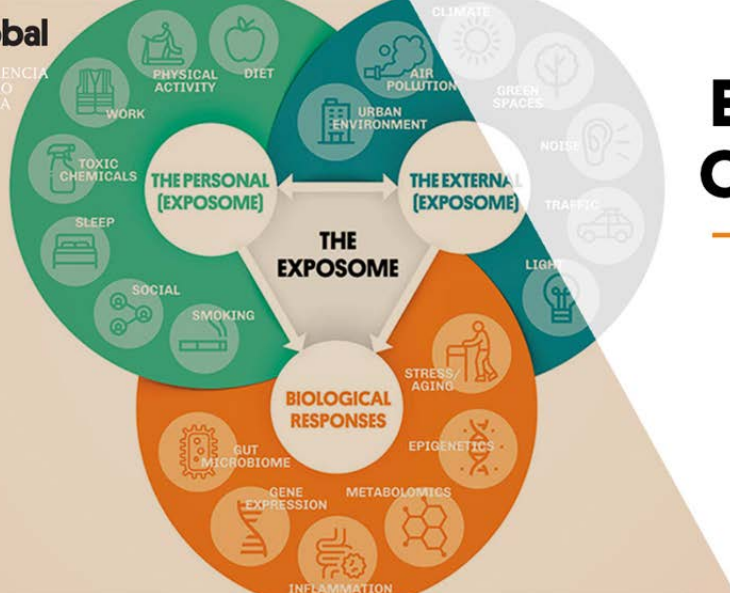


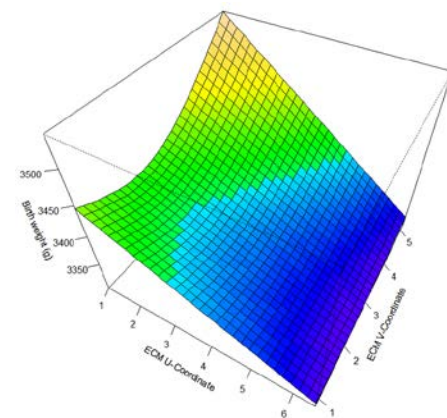
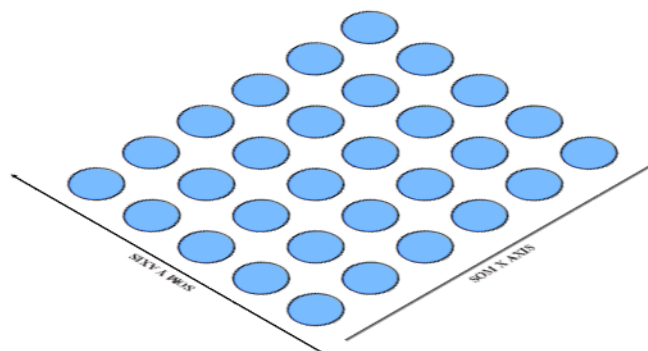
Exposome Data Challenge Event

28-30 April 2021



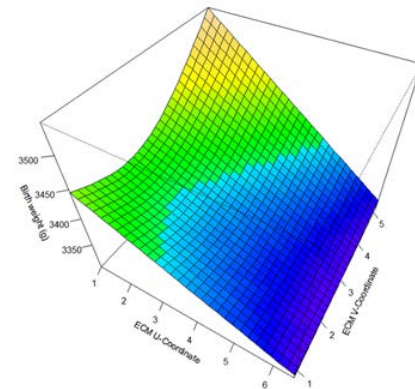
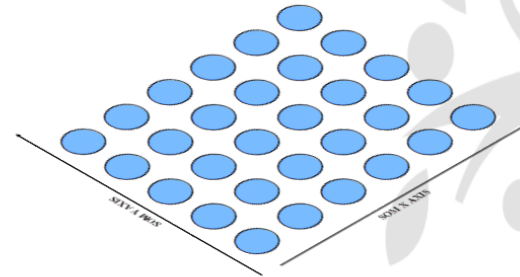
John L. Pearce, PhD
 Department of Public Health Sciences
 Medical University of South Carolina
 Email: pearcejo@musc.edu

Exposure Continuum Mapping for predicting health and disease in Exposome studies



What is Exposure Continuum Mapping (ECM)?

- Addressing the Exposome presents many, many challenges
 - Large data
 - Unique/Rare exposure scenarios
 - Outliers
- We address as a *spatial* problem
 - Spatially-correlated learning is used discover and describe patterns
 - Spatial pooling improves inference
 - Maps provide data visualization
 - Map location is used to assess risk
- Exposure Continuum Map (ECM)
 - A 'spatially' organized exposure model
 - Surface reflects continuous sequence of profiles where adjacent scenarios are similar and extremes are quite distinct
 - Leverage map coordinates for health effects estimation



How do we apply ECM?

- ECM involves two-stages:
 - Step 1: Exposure Mapping
 - *Self-organizing map (SOM)*
 - Identify 'profiles' that reflect prenatal exposure to EDC mixtures
 - Mapping arranges profiles as a continuum of exposure features- the ECM
 - Step 2: Health Effects Mapping
 - *Generalized Additive Model (GAM)*
 - ECM coordinates (U,V) used to construct joint-dose response
 - Tensor product smooth function

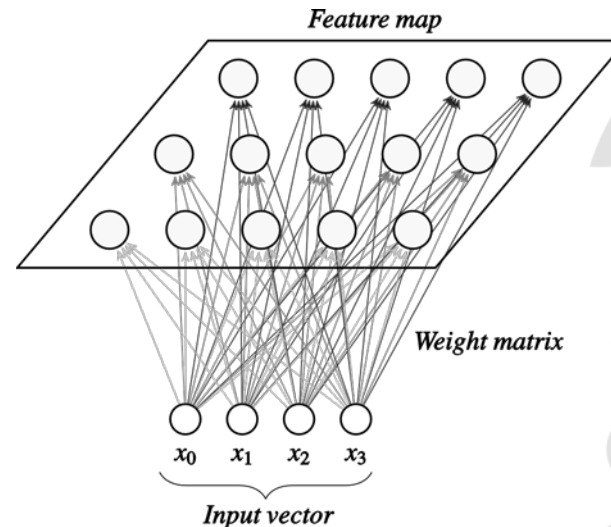


Figure 1. Architecture of the SOM-based exposure continuum map

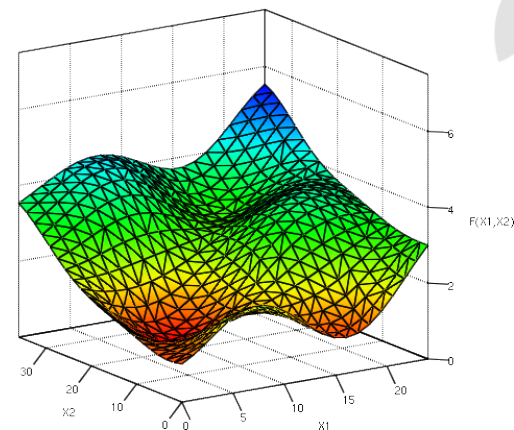
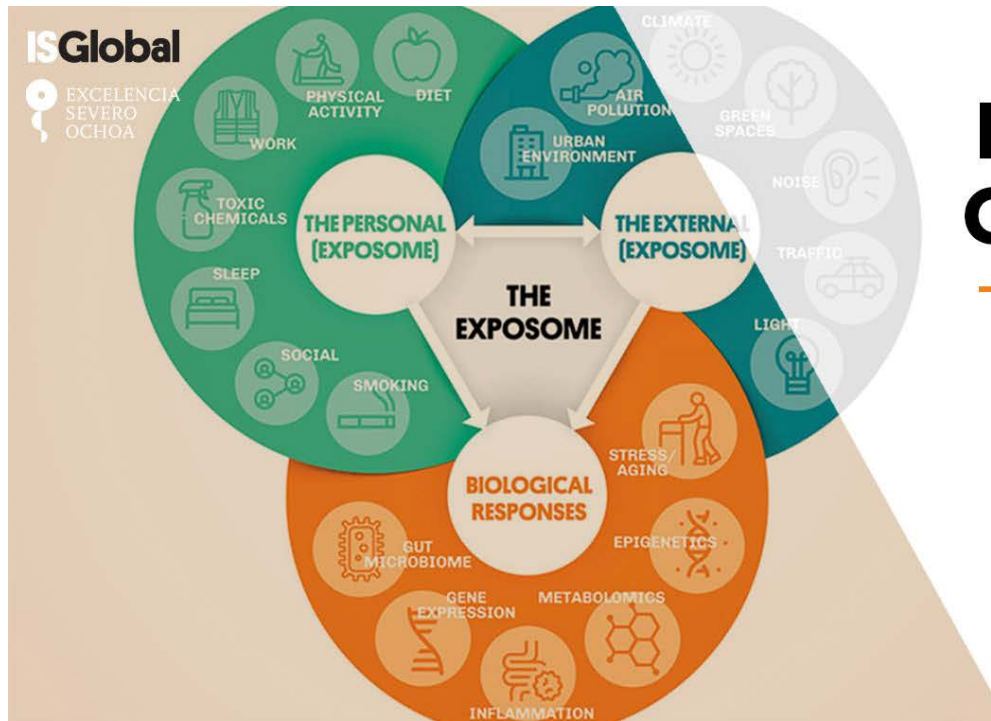


Figure 2. Hypothetical joint-dose response function

Exposome Data Challenge Event

28-30 April 2021



Simulation Study

Exploring associations between prenatal exposure to multiple environmental exposures and birth weight with exposure continuum maps

Study Data

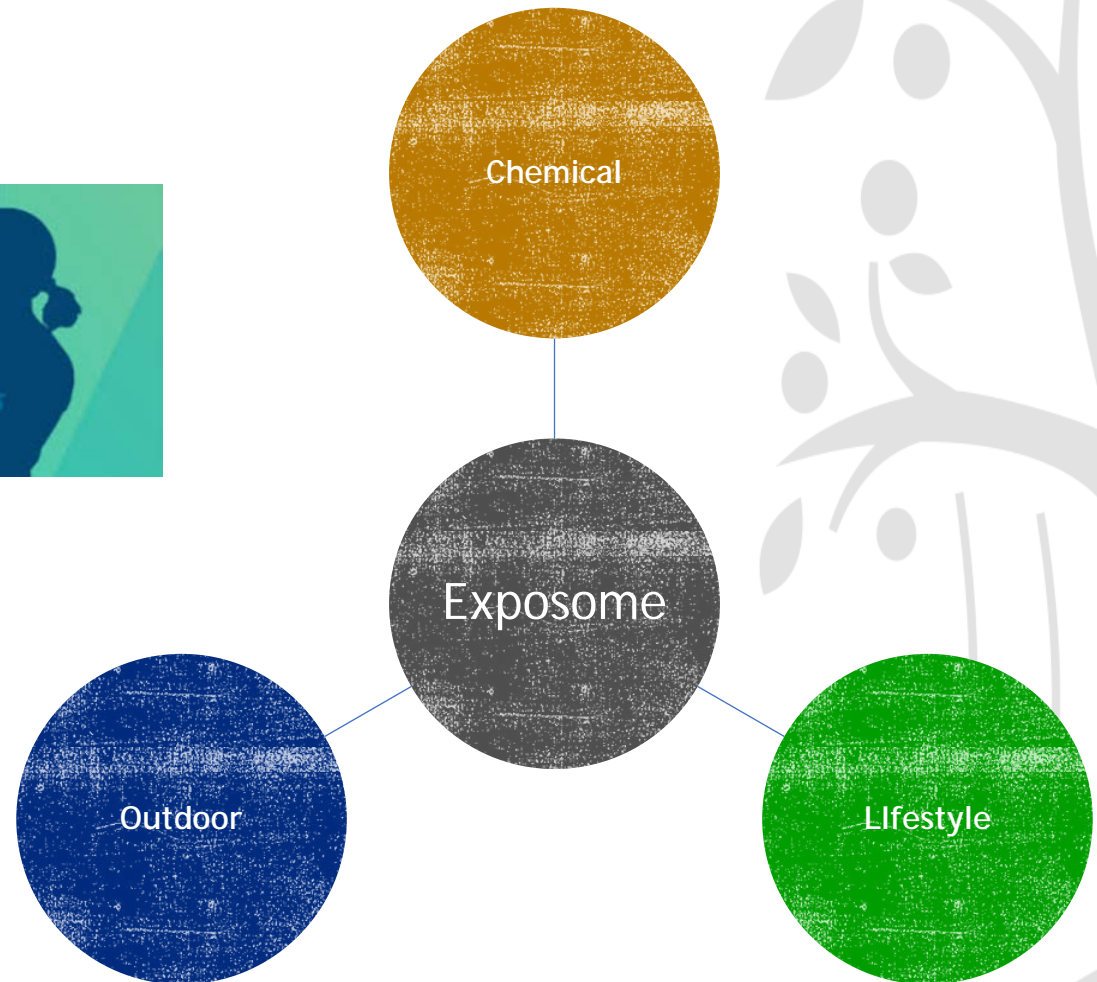
- HELIX EXPO Cohort

- N=1301 mother/child pairs
- 2003 to 2009
- Multiple sites



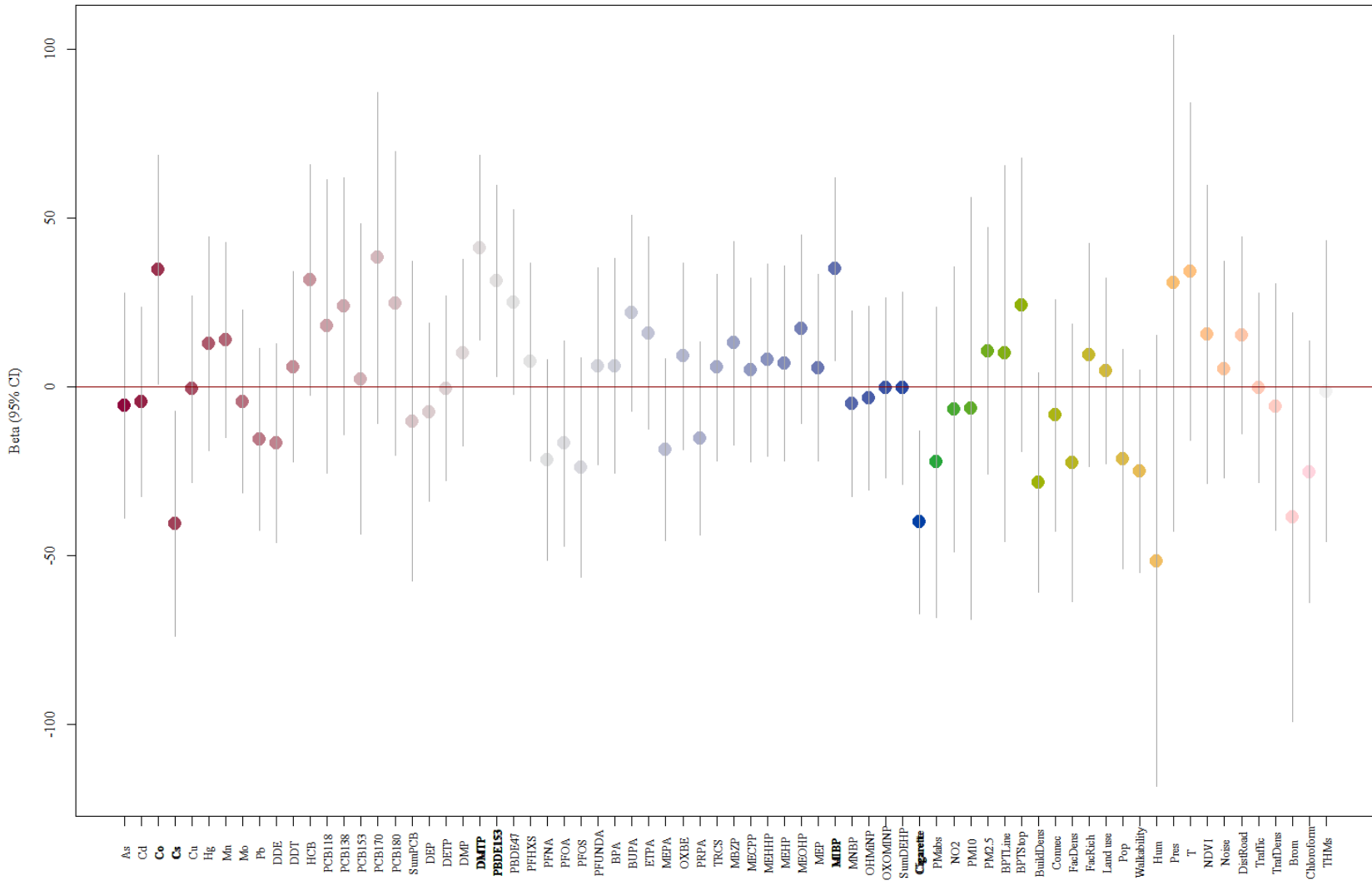
- Analytic data

- Exposome
 - Prenatal Exposures
 - 72 Continuous Exposures
 - 50 Chemical
 - 26 Outdoor
- Covariate
 - Site, Age, Edu, BMI, etc.
- Phenotype
 - Birth Weight



Single Pollutant Models

$$y = \text{cohort} + m_{\text{age}} + m_{\text{edu}} + m_{\text{native}} + m_{\text{parity}} + m_{\text{BMI}} + c_{\text{sex}}$$



*Bold indicates p-value < 0.05

Making the Exposure Continuum Map

- Map Construction

- Step 1: Data Standardization

- Mean=0, unit=SD

- Step 2: Selecting Map Size

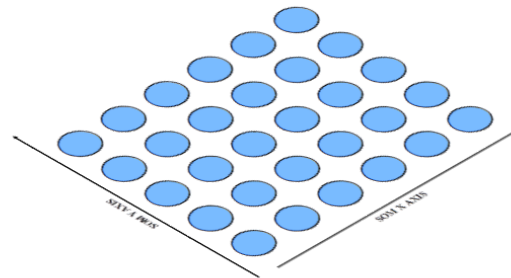
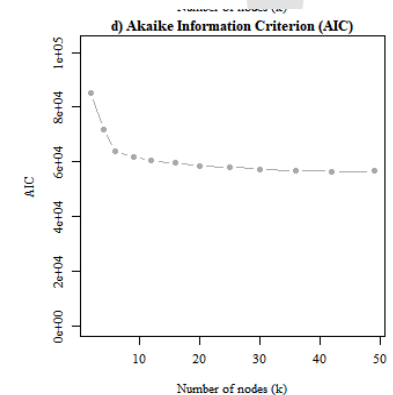
- R², RMSE, AIC, N, WCSS

- Step 3: Map Construction

- Initializations
- Learning parameters

- Step 4: Map Visualization

- Profile Codes
- Arrangement
- Labeling



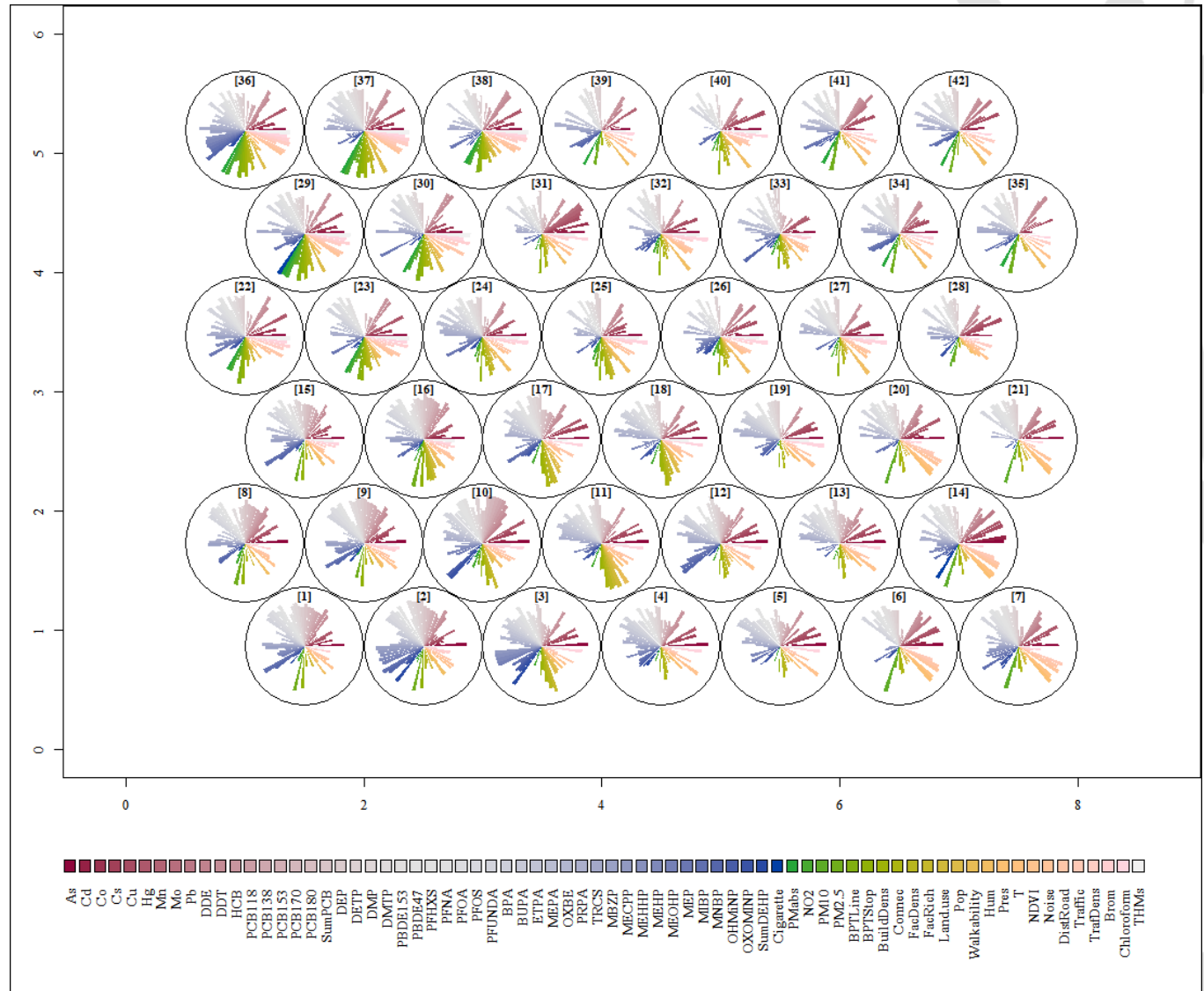
Results: Exposure Continuum Map (ECM)

- Exposure profiles

- Patterns reflect feature patterns discovered in data
- Radial bar reflect normalized values
- Proximity reflects similarity

- ‘Spatial’ Distribution

- Higher values on extremes
- Low-modest near center



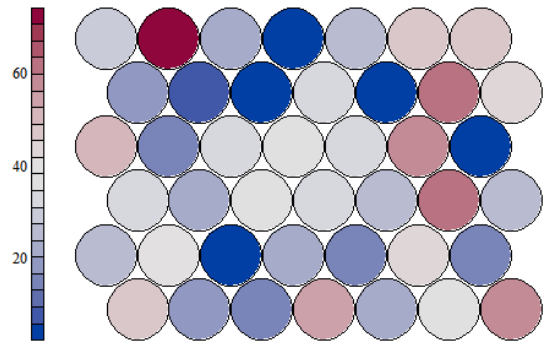
Results: Mapping Characteristics

- The 'map' affords many helpful visualizations

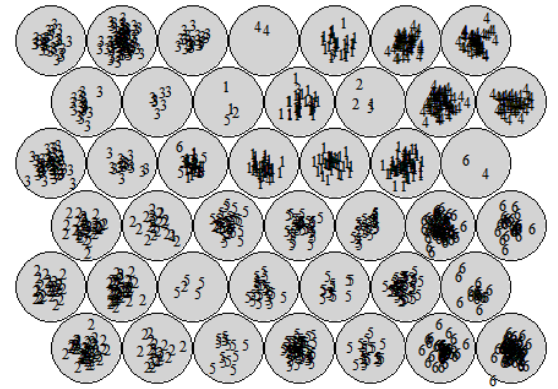
- Frequency distributions
 - Population subgroups

- Profile Characteristics
 - Within-class error
 - Between-class differences

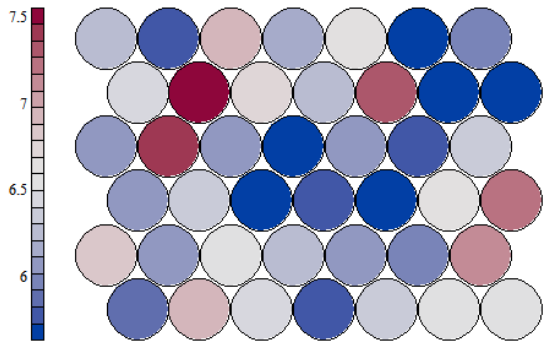
a) Participant Frequencies



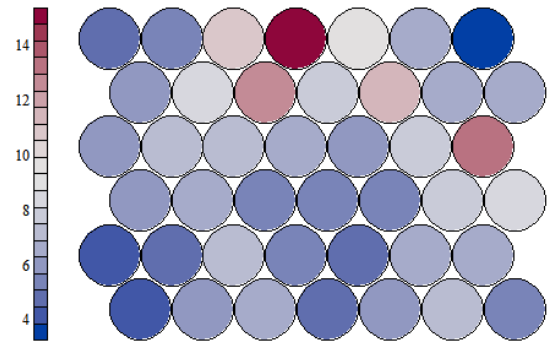
b) Cohort Assignments



c) Profile Error



d) Neighborhood Distances

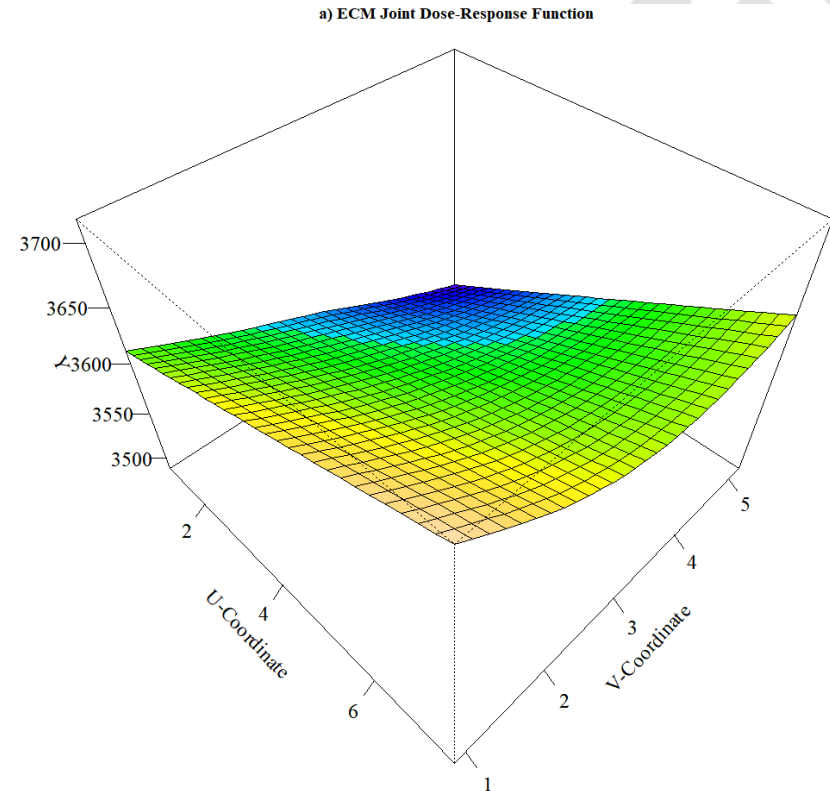


Results: Mapping Individual Components

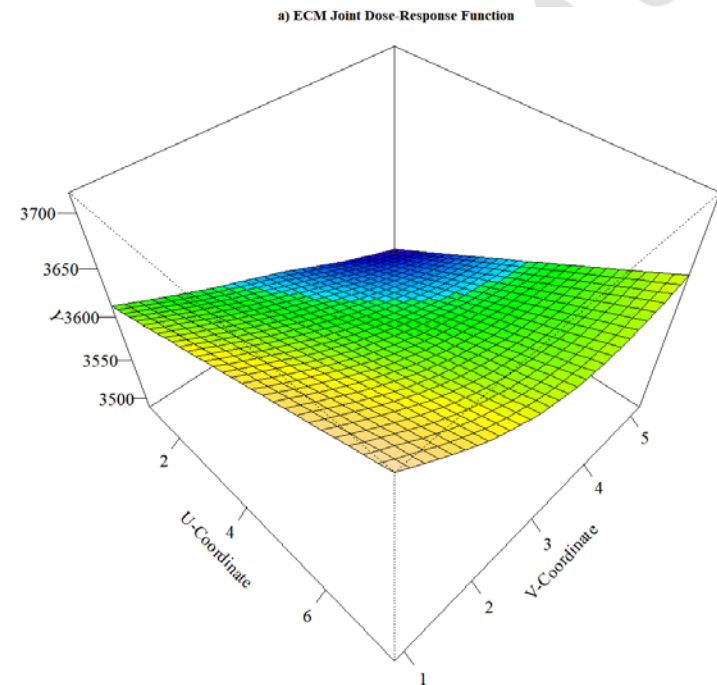
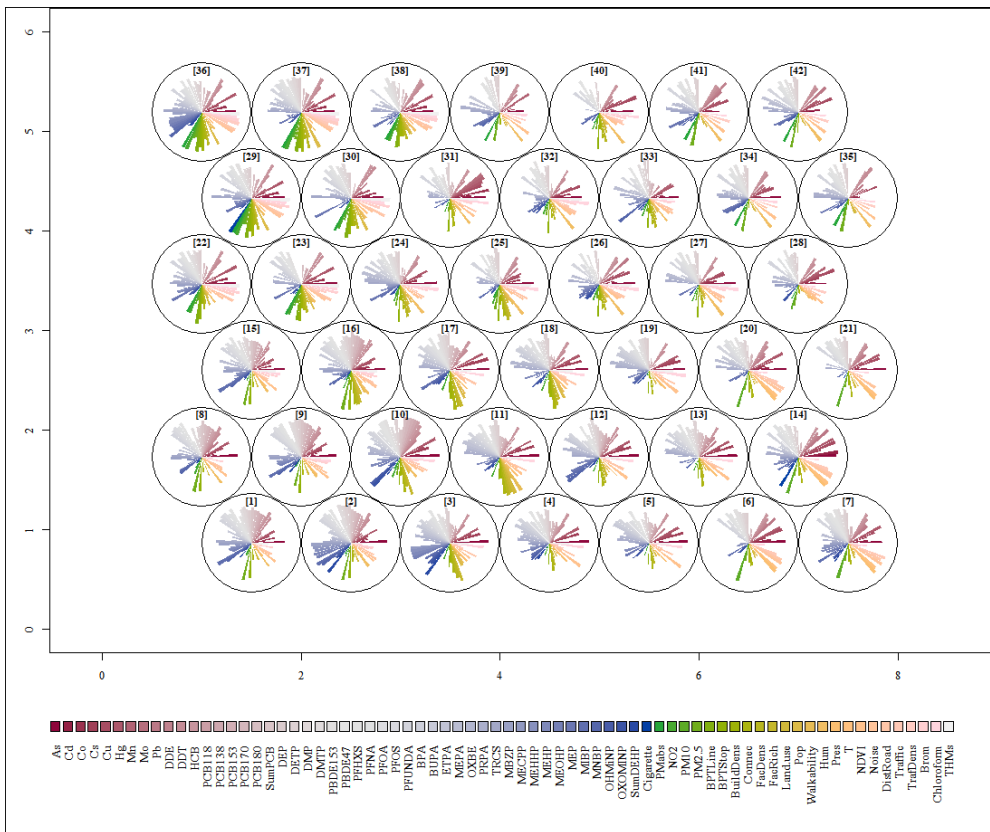


Exploration of Health Effects

- The objective is to assess the response of birth weight across variations in the total exposome
- Construct **generalized additive model** that controls for maternal age, race, education, and infant sex
- $y = cohort + m_age + m_edu + m_native + m_parity + m_BMI + c_sex \dots + \varepsilon$
- ECM (U,V) coordinates modeled as a *tensor product smooth function* in order to assess **joint** effect total mixture
 - $f(U,V)$
- Results provide a ‘dose-response’ function for total exposome



Results: Health Effects



- Lower birthweights corresponded to locations in the upper left corner of the map
 - Higher air pollution, pesticides, and measures of urbanicity
- Higher birthweights occurred in the bottom right and lower axis of the map
 - Hot, dry climate with elevated PM10. Few chemicals higher (e.g., BUPA)

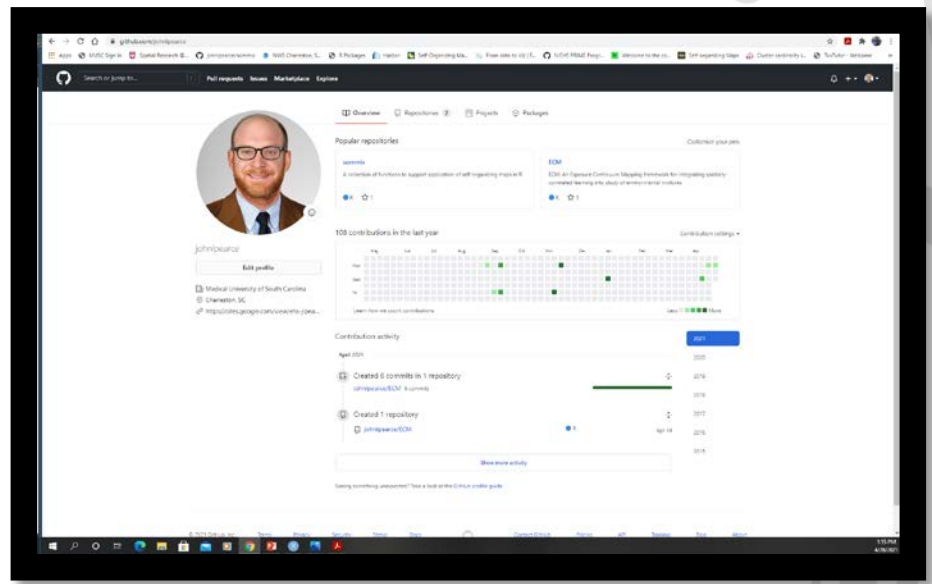
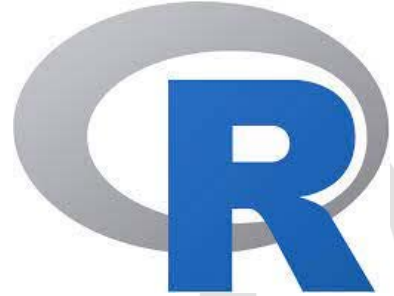
Overall Summary

- Study of Exposome is very difficult and will require multiple tools and perspectives to address
- ECM provides an attractive framework for studies of the Exposome
 - SOM-based exposure characterization is powerful tool
 - GAM-based health effects provide proven regression framework that can handle multiple outcome distributions and complex covariate structures
 - Support hypothesis generation
 - Visualization is a highlight and computational speed is also a benefit
 - Limitations
 - Interpretation of total exposome is difficult
 - Integration of variable selection/importance could help tease out most important components



Resources

- ECM R package currently available on GitHub
- <https://github.com/johnlpearce/>
- Manuscript under review at *Environmental Research*



Acknowledgements

Collaborators:

- Medical University of South Carolina
 - Brian Neelon
 - Alex Alekseyenko
- Johns Hopkins
 - Jessie Buckley
- Dartmouth University
 - Megan Romano
 - Brett Doherty

Funding:

- NIH ECHO
 - U2COD023375 (Smith)
- NIEHS
 - R00ES023475



QUESTIONS???



Exposure Data – Correlation Analysis

- Correlation matrix illustrates complexity of exposure data
 - Majority of correlations are positive
 - Correlations strongest within each exposure domain

