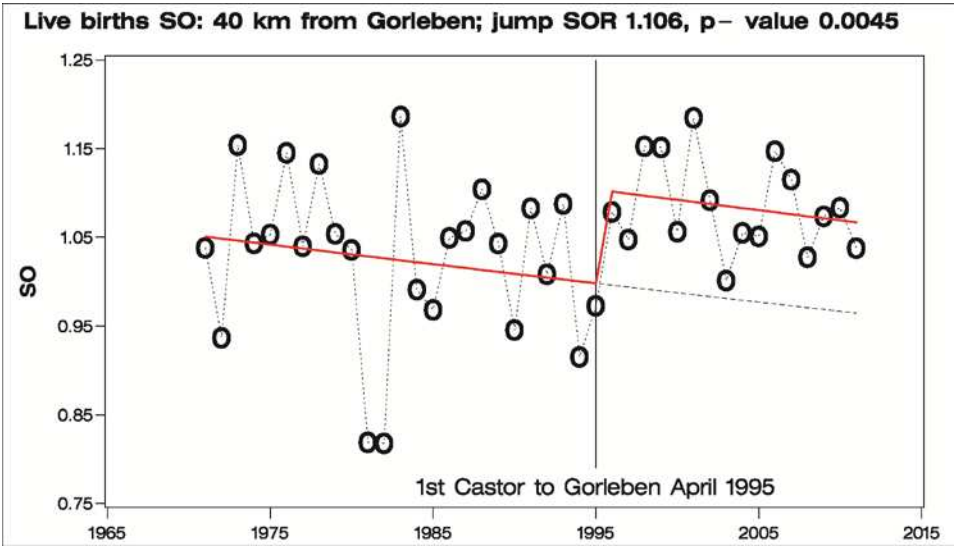


Figure 3: Trend of the live births sex odds around the TBL Gorleben, Germany, 1971 – 2011

We detected increased sex odds around the nuclear storage site TBL Gorleben which went into operation in in the year 1995. The main outcome of our studies is shown in Table 2. Applying a Wald Chi-square two-sided 2×2 table test yields a rather high SOR 1.0838 with 95% CI (1.0391, 1.1305), and a significant p value of 0.000180. These results have been published recently [SVo12b].

Table 2: Births by gender within 40 km from the nuclear storage site Gorleben (TBL) before and after the TBL went in operation in April 1995

Period	Live Births			SO	SOR	ln(SOR) / SE	p-value
	total	m	f				
1981-1995	13 861	6 939	6922	1.0025	1.0838	0.0805	0.000180
1996-2010	23 135	12 047	11088	1.0865		0.0215	
Total	36 996	18 986	18010				

3 Discussion and Outlook

Epidemiological effects of environmental pollution can be assessed and modeled in many different ways. Pollution of human breast milk and the human sex ratio at birth are just two examples. With our briefly introduced modeling and data evaluation approaches, namely the Hasse diagram technique and the linear logistic regression we detected visible and significant results indicating the effects of environmental pollution (with chemicals and ionizing radiation) on human health. The underestimation of the influence of environmental pollution on human health is irresponsible to current and future generation. Not only more research in this area should be initiated and performed but it is an urgent need to act more prudently with handling chemicals and ionizing radiation.

We will continue our research focus in the direction of environmental and health with respect to chemicals and ionizing radiation. This is done in collaboration with experimental scientists at the moment and should be enlarged to other disciplines.

References

[AQN11]Alonso-Magdalena, P.; Quesada, I.; and Nadal, A.: Endocrine disruptors in the etiology of type 2 diabetes mellitus, *Nat Rev Endocrinol*, 2011, 7, (6), pp. 346-353.

[BPa10] Bruggemann, R.; and Patil, G.P.: Multicriteria prioritization and partial order in environmental sciences, *Environmental and Ecological Statistics*, 2010, 17, (4), pp. 383-410.

[BPa11] Bruggemann, R.; and Patil, G.P.: *Ranking and Prioritization for Multi-indicator System*, Springer, 2011.

[EMH10] Ein-Mor, E.; Mankuta, D.; Hochner-Celnikier, D.; Hurwitz, A.; and Haimov-Kochman, R.: Sex ratio is remarkably constant, *Fertil Steril*, 2010, 93, (6), pp. 1961-1965.

[KVS10] Kusmierz, R.; Voigt, K.; and Scherb, H.: Is the human sex odds at birth distorted in the vicinity of nuclear facilities (NF)? A preliminary geo-spatial-temporal approach. In (Greve, K. Ed): *Proceedings EnviroInfo 2010*, Shaker Verlag, Aachen, 2010, pp. 616-626.

[KSV12] Kusmierz, R., Scherb, H.; and Voigt, K.: Overview on data sources for modelling epidemiologic effects of environmental pollution. In (Seppelt, R., Ed.) *Proceedings of iEMSS 2012*, Leipzig, pp. 1-10.

[LNK11] Lyche, J.L.; Nourizadeh-Lillabadi; R., Karlsson, C.; Stavik, B.; Berg, V.; Skare, J.U.; Alestrom, P.; and Ropstad, E.: Natural mixtures of POPs affected body weight gain and induced transcription of genes involved in weight regulation and insulin signaling, *Aquat Toxicol*, 2011, 102, (3-4), pp. 197-204.

- [MSh07] Muller, M.; and Schwarzer, S.: Python im deutschsprachigen Raum, In Book Python im deutschsprachigen Raum, Lehmanns Media, 2007.
- [NSh91] Neel, J.V.; and Schull, W.J.: The children of atomic bomb survivors: a genetic study National Academy, Washington, 1991.
- [NLG12] NLGA, Niedersächsisches Landesgesundheitsamt, Sekundaeres Geschlechterver-haeltnis in der Umgebung des Transportbehaelterlagers TBL Gorleben-Fachgespraech am 12 Maerz 2012, <http://www.nlga.niedersachsen.de/download/65642>.
- [PIL12] Pillmann, W.: Interlink between environment and health information. In: (Arndt, H.K.; Knetsch, G.; Pillmann, W., Eds.) Proceedings of EnviroInfo 2012, Shaker Verlag, Aachen, 2012, pp. 61-68.
- [RAP11] Rappaport, S.M.: Implications of the exposome for exposure science, J Expo Sci Environ Epidemiol, 2011, 21, (1), pp. 5-9.
- [RSm10] Rappaport, S.M.; and Smith, M.T.: Epidemiology. Environment and disease risks, Science, 2010, 330, (6003), pp. 460-461.
- [ROV13] Reis, S.; Oxley, T.; and Voigt, K.: The role of models in assessing environmental impacts on human health, Position paper, Environ Model and Software 2013, submitted.
- [SVo07] Scherb, H.; and Voigt, K.: Trends in the human sex odds at birth in Europe and the Chernobyl Nuclear Power Plant accident, Reprod Toxicol, 2007, 23, (4), pp. 593-599.
- [Svo11] Scherb, H.; and Voigt, K.: The human sex odds at birth after the atmospheric atomic bomb tests, after Chernobyl, and in the vicinity of nuclear facilities, Environ Sci Pollut Res Int, 2011, 18, (5), pp. 697-707.
- [Svo12a] Scherb, H.; and Voigt, K.: Response to W. Kramer: The human sex odds at birth after the atmospheric atomic bomb tests, after Chernobyl, and in the vicinity of nuclear facilities: comment, Environ Sci Pollut Res Int, 2012a, 19, (4), pp. 1335-1340.
- [SVK12] Scherb, H.; Voigt, K., and Kusmierz, R.:Modelling the human radiation risks around nuclear facilities in Germany and Switzerland: A case study. In (Seppelt, R. Ed.) Proceedings iEMSs 2012, Leipzig, 2012, pp. 1-12.
- [SWe03] Scherb, H.; and Weigelt, E.: Congenital malformation and stillbirth in Germany and Europe before and after the Chernobyl Nuclear Power Plant accident, Environ Sci Pollut Res Int, 2003, pp. 117-125.
- [TAM11]Turgut, C.; Atatanir, L.; Mazmanci, B.; Mazmanci, M.A.; Henkelmann, B.; and Schramm, K.W.: The occurrence and environmental effect of persistent organic pollutants (POPs) in Taurus Mountains soils, Environ Sci Pollut Res Int, 2011, 19, (2), pp. 325-334.
- [VBS13a] Voigt, K.; Bruggemann, R.; Scherb, H.; Cok I.; Mazmanci, B.I.; Mazmanci, M.A.; Turgut, C.; and Schramm, K.-W.: Evaluation of organochlorine pesticides in breast milk samples in Turkey applying features of the partial order technique, International Journal of Environmental Health Research, 2013a, 23 (3), pp. 226-246.
- [VBS13b] Voigt, K.; Bruggemann, R.; Scherb, H.; Cok, I.; Mazmanci, B.; Mazmanci, M.A.; Turgut, C.; and Schramm, K.-W.: Organochlorine Pesticides in the Environment and Humans: Necessity for Comparative Data Evaluation. In (Wittmann, J. Ed.) Proceedings Simulation, Leipzig 10.-12. April 2013b, Shaker Verlag, Aachen, accepted.
- [WHO12] WHO, World Health Organization: Environment and health, <http://www.euro.who.int/en/what-we-do/health-topics/environment-and-health>, 2012.
- [WIL05] Wild, C.P.: Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology, Cancer Epidemiol Biomarkers Prev, 2005; 14, (8), pp. 1847-1850.