Heat Map Visualization of Complex Environmental and Biomarker Measurements

Joachim D. Pleil\textsuperscript{a*}, Matthew A. Stiegel\textsuperscript{b}, Michael C. Madden\textsuperscript{c}, and Jon R. Sobus\textsuperscript{a}
\textsuperscript{a}Human Exposure and Atmospheric Sciences Division, NERL/ORD, \textsuperscript{b}SSA, Human Exposure and Atmospheric Sciences Division, \textsuperscript{c}Environmental Public Health Division, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711

Abstract

Over the past decade, the assessment of human systems interactions with the environment has permeated all phases of environmental and public health research. We are invoking lessons learned from the broad discipline of Systems Biology research that focuses primarily on molecular and cellular networks and adapting these concepts to Systems Exposure Science which focuses on interpreting the linkage from environmental measurements and biomonitoring to the expression of biological parameters. A primary tool of systems biology is the visualization of complex genomic and proteomic data using “heat maps” which are rectangular color coded arrays indicating the intensity (or amount) of the dependent variable. Heat maps are flexible in that both the x-axis and y-axis can be arranged to explore a particular hypothesis and allow a fast overview of data with a third quantitative dimension captured as different colors. We are now adapting these tools for interpreting cumulative and aggregate environmental exposure measurements as well as the results from human biomonitoring of biological media including blood, breath and urine. This article uses existing EPA measurements of environmental and biomarker concentrations of polycyclic aromatic hydrocarbons (PAHs) to demonstrate the value of the heat map approach for hypothesis development and to link back to stochastic and mixed effects models that were originally used to assess study results.

*corresponding author: pleil.joachim@epa.gov
Introduction

The most promising new ideas for advancing public health and medical science are coming from genome- and proteome-based investigations for estimating individual responses to pharmaceuticals and other exogenous compounds (Manolio 2010, Zhou et al. 2009). There have been numerous successes in identifying genetically linked drug effectiveness and genetic predispositions to diseases (Gonzales-Angulo et al. 2010, Knight 2009). As an outgrowth to this work, researchers are further hypothesizing that adverse health outcomes previously considered “bad luck” might in reality have subtle environmental triggers. Termined the “gene-environment interaction”, the hypothesis is that autoimmune disease, cancer, and other chronic or rare disease conditions are the result of more than just chance (Ekins et al. 2005, Smith and Rappaport 2009, Wild 2005). This concept has been captured succinctly by Prof. Judith Stern from University of California, Riverside, who said:

"Genetics loads the gun, but environment pulls the trigger."

This simple expression (which is oft quoted across the internet) is now shaping the dialog and understanding of environmental exposure science (e.g. Suk et al. 2002) although the reality is undoubtedly much more complicated. For example, using the analogy for building a house, the genetic code serves only as the “blueprint” for the human organism; further complexity is developed through the expression and modification of proteins that can be thought of as the construction materials and tools, and epigenetic factors that serve as plan modifications after construction has begun. Much as houses built from the same blueprint can differ at completion due to the owners’ tastes and the workmen’s talents, the biochemical implementation of the genetic information can be directly or indirectly affected by a variety of environmental factors resulting in different human phenotypes, susceptibilities to disease, and health outcomes.

Genomics and proteomics (and other ‘omic specialties) have evolved into the broader academic undertaking referred to as systems biology that endeavors to include the networks of interactions among genes, proteins, and other bio-chemicals that define life-processes (Aggarwal and Lee 2003, Bruggeman and Westerhoff 2007, Edwards and Preston 2008, Stark 2008). Recently, the concept of the “human exposome” has been added to explore the external influences on internal human biochemistry (Rappaport and Smith 2010, Wild 2005). A parallel framework of systems exposure has been defined that explores the conceptual and statistical linkages between environmental exposures and internal dose using network diagrams and variance components (Pleil 2009). Based on the exposome concepts for toxicity testing, public health, and risk assessment (Birnbaum 2010, Hubal 2009, Sheldon and Hubal 2009), we have recently proposed the exploitation of empirical measurements of biological parameters that reflect systems responses and could conceivably be mapped forward to adverse perturbations in the normal operation of the human organism (Pleil and Sheldon 2010). Thus, in using systems biology approaches for systems exposure research, we expect to adapt some of the ‘omics tools such as network interpretation, grouping and clustering of variables, and heat map visualization (Auman et al. 2007, Eisen et al. 1998, Kohn et al. 2006).

In this article, we focus on heat maps as the first step for interpreting complex environmental exposure data. We present a series of heat map applications to environmental
and human biomarker data and subsequently show how to interpret visual trends to infer
underlying structure in data. We also demonstrate how to assess and develop specific
hypotheses using heat map visualization, and subsequently describe statistics and modeling to
confirm these observations.

**Experimental**

Heat map visualization is implemented herein for two different types of studies. The first
is a purely environmental media sampling application that was used to assess the impact of the
World Trade Center (WTC) disaster fires that contributed particulate matter and smoke to the
New York City air shed from 9/11/2001 to about 12/20/2001 via measurement of PM2.5 and
associated particle bound polycyclic aromatic hydrocarbons (PAHs) (Pleil et al. 2004). The
second is an assessment of bloodborne PAHs biomarkers from a human chamber study of diesel
exhaust exposures (Pleil et al. 2010, Sobus et al. 2008). In both cases, the data structure is
complex in that there are multiple chemical analyses, different time frames, and fixed effects
variables to consider. We note that PAHs are of interest as they comprise a ubiquitous
environmental chemical group with known human carcinogenicity.

**Environmental study:**

One of the many data sets from the WTC disaster was derived from a series PM2.5
Teflon particle filter samples collected by EPA personnel at three fence line locations close in to
WTC “ground zero” and one location about 0.5 km north of ground zero at 290 Broadway
(Figure 1). Samples were collected daily at each site (~24 hr duration) between Sep. 23, 2001
and Mar. 27, 2002. The filters were initially (non-destructively) assayed for PM2.5 mass and
subsequently solvent extracted to measure particle bound PAHs. The method for PAHs analysis
has been published (Pleil 2004); briefly, filters were extracted with 10 ml of dichloromethane
containing deuterated PAHs as internal standards. After concentration of the extracts to 50 µl
under N2, 2-µl aliquots were analyzed by gas chromatography – mass spectrometry in selective
ion monitoring mode. Results were reported for 9 carcinogenic and/or mutagenic PAHs, namely,
benz(a)anthracene (BaA), chrysene (Chr), benzo(b)fluoranthene (BbF), benzo(k)fluoranthene
(BkF), benzo(e)pyrene (BeP), benzo(a)pyrene (BaP), indeno(1,2,3-c,d)pyrene (Ind),
dibenz(a,h)anthracene (DaA), and benzo(g,h,i)perylen (BghiP).

The resulting data structure was comprised of air volume sampled (m³), absolute PM2.5
mass (µg), ambient mass concentration (µg/m³), nine individual PAHs concentrations (ng/m³),
four sampling sites (a,b,c, and k), and a sampling date for each sample. To discover if the WTC
fires influenced the composition of the normal NYC particulate matter, we constructed an
additional parameter for each sample to calculate enrichment factors for individual PAHs: PAHsi
/PM2.5 (ng/m³)/(mgPM2.5) where subscript i refers to the i^{th} PAH from 1 to 9 (Pleil et al. 2006).

**Biomarkers of exposure study:**

The general public is exposed to PAHs on a regular basis via ingestion of charbroiled foods,
dermal contact with fuels, oils, and lubricants, and inhalation of products of incomplete
combustion. As such, there is always some chronic background level of PAHs in the circulating human blood (Pleil et al. 2010). Diesel exhaust (classified as a known human carcinogen) is an important source of inhalation exposure to PAHs (US EPA 2002). To determine the importance of this source with respect to the baseline, we performed a controlled chamber study wherein human subjects were exposed for two hours to a typical “busy urban intersection” level of diesel exhaust set at 100 ug/m3 particulate matter (DPM) plus the associated combustion gases. The study also included randomly assigned control experiments wherein the subjects were exposed only to purified (clean) ambient air. In addition to the exposures, the subjects performed 15 minute long scripted aerobic exercise segments (75 watts on stationary bicycle) to elevate respiratory rates. Blood samples were collected before and immediately after the exposures and also about 24 hours later. All exposures were feedback controlled to achieve consistency and environmental samples were collected to establish precise PAHs levels in the chamber air (Sobus et al. 2008). All samples were collected and analyzed under Institutional Review Protocols (IRB) approved by University of North Carolina and US EPA with informed consent of the subjects (IRB# 99-EPA-283: Physiological, Cellular, and Biochemical Effects of Diesel Exhaust in Healthy Young Adults).

Blood samples were centrifuged and the plasma fractions were analyzed for 21 PAHs: naphthalene (Nap), acenaphthylene (Acl), acenaphthene (Ace), fluorene (Flu), phenanthrene (Phe), anthracene (Ant), fluoranthene (Flt), pyrene (Pyr), benzo[an]thracene (BnP), benz(a)anthracene (BaA), chrysene (Chr), 7,12-dimethbenz[a]anthracene (DmB), benzo(b)fluoranthene (BbF), benzo(j)fluoranthene (BjF), benzo(k)fluoranthene (BkF), benzo(e)pyrene (BeP), benzo(a)pyrene (BaP), 3-methylcholanthrene (MeC), indeno(1,2,3-c,d)pyrene (Ind), dibenz(a,h)anthracene (DahA), and benzo(g,h,i)perylene (BgaP). The methodology for blood analysis is based on liquid-liquid extraction with hexane containing isotopically labeled standards, volume reduction, and subsequent GC-MS analysis (Pleil et al. 2010).

**Construction of Heat maps**

Heat maps are visual representations of quantitative data on two axes; the x-axis generally reflects individual samples and the y-axis consists of groups of measured parameters. The field between the axes is comprised of an array of contiguous boxes color coded to reflect quantitation. As such, heat maps are a flexible visualization tool for grouping data and exploring patterns. They can be created using various software products; in our case, we employ MatLab (MathWorks, Natick, MA). The basic data structure underlying a heatmap is comprised of an independent variable (human subjects, environmental samples, days or dates, etc.) and an array of dependent variables (environmental chemicals, biomarker chemicals, biological parameters, etc.). The dependent parameters are usually continuous variables that are further coded by color to reflect their quantitative value. Both axes are generally further sub-grouped. For example, the samples on the x-axis can be ordered by time collected, by distance from a source, or sub-divided by human host factors such as gender or age; the dependent variables can be grouped by parameters such as pre- and post-treatment, geographical site, season, or compound class.

The construction of any particular heat map is based purely on choices made by the investigator to explore a relationship between sample groups and measurements. For example,
in the environmental data example, we were primarily interested in the relationship between time elapsed after 9/11 and the level of PAHs in air at the different sampling sites. As such, the samples were ordinarily grouped by date on the x-axis and the compound groups arranged in blocks by location on the y-axis. The quantitative color scheme was chosen in a log scale (blue = lowest, red = highest) to achieve optimal range in the display; the annotated color quantitation scale is seen as a vertical column at the right side of the heat map. Subsequent refinements could then be made to explore more detailed observations (compositing sample sites) and other data calculations (PAHs enrichment, compound differences).

Results and Discussion

Interpretation of environmental data: There are a series of questions or hypotheses of interest regarding the impact of the WTC disaster, especially with regard to PAHs and the fugitive dusts and fires.

1. Did the WTC disaster indeed contribute excess PAHs to the local air?
2. Are there absolute differences and distributions (patterns) among specific PAHs that could be used to assess their source?
3. Are PAHs enriched in particulate matter over normal NYC baseline?
4. How does the behavior of individual PAHs vary with respect to all PAHs and elapsed time after the disaster?

Because the unifying feature of this data set is sample date, these questions can all be explored with a series of simple heat maps where the x-axis is the number of days after 9/11/2001.

Heat map visualization: To address question #1, we plot the individual PAHs air concentrations (ng/m³) partitioned into four horizontal blocks based on sampling site on the y-axis. We coded the measured concentrations on a sliding logarithmic color scale where the lowest values are represented by dark blue = 0.001, midrange values by yellow = 0.1, and the highest values dark red = 10 ppbv. Figure 2a shows this heat map where we can immediately see that the absolute airborne PAHs concentrations do indeed decrease in time to some stable background level and that there is appreciable scatter within- and between-day. We also can answer part of question #2 and observe that there are absolute differences among compounds and that the B site is appreciably lower in PAHs concentration than the K, A, and C sites. More specifically, dibenz(a,h)anthracene (DaA) is consistently lower than the other 5- and 6-ring PAHs and benzo(g,h,i)perylene (BgP) shows little, if any, decrease over time.

We then redraw this map in Figure 2b using individual PAHs divided by sample PM2.5 mass (ng/m³)/(mgPM2.5) to explore question #3. In this graph, we also represent concentrations with a sliding color scale, but due to the change in units, lowest values are now represented by dark blue = 0.01, midrange values by yellow = 0.9, and the highest values dark red = 10 (ng/m³)/(mgPM2.5). We observe an additional enrichment effect of the particle composition as seen by the starker color contrast from left to right (that is, there is more intense red color in the initial 60 days). The difference of PAHs enrichment between the remote site B and the
fenceline sites K, A, and C is even more exaggerated than the pure concentration data from Figure 2a which confirms that excess particles and particle bound PAHs are likely coming from the WTC site.

To gain insight into the overall source changes over time, we use the per day averages of the three fence line sites (K, A, and C) to smooth out scatter from micro- and meso-scale wind shifts. Figure 3a shows the absolute average PAHs concentrations at the WTC fenceline and Figure 3b shows the average PAHs concentrations proportional to PM$_{2.5}$. Again we see that there is both an absolute and a relative per mass decrease of PAHs with time; the smoother appearance of these heat maps supports the conjecture that local heterogeneity in wind was at least partially responsible for the scatter among sites K, A, and C seen in Figures 2a,b.

Statistical confirmation: Using a mixed regression model approach, previous work with WTC data had shown that the initial decrease in time of airborne PAHs and PM$_{2.5}$ was exponential in nature and that the fenceline measurements were indistinguishable by site (K, A, and C), and also significantly higher than the measurements made at the more remote site B. Specifically, the median total PAHs values for the fenceline sites were 2.6 (s.d. 1.12) times higher than for site B for the first 100 days, and the mean values for day 3 vs. day 100 were 11.0 (s.d. 6.64). ANOVA tests of log-space transforms showed no significant difference among measures from the fenceline sites when treated as replicates (p = 0.13) but showed significant difference from site B (p < 0.0001) (Pleil et al. 2004). Additionally, the linkage between hard data of PAHs and PM$_{2.5}$ was exploited to make geospatial heat maps of the major metropolitan area of New York City. Estimated PAHs concentrations were calculated with a Bayesian maximum entropy approach using abundant PM$_{2.5}$ measurements coupled with sparse PAHs data and confirmed both the spatial and temporal behavior observed for the PAHs measurements alone (Allshouse et al. 2009). In fact, the geospatial and the temporal maps from the Bayesian approach are remarkably similar to the data based heat maps in Figs. 2 and 3 herein. As such, the simple and fast observations from the heat maps described above are validated by very sophisticated mathematical methods. In contrast to rigorous modeling and statistical methods, the heat maps are a qualitative, or non-parametric, approach of analysis. In the case of these environmental measurements, not every day is represented for all sites, as such, the x-axis is ordinal in time, but not in the form of a continuous variable. Furthermore, the colors appear discrete to the eye rather than smoothed. However, we found that this modular presentation is preferable for elucidating patterns as it is difficult to present missing data without skewing the visual effects.

Interpretation of blood biomarker data: The primary issue in this experiment was to determine the relationship, if any, between pre-existing PAHs baseline levels in blood and incremental contributions from diesel exhaust. Specifically, the initial questions were:

1. Are there measurable quantities of PAHs in circulating blood?
2. Is there a “treatment” effect from the diesel exposure and/or the chamber exercise experience?
3. How do individual PAHs vary between- and within- persons?
The first question was answered by analytical methods development and comparison to “real world” samples; we found that we could estimate median values for all compounds but that for the heavier 5- and 6-ring compounds, normal control subjects often had levels that needed to be estimated below objective detection limits based on subjective interpretation (Pfeil 2010). Imputation of these values did not affect the non-parametric interpretation via subsequent heat mapping. Unlike the previous example where samples were dependent on time after an event, there was no obvious way to order the x-axis. As such, we used broad grouping strategies to arrange the subjects.

Heat map visualization: Because there was no over-arching time component to the individual subjects, the x-axis was arranged in ordinal grouped segments as follows: pre-exposure, post-exposure, and 24hr post-exposure within which subjects were grouped by gender. The vertical axis was arranged into two groups by experiment: clean air and diesel exhaust in matched pairs. Furthermore, each group was subdivided by compound list based on PAHs volatility (naphthalene at the top, down to dibenzopyrene). Due to some data dropouts from loss to followup of certain subjects, we show data only from 8 individuals (5 female, 3 male) for which complete data sets were available (both “diesel” and “clean” exposures for all three time points). Again, the relationship of color gradation to quantitation is given in a column at the right side of the heat maps. Figure 4a shows the heat map for the circulating blood-borne PAHs concentrations in ng/ml. At first glance, the answer to question #2 is apparent: there seems to be no obvious treatment effect (that is, no difference between clean air exposure and diesel contaminant exposure), and no remarkable differences among, pre, post and 24hr post exposure segments for either group. There are, however consistent differences between subjects and between compounds. The internal groups show a gender effect wherein four of the five females consistently carry greater body burden of PAHs than the male subjects regardless of treatment or timing. Although we have access to other host factor parameters, there are not enough representatives in sub-categories to make statistical comparisons.

To address question #3 regarding obvious between- and within-subject variance, we redraw the heat map (Figure 4b) for the same data but rearranged to treat all sample times and treatment groups as independent. Now all available data are grouped by subject in the following order: pre clean, pre diesel, post clean, post diesel, 24 hr post clean, 24 hr post diesel. Here we see that within subject variance is definitely smaller than between subject variance, again with the biggest differences found between genders.

Statistical confirmation: Preliminary analyses using mixed model regression (proc MIXED, SAS, Cary, NC) indicated that there was no significant time effect (pre-exposure vs. post-exposure vs. 24-hr post-exposure; p = 0.5), treatment effect (diesel exhaust vs. clean air; p = 0.7), or time×treatment interaction effect (p = 0.9) on total PAH levels in blood. We have recently published a similar analysis from this study using measurements of endogenous metabolites in exhaled breath condensate (EBC) where we confirmed this observation and also demonstrated a significant gender effect and appreciable between subject differences (Hubbard et al. 2009). We point out that only 8 of the 10 subjects completed all 12 of the blood draws successfully and so some data could not be plotted as there were no visual comparisons possible. In contrast to the visual approach of the heat maps, with parametric statistics we can include all data in the models regardless of their completeness because the error estimates are included in
the significance calculations (Sobus et al. 2010). However, we note that the visual inspection of
the heat map data accurately and quickly predicted the eventual statistical results.

*Future work and heat map applications:* We are currently embarking on a second diesel exhaust
study at three-times the particulate levels wherein an additional 15 subjects will be tested. We
expect to confirm the gender effect, add in other host factors such as ethnicity, height, weight
and age, and will further explore the treatment effect using heat maps and statistical analyses.

Based upon what we have learned so far, we now understand that there is a basic process
with which heat maps can be drawn to construct or explore patterns. The first step should be to
group data for the y-axis based on experimental design; that is, by sampling locations (e.g.
indoor vs. outdoor, rural vs. urban, downwind vs. upwind, etc.), by treatment or intervention
(clean air vs. diesel exhaust, job category, etc.) or by event (e.g. year 1 vs. year 2, summer vs.
winter, high traffic vs. low traffic, etc.). The y-groupings should represent some major divisions
that are expected to answer the broader questions of the experiment with the fine structure
comprised of individual chemical measurements in parallel order. The x-axis, on the other hand,
should be constructed more to develop trend information or minor grouping effects. For
example, in Figures 2 and 3, we use time after the WTC event to explore changes in a semi-
continuous sense; in Figure 4 we use broad structure (male vs. female) with a fine structure of
ordinal sample grouping (pre-, post, and 24-hr post) intervention. This general guidance is not
necessarily the only way to construct heat maps. Often, the nature of the experiment itself guides
the best way to interpret data in that the hypothesis helps decide which parameters are used on
the y-axis vs. the x-axis.

**Conclusions and Recommendations**

The visualization of complex measurement data via a heat map approach is a valuable
screening tool for quickly testing broad hypotheses regarding relationships among exposure
measurements, biomarkers, meta-data, and host factors before computational efforts are
expened. Certainly, subsequent detailed statistical analyses are always required to determine
the quantitative aspects (e.g. correlations, p-values, etc.), however, the examples presented here
demonstrate that heat map visualization interpretation is useful and robust with respect to
computational methods. An inverse benefit of the three dimensional heat map style graphics is
to concisely illustrate pre-existing results from statistical and multivariable analyses that would
otherwise be relegated to complex numerical tables and/or a series of two-dimensional graphs.

We caution that data visualization is a qualitative tool; missing or unbalanced data cannot
be properly represented or imputed as in quantitative statistics or regression analysis. We have
noted in the WTC example that the x-axis is ordinal in time, but due to missing data days, time
cannot be represented as a continuous variable. In the chamber study experiment, we noted that
unbalanced comparisons could not be presented without disturbing the overall effect. We
therefore recognize that visualization techniques, regardless of their elegance and power to
communicate complex results, must always be confirmed with robust computation.
We recommend that heat map representations become a mainstream tool to communicate the interactions among environmental measurements or classifications, biomarker measurements, and other meta-data in systems exposure science. As shown in the examples in this article, data can be rearranged to observe potential relationships among variables, groups of variables, and groups of samples. Numerical data can be normalized by color range to show trends within different absolute ranges and in response to underlying distributions (normal, lognormal). The resulting patterns (or lack thereof) can be readily observed and the qualitative results can be efficiently communicated to the ultimate user of the information.

Acknowledgements

The authors wish to thank Tzipporah Kormos and Myriam Medina-Vera of US EPA for their expert advice. This research has been subjected to (EPA) Agency review and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Interested readers are invited to contact the authors for assistance and advice in constructing specific heat maps.
References


Figure 1. Aerial photograph of lower Manhattan taken on Sept. 23, 2001 from an altitude of 3,300 feet. Ground zero of the World Trade Center (WTC) and 4 sampling sites are labeled. Site A: Park Place & W. Broadway; Site C: Trinity & Cedar; Site K: West & Albany; Site B: 290 E. Broadway between Reade and Duane. Photo: courtesy National Oceanographic and Atmospheric Administration (NOAA).
Figure 2a, b. Heat maps for concentrations of nine particle-bound PAHs data in ordinal space after WTC disaster on 9/11/2001 by sampling sites K, A, C, and B (see Figure 1). 2a shows absolute concentrations in air (ng/m³) and 2b shows PAHs enrichment by particle-mass in air (ng/m³)/(mgPM_{2.5}). Heat maps indicate that PAHs concentration and enrichment decrease with time after 9/11, that there are differences between compounds, and that there is appreciable within day variance. Furthermore, enrichment data demonstrate a starker contrast while concentration data have a wider dynamic range and site B is distinctly different from sites K, A, and C.
Figure 3a, 3b. Heat maps for concentrations of nine particle-bound PAHs data in ordinal space after WTC disaster on 9/11/2001 by daily averages of the data from three fenceline sites K, A, and C. 3a shows the daily absolute average PAHs concentrations at the WTC fenceline and 3b shows the average PAHs concentrations proportional to PM$_{2.5}$. Heat maps indicate an absolute and a relative per mass decrease of PAHs with time; the smoother appearance supports the conjecture that local heterogeneity in wind was responsible for the within-day variance among sites K, A, and C seen in Figures 2a,b.
Figure 4a, b. Heat maps for circulating blood-borne concentrations in ng/ml for 21-PAHs. 4a shows sample grouping along the x-axis by time with respect to chamber exposure with subgrouping by subject gender. The vertical axis is grouped by experiment type (clean air or diesel exhaust at 100 µg/m³) with internal grouping by compound. There are no apparent time or exposure effects but there is a perceived gender effect in that females demonstrate higher levels on average. 4b shows the same data re-arranged by individual subjects that demonstrates that within subject variance is definitely smaller than between subject variance for all samples, again with a main trend for between-gender variance.