UW CCAR Scientific Advisory Committee Meeting

October 6 & 7, 2014

CCAR
Center for Clean Air Research

UW wireless connection
NetID: **event0521**
Password: **34rW-67zT-96dN**

SCHOOL OF PUBLIC HEALTH
UNIVERSITY of WASHINGTON

COLLEGE OF ENGINEERING
UNIVERSITY of WASHINGTON

Lovelace

WASHINGTON STATE UNIVERSITY

THE UNIVERSITY of NEW MEXICO

CCAR
Center for Clean Air Research

the SAC

John Balmes (chair)

Jesus Araujo

Mike Brauer

Brent Coull

Ian Gilmour

Nick Mills

Tom Peters

Arden Pope

Sanjay Rajagopalan

Barbara Turpin

the investigators

Sverre Vedal (PI)

Matt Campen

Tom Jobson

Joel Kaufman

Jake McDonald

Tim Larson

Mike Rosenfeld

Paul Sampson

Lianne Sheppard

Chris Simpson

Adam Szpiro

Tim VanReken

Mike Yost

CCAR projects & cores

Project 1
roadway exposure characterization

M Yost (PI), T Larson, C Simpson, T Jobson, T VanReken

Project 2
exposure atmosphere generation

J McDonald (PI), T Larson

Project 3
toxicology

M Campen (PI), M Rosenfeld, J McDonald

Project 4
human clinical studies

J Kaufman (PI)

Project 5
epidemiology cohort study

J Kaufman (PI), S Vedal

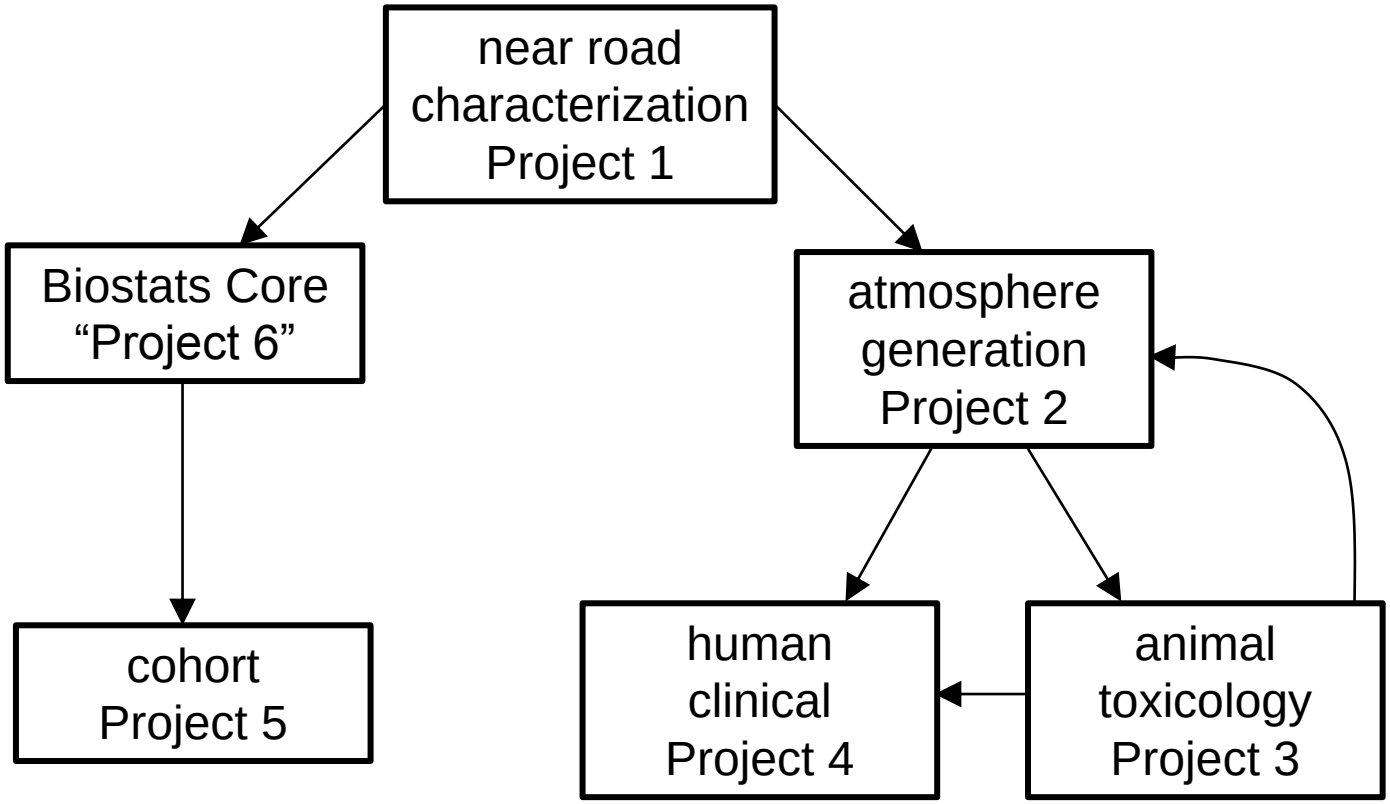
Project 6
multipollutant exposure modeling

L Sheppard (PI), A Szpiro, P Sampson

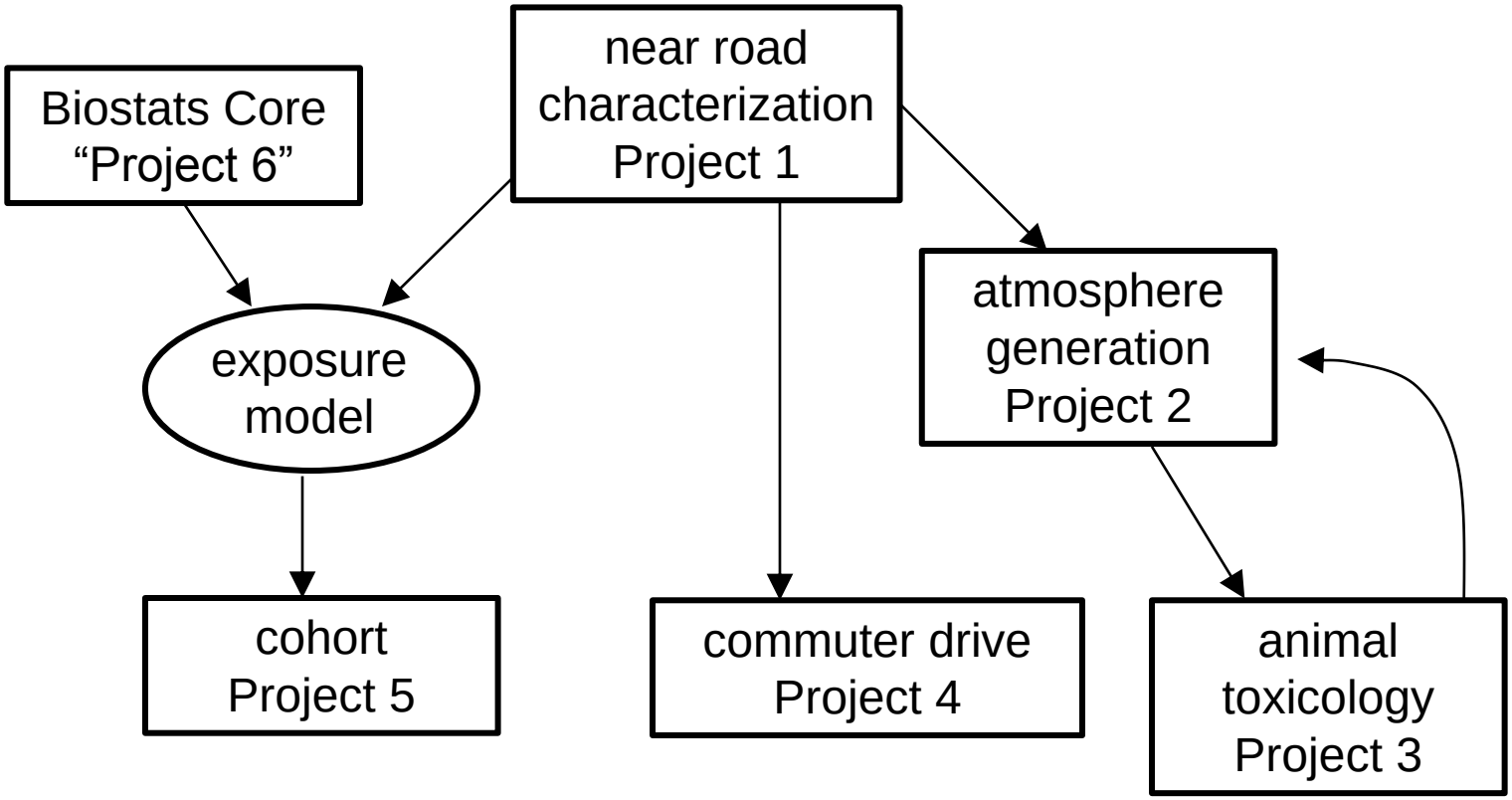
Biostats Core

Admin Core

project integration: proposed

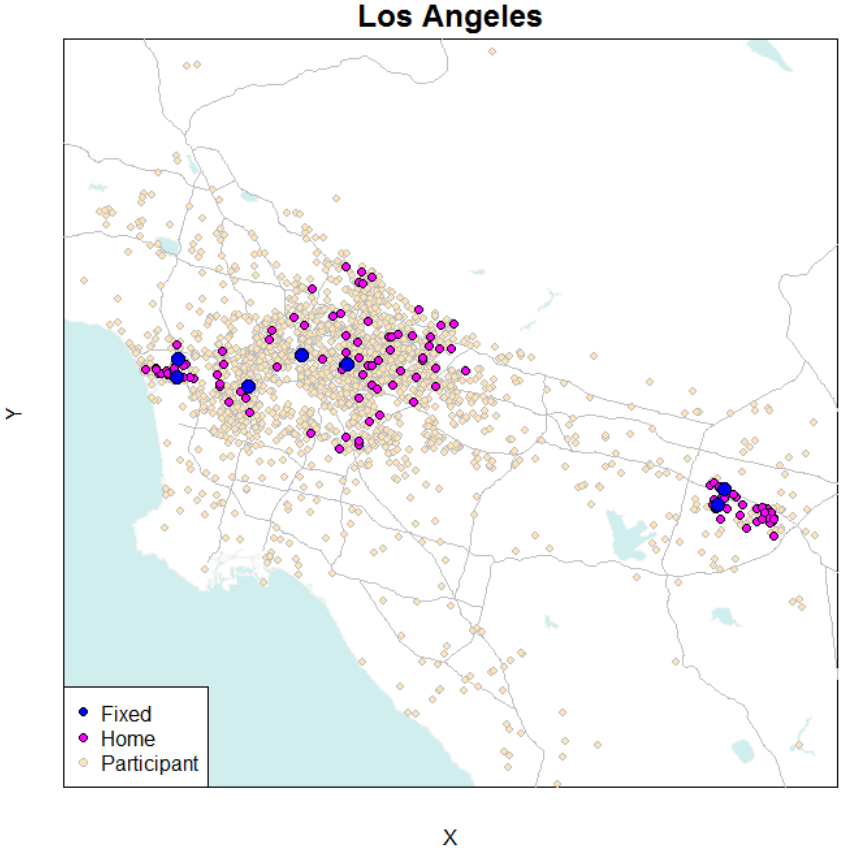


project integration: current



MESA Air/ NPACT monitoring and cohort locations

- Monitoring campaign dedicated to MESA cohort
- 2-week samples for 2005-2009
- 3-7 fixed sites and about 50 rotating home-outdoor sites in each of two seasons



Since we last met:

1. responses to SAC review

2. Clean Air Research Centers (CLARC) and CCAR

    - EPA center webinar, January 2014 – R Jandarov & A Szpiro (Biostatistics Core)

    - CLARC annual meeting, Atlanta, Sept 18-19, 2014

    - Planned papers to highlight CLARC work
        - Workshops and collaborations

Since we last met:

3. Projects
   - P1 – Atlanta collaboration monitoring; LRRI chamber analyses (with P2), multiple approaches to mobile data
   - Biostatistics Core – sparse P-PCA, monitor network and health effects; utilizing mobile monitoring data
   - P2 – multiple atmospheres
   - P3 – proteomics, metabolomics, serum bioactivity
   - P4 – study design & commute, outfitted vehicle for air filtration, recruited subjects
   - P5 – in transit micro-environmental monitoring: two seasons completed in Winston-Salem and Los Angeles

SAC input especially on:

1. Using the mobile monitoring data for generating exposure predictions for our cohort study
2. How to best make use of micro-environmental monitoring data
3. Health endpoints for the cohort study (epigenetics and gene expression)
4. Design, methods, and endpoints for the commute drive
5. New directions for toxicology
6. AMS characterization of exposure chamber atmospheres

AM - Monday, October 6, 2014 (UW Tower Boardroom)		
Time	**Topic**	**Speaker**
9:45-10:00	Overview of UW CCAR, Meeting Objectives and Agenda	Vedal
10:00-10:25	Mobile monitoring update (Project 1)	Yost
10:25-10:45	Chamber characterization – AMS findings (Projects 1 & 2)	Jobson
10:45-11:05	Micro-environmental exposures (Project 5)	Hazlehurst
11:05-12:05	Using the mobile monitoring data for epidemiology (Project 1 & Biostatistics Core) – Part 1: -- Preparing the data: 2 approaches -- Plans for analysis	Riley, Austin Sheppard
12:05-1:00	**LUNCH + poster viewing**	

PM - Monday, October 6, 2014 (UW Tower Boardroom)		
Time	**Topic**	**Speaker**
1:00-1:30	Using the mobile monitoring data for epidemiology (Project 1 & Biostatistics Core) – Part 2: -- Plans for analysis (continued)	Sheppard / Szpiro
1:30-2:00	Commuter drive (Project 4)	Kaufman
2:00-2:40	Cohort study: design and health endpoints (Project 5)	Kaufman / Chi
2:40-2:55	**Coffee Break**	
2:55-3:40	Controlled exposure metabolomics and serum bioactivity (Project 3)	Campen
3:40-4:05	Special topic: Tutorial on UW spatio-temporal modeling	Sampson
4:05-4:25	Center Collaboration Projects	Vedal & Others
4:25-4:55	General Discussion of Day 1	
6:00-8:00	**Dinner at Portage Bay Cafe**	

Tuesday, October 7, 2014 (Watertown Hotel)		
Time	**Topic**	**Speaker**
8:00-10:30	SAC Closed Discussion	
10:30-11:45	SAC Report, Recommendations and Discussion	
11:45-11:50	EPA Closing Remarks	Costa
11:50-12:00	CCAR Final Comments	Vedal
12:00	Adjourn	

overview (selected) of SAC comments:

- integration of mobile and chamber characterization data, and with experimental and observational exposures
- distinguishing roadway pollution from other sources
- more sensitive toxicologic endpoint(s)
- scripted study – prefer crossover study
- drop use of “eigenpollutant”; keep in mind spatial scales of contrasts

CCAR Project 1: Mobile Monitoring Update

Mike Yost

10/6/2014

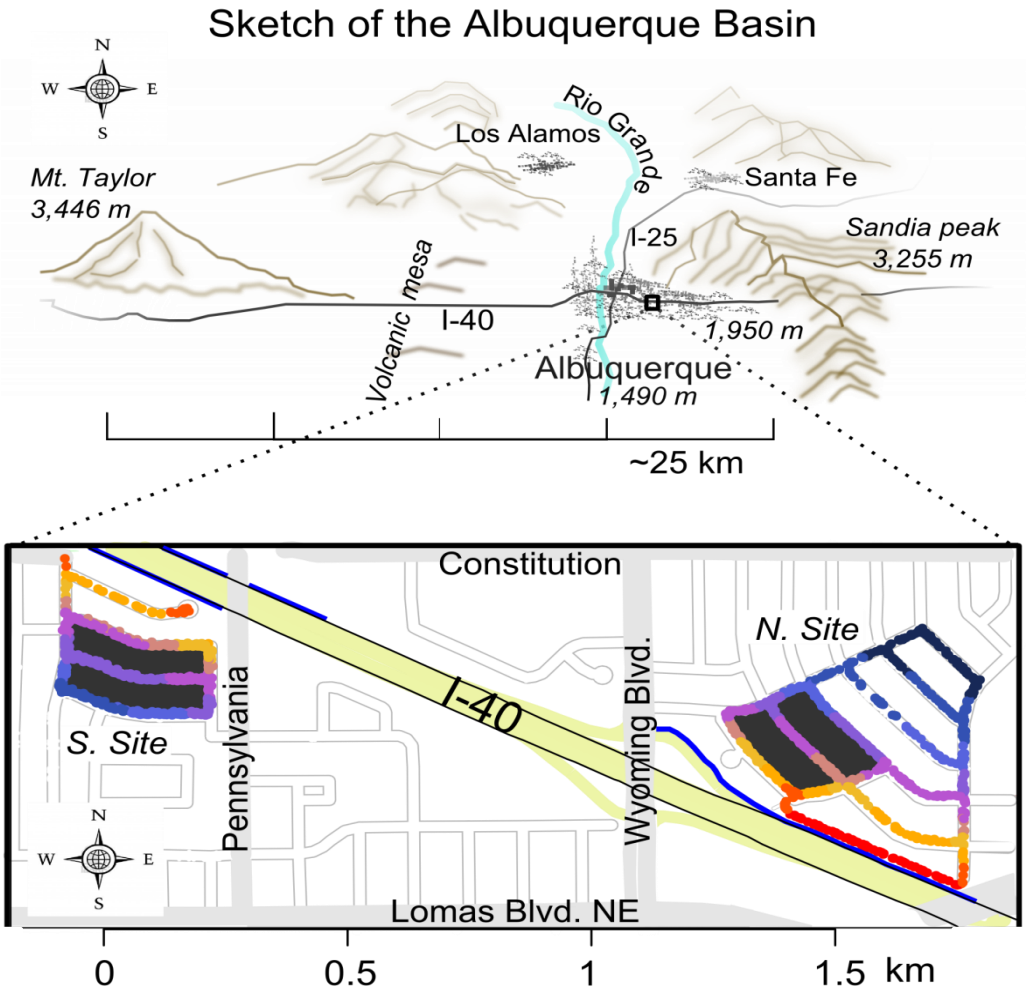
Outline

- Introduction
- Some applications of mobile monitoring
  - #1 Characterization of small-scale spatial gradients
  - #2 Real time video analysis of multi-pollutant data
  - #3 Cluster analysis of fuzzy point locations
- Are fuzzy points just high-traffic intersections?
- Conclusions /summary thoughts

Some important features of mobile monitoring

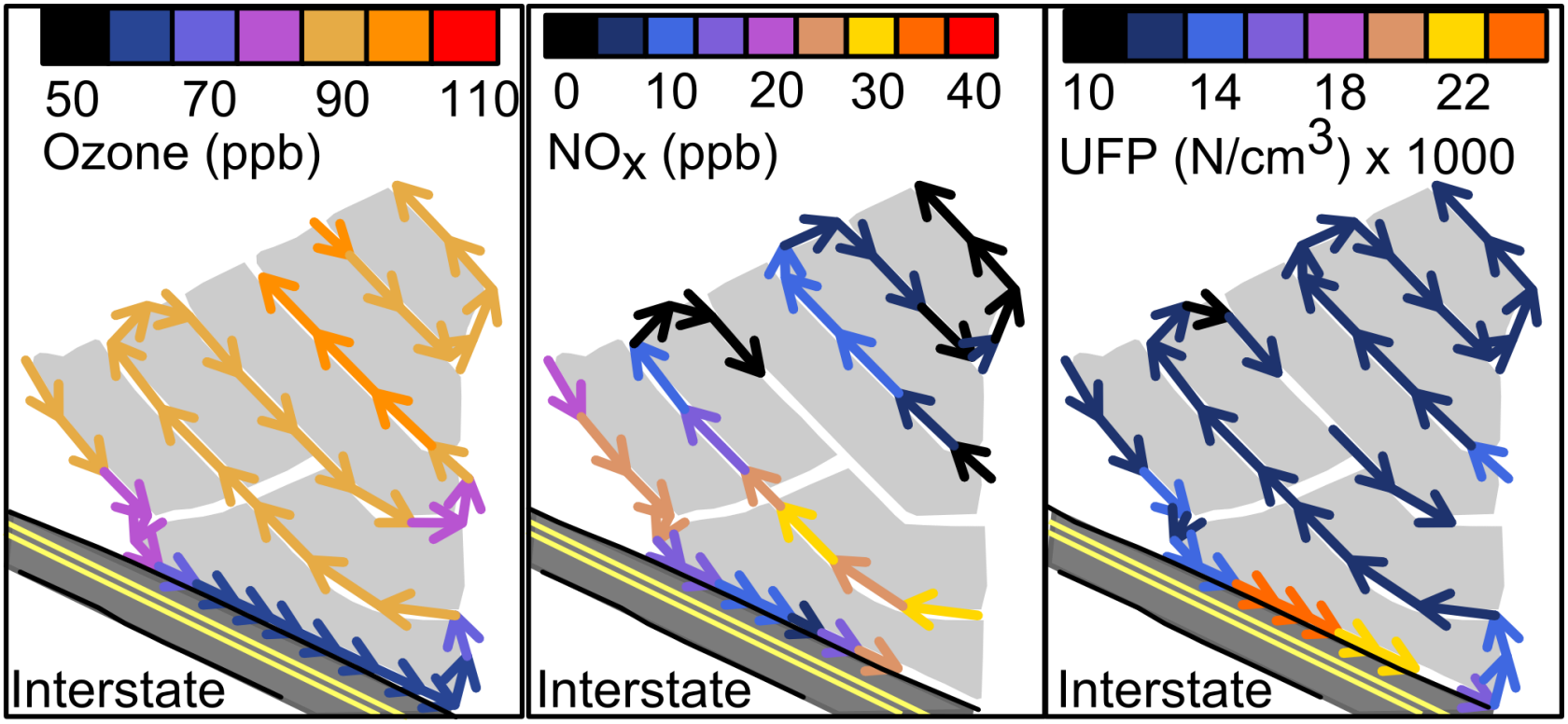
- Inherently a spatio-temporal sampling scheme
- Data are spatially diverse and temporally sparse
- Simultaneous multi-pollutant data
- Data are sampled over multiple scales of time and space

#1 Multi-pollutant characterization of a near roadway gradients



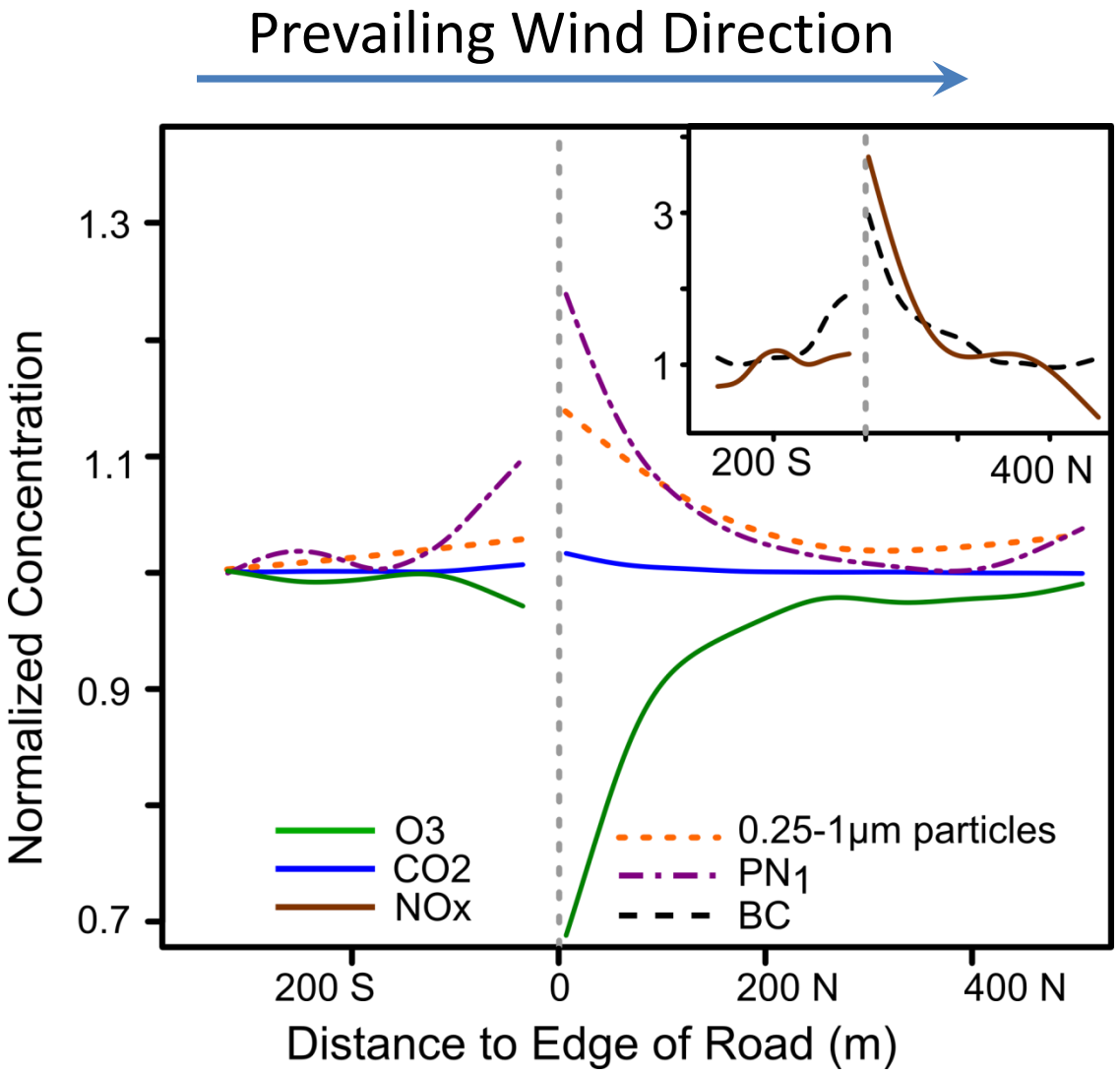
Riley, E.A., Banks, L., Fintzi, J., Gould, T.R., Hartin, K., Schaal, L., Davey, M., Sheppard, L., Larson, T., Yost, M.G., Simpson, C.D., Multi-pollutant mobile platform measurements of air pollutants adjacent to a major roadway. Atmos. Environ. DOI: 10.1016/j.atmosenv.2014.09.018

#1 Gradient sampling route was driven in alternating directions and recorded 10-second concentrations

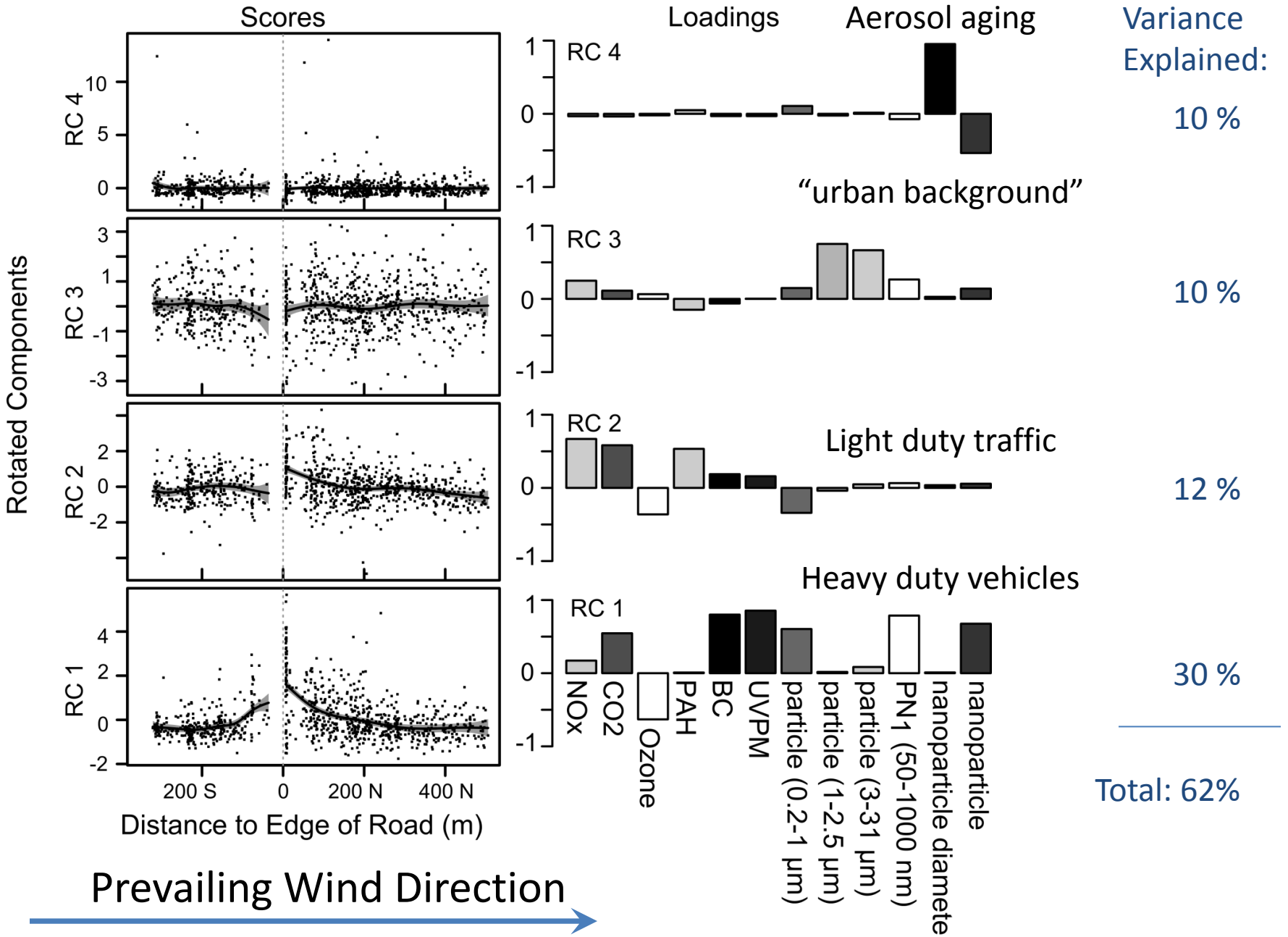


#1 Single Pollutant Gradients

- 10 second measurements
- Adjusted for background using data > 250 m from roadway.
- Modeled mean normalized by campaign background
- Principal Component Analysis (PCA) on 10 s measurements.



#1 Multi-pollutant gradients



#2: On-roadway multi-pollutant characterization

- Data in Industrial Seattle
- 1 day of data, four drives of route

30 s average of 10 s measurements

↓

"Active Subset" made by finding peaks in pollutant concentrations (> 2 s.d. threshold above 95% trimmed mean)

↓

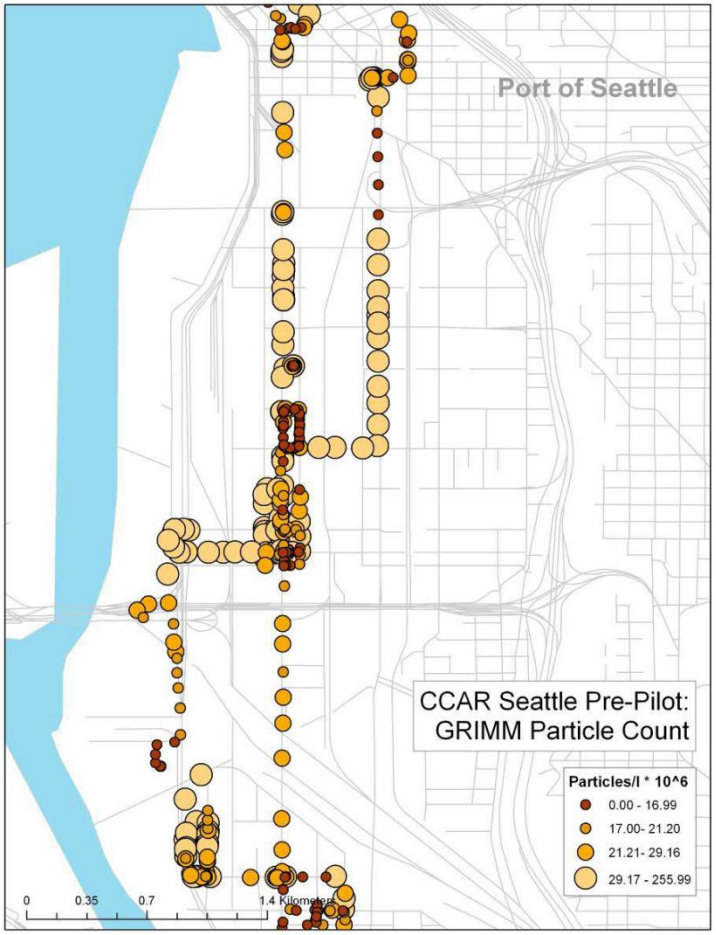
Summarize video at 1 min around peaks into categories: road type and traffic composition.

↓

Compute PCA on "peaks" subset of data

↓

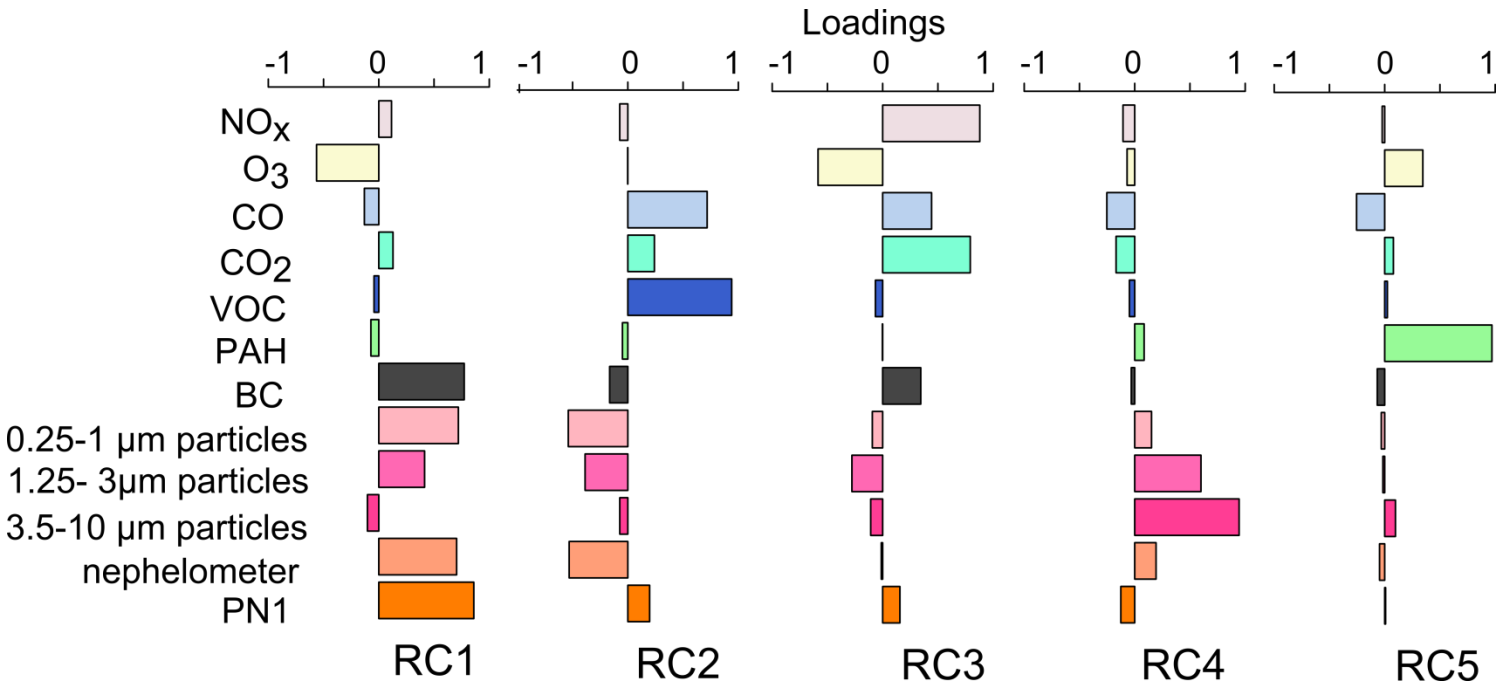
Calculate average score of components for video categories

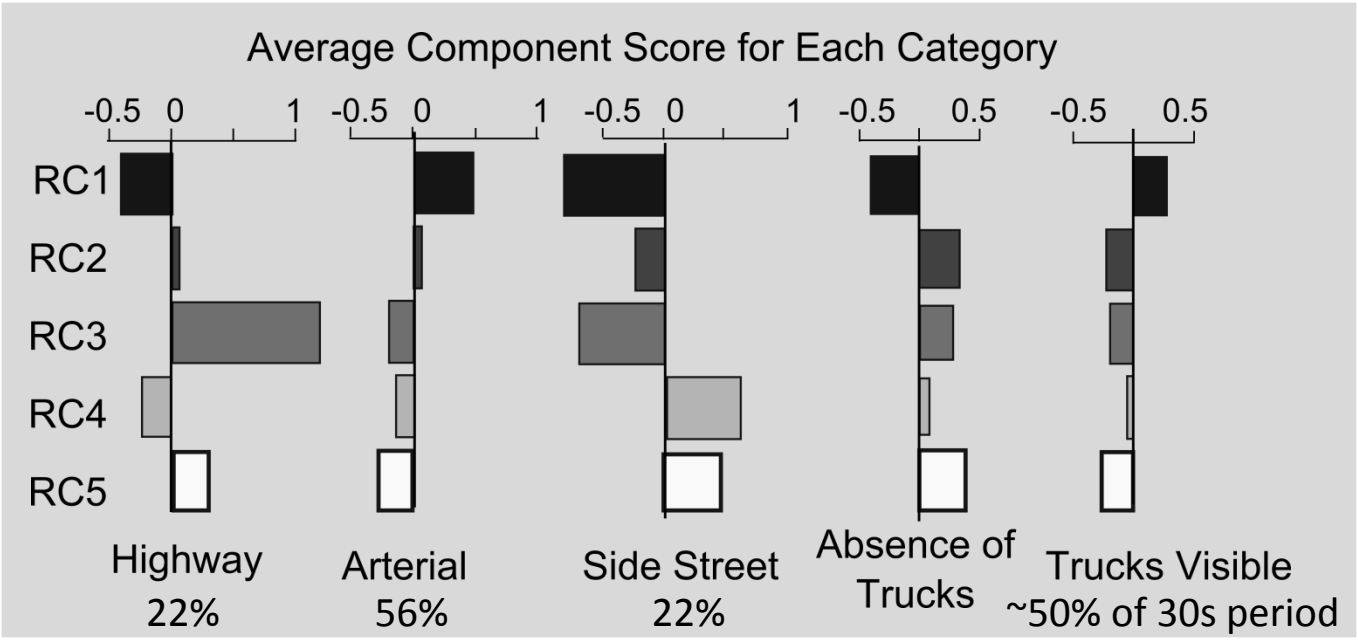


#2 Analysis of DEEDS data using PCA and video

- "Active peak subset": measurements for which two or more instruments have values at least two standard deviations away from their means.
- Video analysis of 64 thirty-second average values (~13% of the data)
  - Video inspected *prior* to statistical analysis. Description logged for one minute preceding the time stamp, and thirty seconds after.
  - PCA with varimax rotation, five components retained.

#2 Multi-pollutant features derived from the DEEDS “peaks” subset





#3 Cluster Analysis Baltimore

- Describe spatial and temporal features of the CCAR mobile monitoring data in Baltimore, MD

- Present a method to identify multivariate patterns in this data using cluster analysis
  - Important features
  - Identify locations with similar multivariate pollutant distributions

#3 Cluster Analysis Baltimore

30 s average data @ fuzzy points

↓

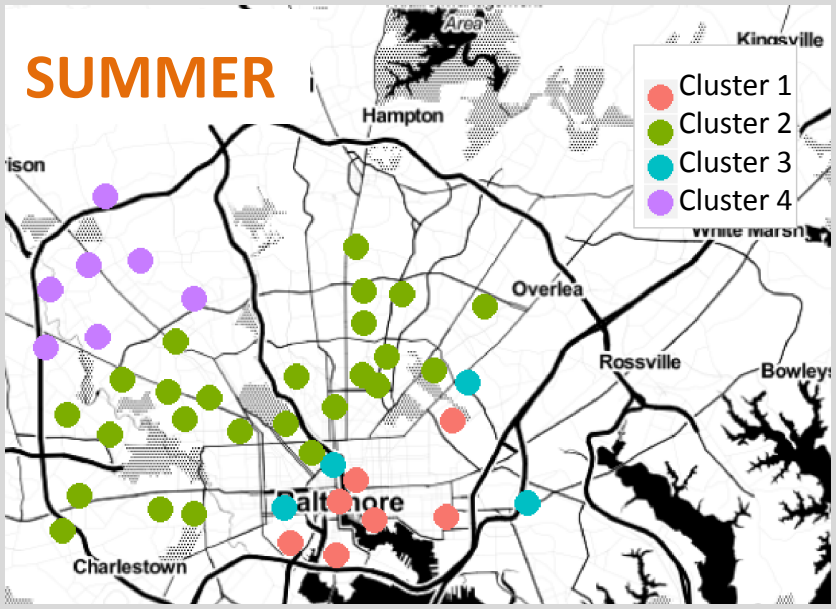
Temporally corrected for changes in Background concentration (see poster)

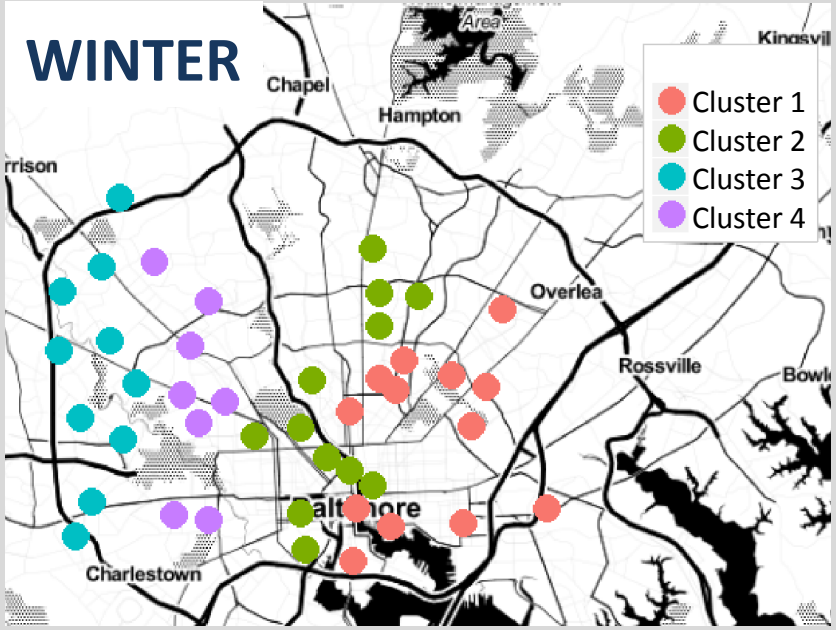
↓

Converted Z-score

↓

K- means clustering algorithm





#3 Cluster Analysis Baltimore:
Clusters by pollutant species

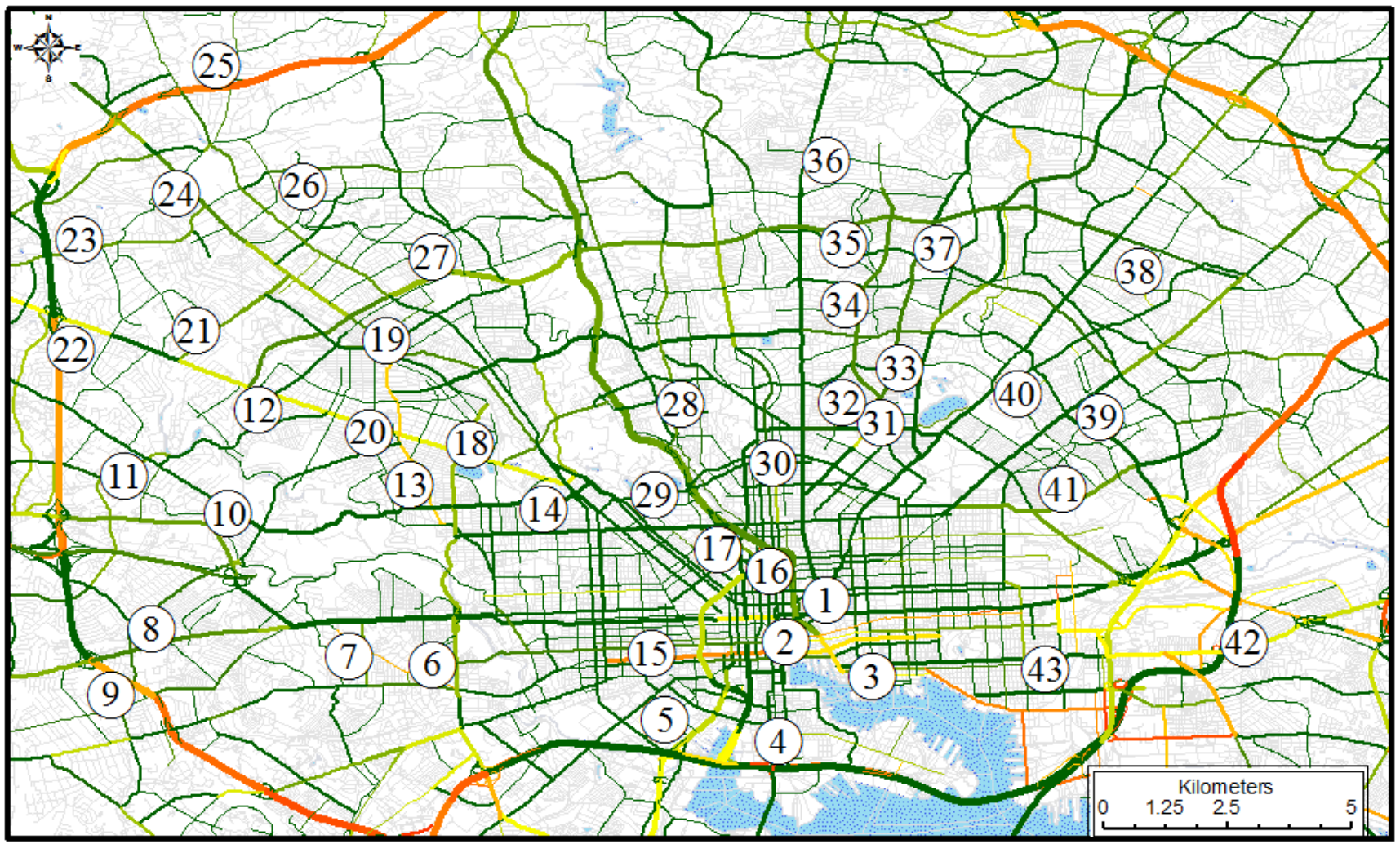
Z-scores

WINTER										
	OZONE	**NO2**	**Nox**	**PN1**	**PM2.5**	**UF**	**Fine**	**VOC**	**BC**	**PAH**
Cluster 1	-1.0	0.2	0.4	1.0	1.2	1.2	1.1	1.0	0.9	-0.6
Cluster 4	-0.4	1.0	0.8	0.4	0.0	0.1	0.4	-0.1	0.1	0.4
Cluster 2	1.2	-0.8	-1.2	-1.3	-0.9	-0.9	-1.3	-0.9	-0.9	-0.8
Cluster 3	0.8	-0.9	-0.3	-0.7	-0.8	-0.9	-0.6	-0.3	-0.5	1.5
SUMMER										
	OZONE	**NO2**	**Nox**	**PN1**	**PM2.5**	**UF**	**Fine**	**VOC**	**BC**	**PAH**
Cluster 1	-1.2	1.8	1.8	1.9	0.8	-0.5	1.3	1.1	1.9	-1.6
Cluster 5	-0.2	0.7	0.4	0.9	1.1	0.0	1.1	0.9	0.9	0.5
Cluster 2	-0.1	-0.5	-0.1	-0.4	-0.5	-0.3	-0.1	0.1	-0.5	0.2
Cluster 3	1.7	-0.3	-1.5	-1.1	0.2	1.7	-1.6	-1.8	-0.8	0.8

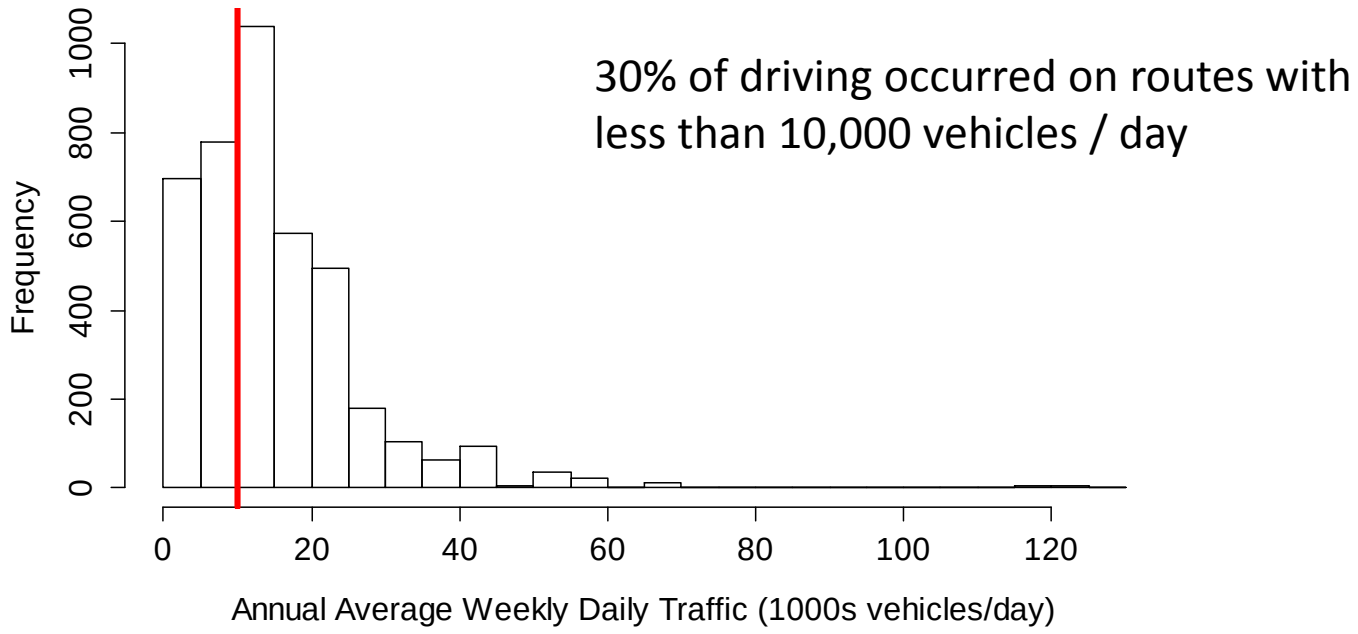
Are Fuzzy Points just high-traffic intersections?

- Mobile monitoring captures a distribution of multi-pollutant values around fuzzy points
- Typical spatial scale of fuzzy points ~300m
- Mobile data sample a variety of covariates, such as traffic, road conditions and housing density
- Mobile data depend on at least 2 local factors, type of roadway, and the traffic conditions
- Mobile data also appear to depend on larger urban /suburban /regional features (mixing, transport, atm. chemistry, etc.)

Baltimore, MD



Traffic Distribution in Baltimore on Mobile Monitoring Routes



Features in DEEDS video results

Characteristic	Occurrence
Buses visible in the frame	20% of measurements
Heavy trucks visible in the frame	60% of measurements
% of time trucks/buses are visible	~50% of each 30-sec period on avg.
Density of car/light truck traffic	14% none; 27% light; 47% medium; 9% heavy
Max number of lanes in each direction	2.47 on average
Accelerating from stop behind cars	23% of measurements
Accelerating from stop behind trucks	27% of measurements
Dominant road type	22% highway; 56% arterials; 22% side roads
Platform stopped at red light	30% of measurements
Platform stopped at stop sign	6% of measurements
Platform stopped at intersection	36% of measurements
Uphill gradient visible	19% of measurements
Downhill gradient visible	13% of measurements

Type of Traffic

Type of Roadway

Conclusions and Significance

- Single and multi-pollutant gradients are captured
- Traffic related multi-pollutant features generally are consistent across 3 cities, and manifest different roadway /source influences
- Multivariate clusters of fuzzy point monitoring locations identify distinct regions of urban air quality
- Fuzzy point data capture multivariate distributions in space and time
- Departures from background represent a possible pathway to test downscaling of air quality models

Thank You!

CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

EPA Clean Air Research Center

Project 1:

Aerosol Characterization of LRRI Exposure Chamber

External Science Advisory Meeting October 2014

Investigators:
Tom Jobson, Tim VanReken, Courtney Herring, Matt Erickson,
Mylene Gueneron, WSU
Michael Yost, Tim Larson, Chris Simpson, UW
Jake MacDonald, LRRI

Lovelace Respiratory Research Institute Exposure Chamber Study

April – May, 2012

Task 1. Characterize gas and particle composition in the 1-m^3 engine exhaust exposure chambers. Sample mixtures of diesel and gasoline engine exhaust.

Task 2. Characterize 11.5-m^3 Teflon chamber for engine exhaust irradiation → SOA

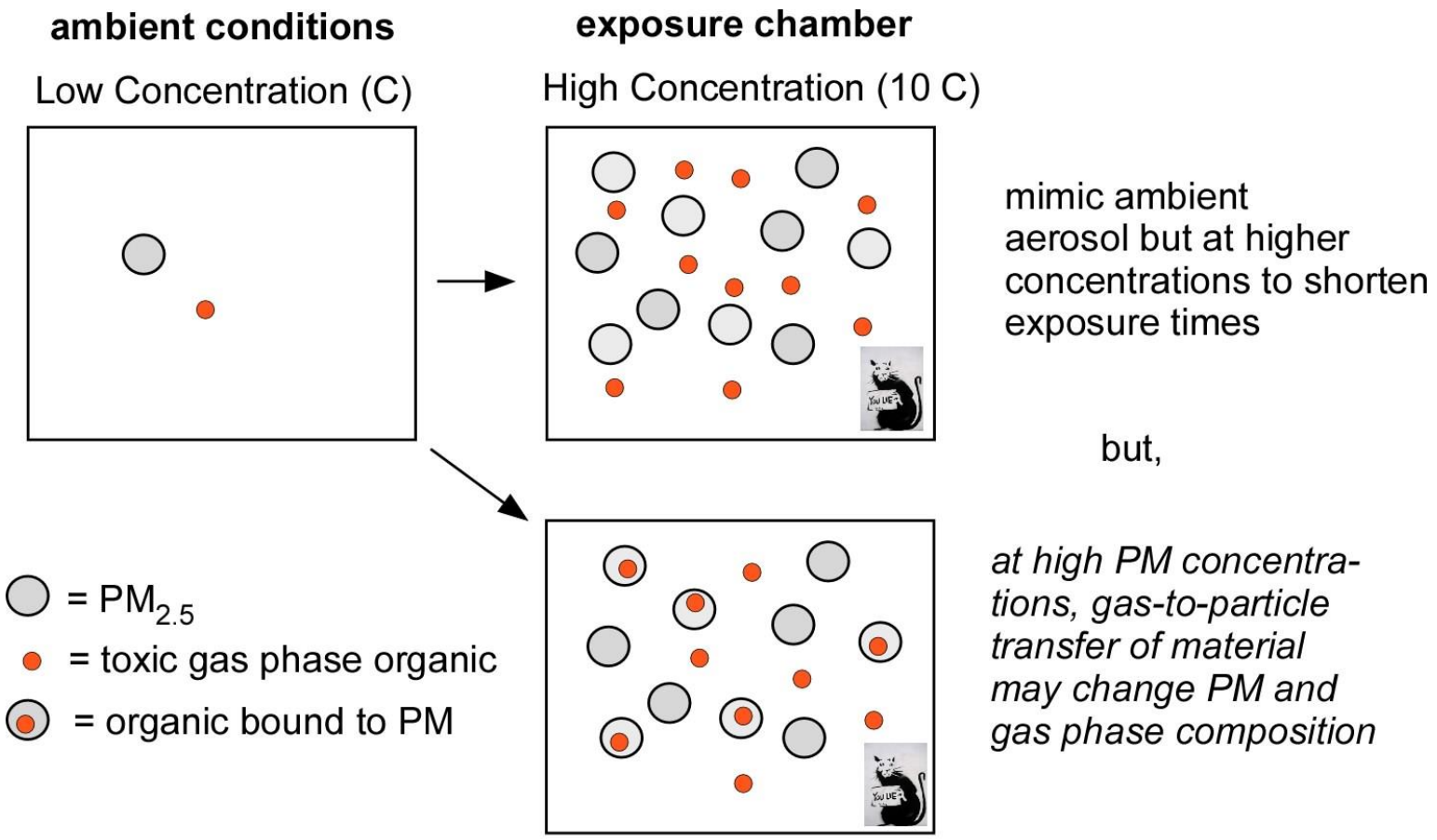
Task 1

Task 2

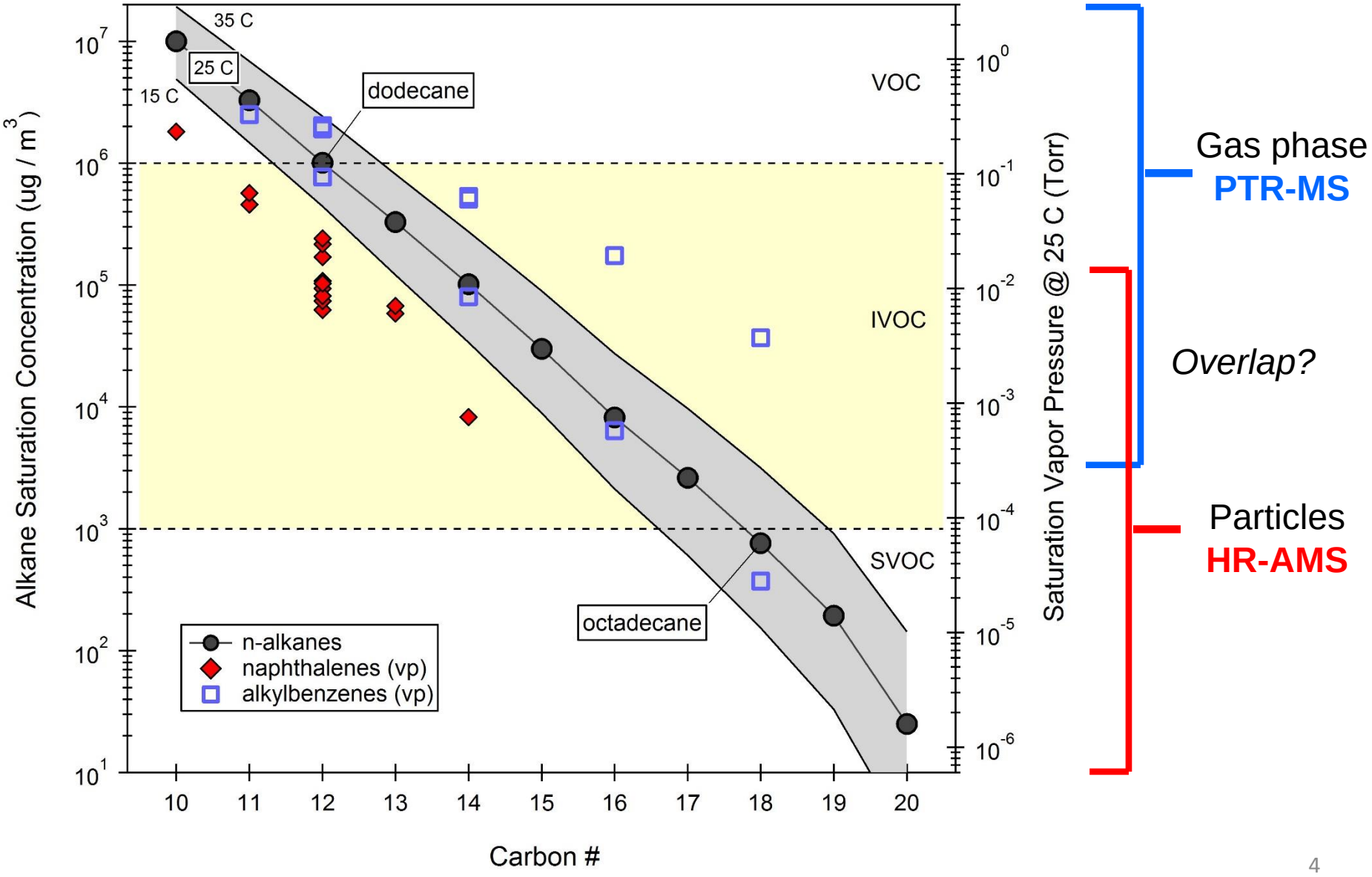
Purpose

Examine engine exhaust aerosol composition measured in chambers to real-world scenarios to provide guidance on generating the most realistic exposures for toxicologic and human clinical trials.

Do high concentration exposures reflect real world aerosol composition?

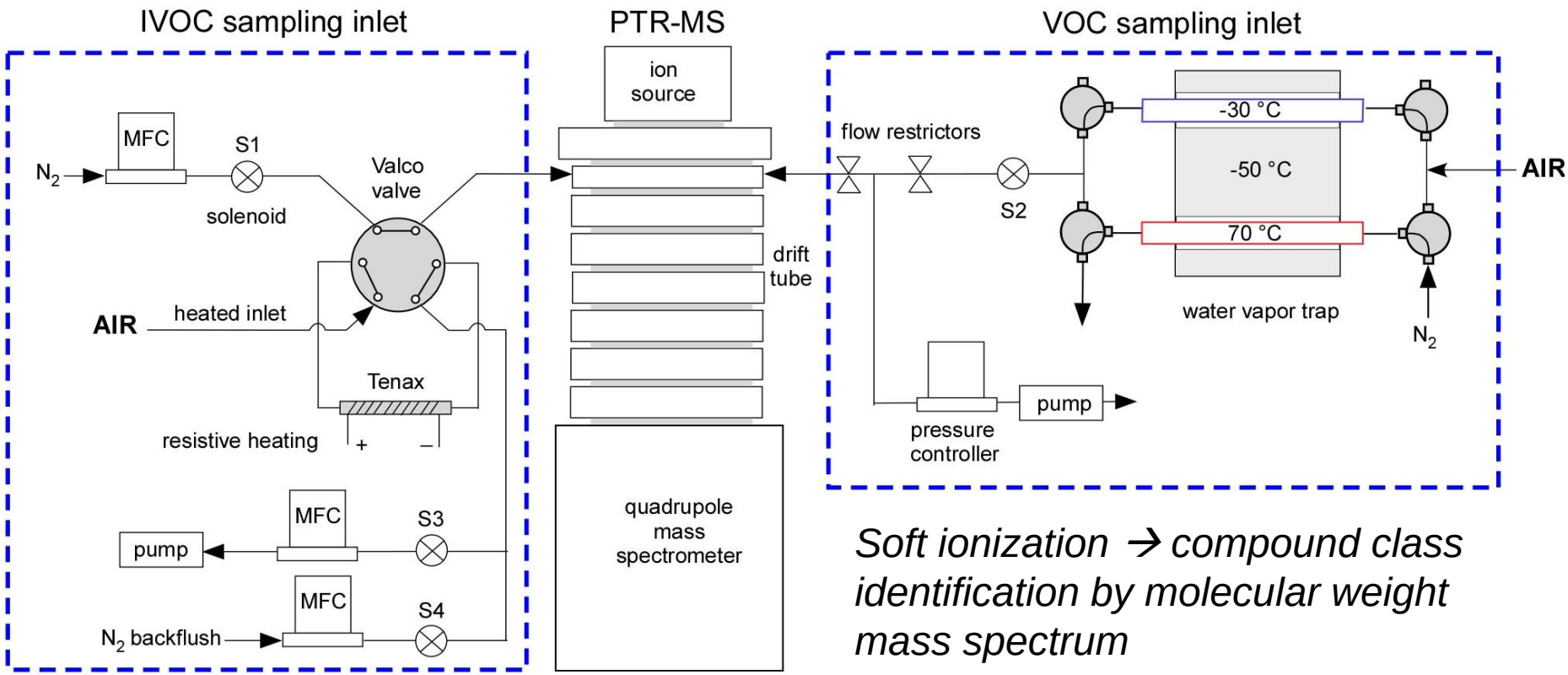


What compounds are impacted by gas-to-particle partitioning in exposure chambers?



Gas phase organic compounds by PTR-MS

tested new sampling scheme to measure larger compounds by thermal desorption

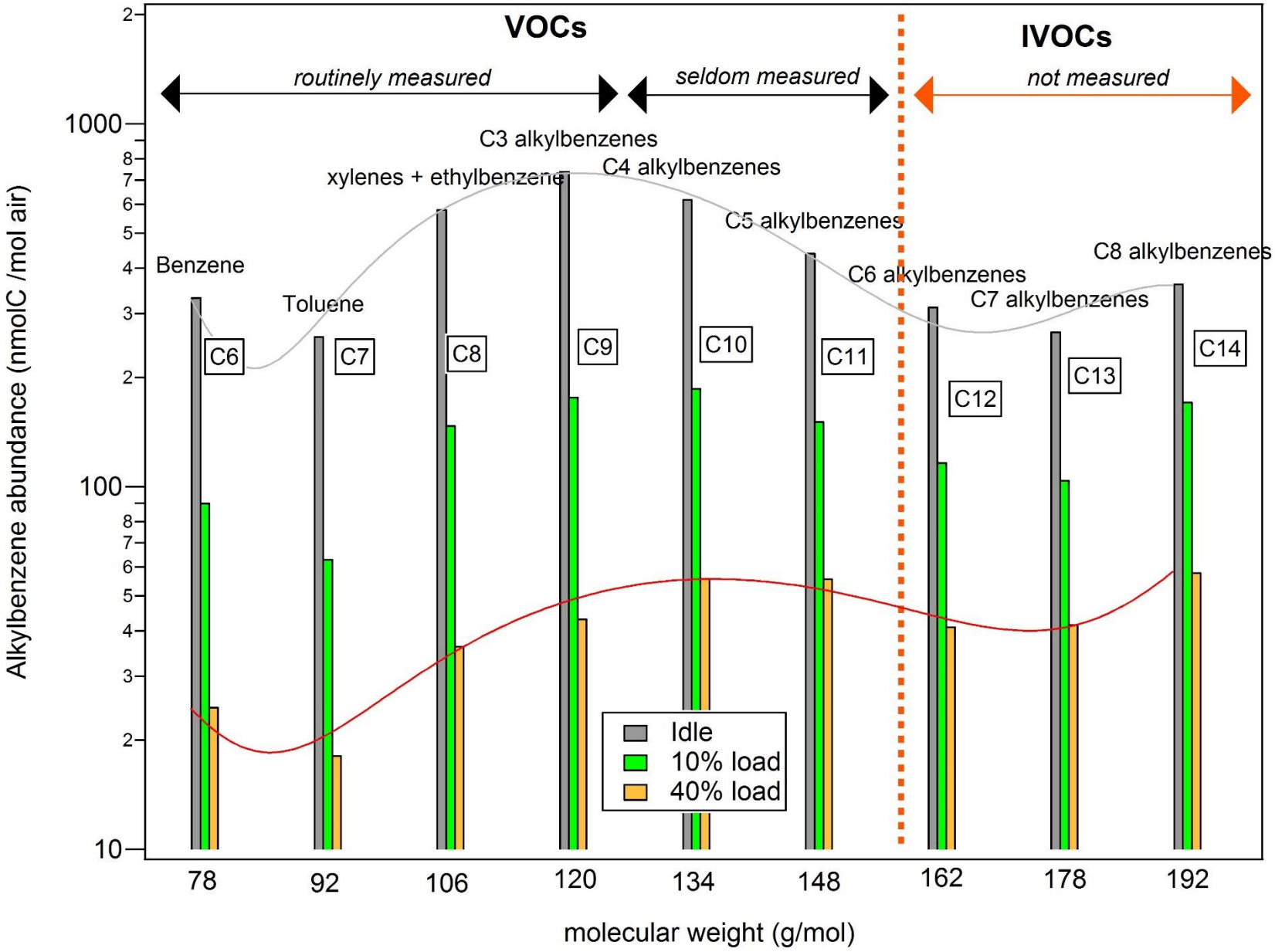


Soft ionization → compound class identification by molecular weight mass spectrum

Measuring Long Chain Alkanes in Diesel Engine Exhaust by Thermal Desorption PTR-MS, M.H. Erickson, M. Gueneron, B.T. Jobson, *Atmospheric Measurement Techniques*, 7, 225-239, 2014.

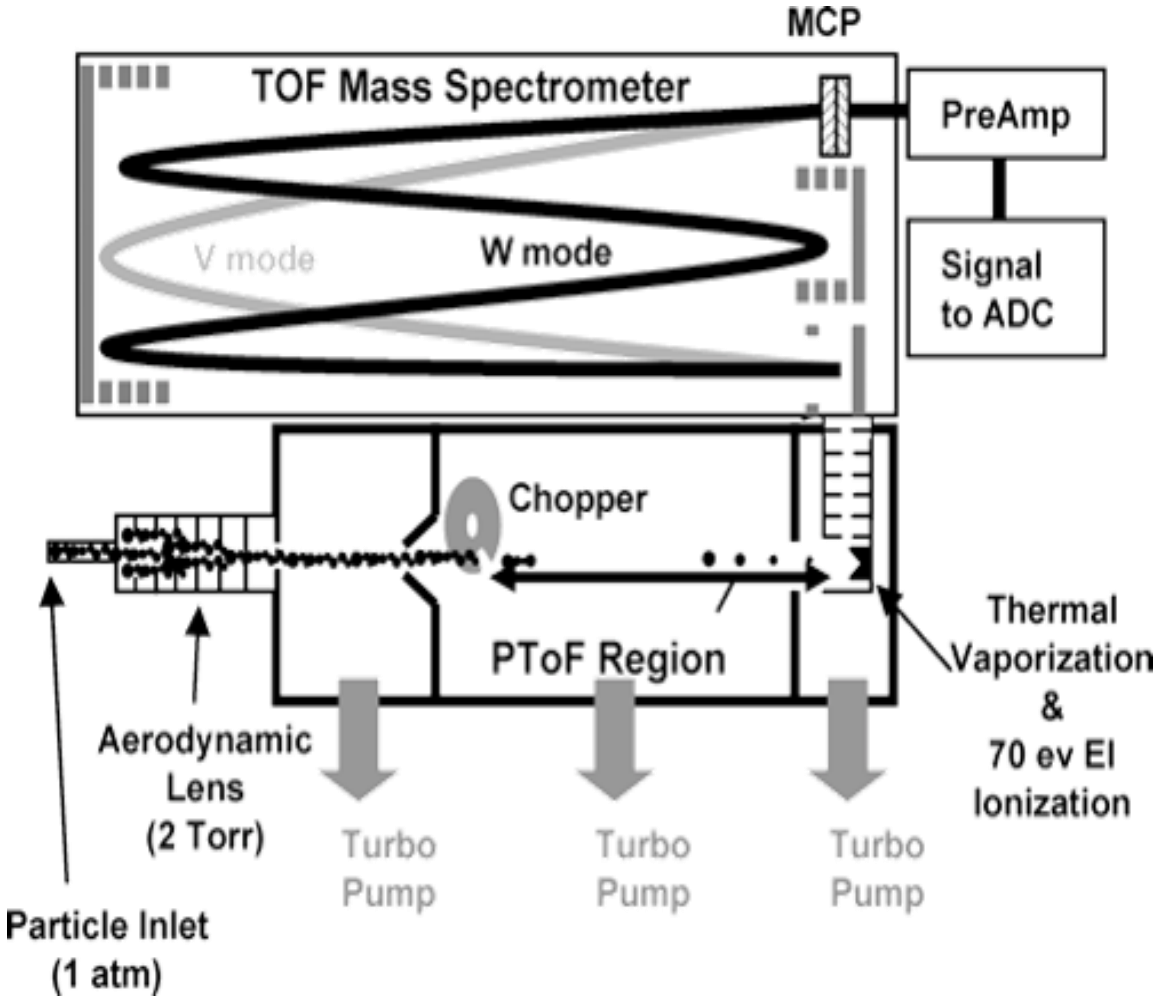
PTR-MS Fragmentation Patterns of Gasoline Hydrocarbons, M. Gueneron, M.H. Erickson, G. VanderSchelden, B.T. Jobson, *Int. J. Mass Spectrometry, in press.*

Alkyl Aromatic Compound Abundance in WSU Diesel Generator Exhaust



Particle phase organics by HR-AMS

measures the composition of non-refractory particulates (50 and 1000 nm).
Only material that volatilizes below ~600 °C is measured



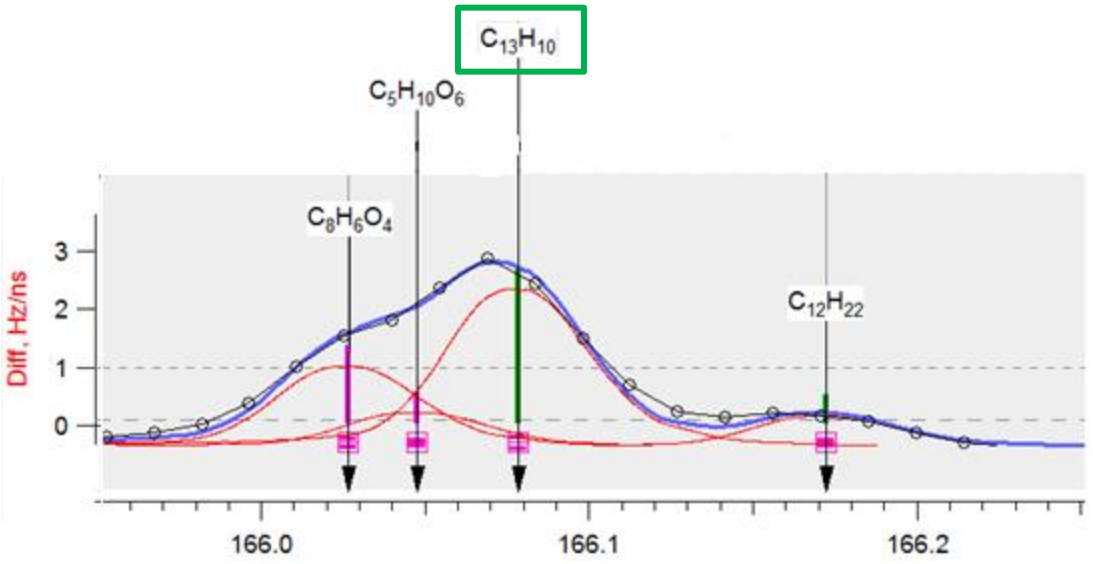
electron impact ionization
mass resolution ~ 5,000

Big job to de-convolve mass spectra to identify constituents.

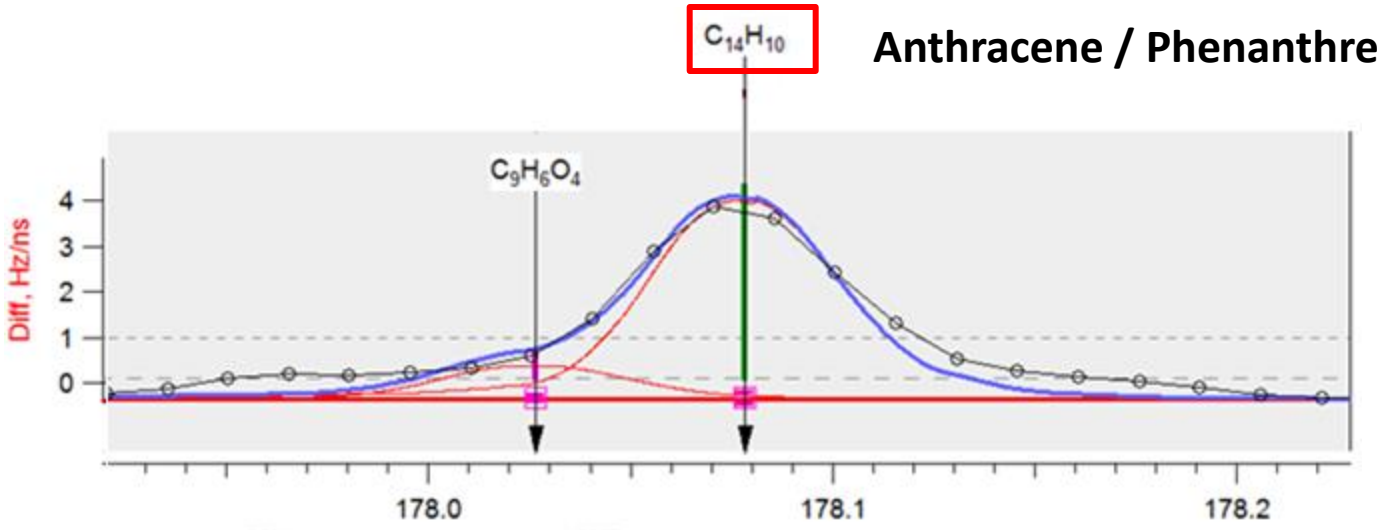
Lower mass PAH compounds

Name	Formula	MW
Naphthalene	$C_{10}H_8$	128
Acenaphthylene	$C_{12}H_8$	152
Acenaphthene	$C_{12}H_{10}$	154
Fluorene	$C_{13}H_{10}$	166
Anthracene	$C_{14}H_{10}$	178
Phenanthrene	$C_{14}H_{10}$	178

Flourene

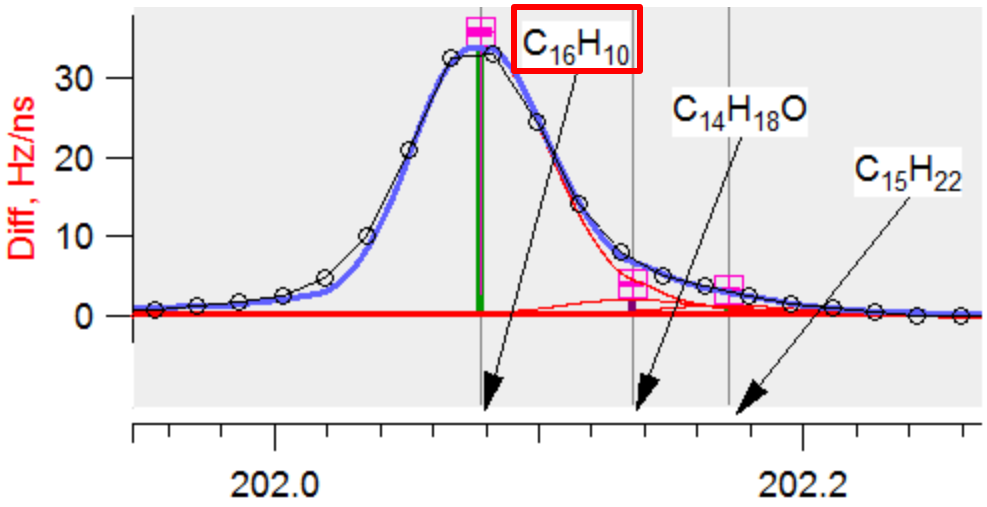


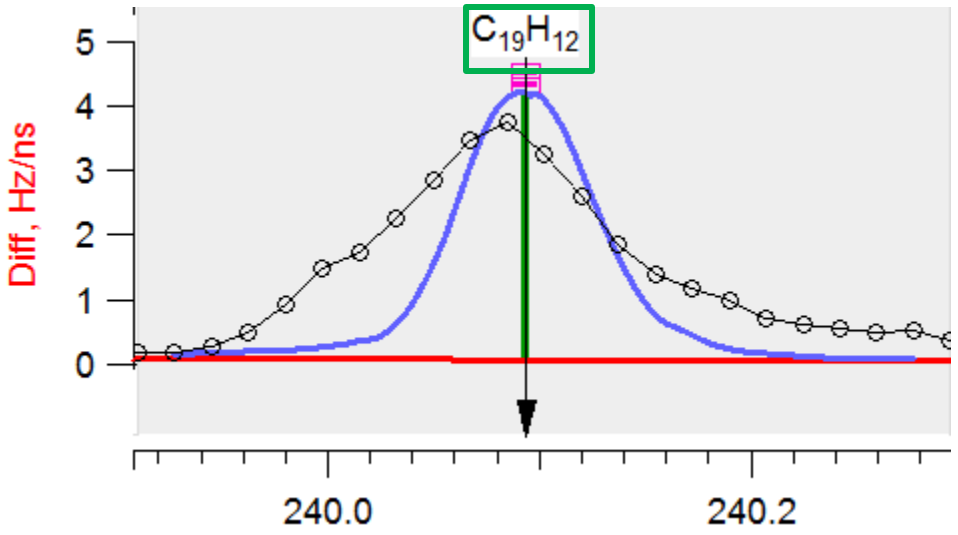
Anthracene / Phenanthrene



Higher mass PAH compounds

Name	Formula	MW
Pyrene Fluoranthane Acephenanthrylene	$C_{16}H_{10}$	202
1,2-Benzofluorene 2,3-Benzofluorene 1-Methylpyrene	$C_{17}H_{12}$	216
Benzo[ghi]fluoranthene Cyclopenta[cd]pyrene	$C_{18}H_{10}$	226
Benz[a]anthracene Chrysene Triphenylene	$C_{18}H_{12}$	228
Methylbenzo[ghi]- fluorathene	$C_{19}H_{12}$	240
…	…	…
Phenanthro[3,4-c]- phenanthrene	$C_{26}H_{16}$	328



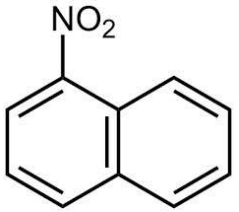


m/z	Chemical Formula	Compounds	Category
128	$C_{10}H_8$	Naphthalene	PAH
142	$C_{11}H_{10}$	Methyl-naphthalene	MPAH
152	$C_{12}H_8$	Acenaphthylene	PAH
154	$C_{12}H_{10}$	Acenaphthene	PAH
156	$C_{12}H_{12}$	Dimethyl-naphthalene	MPAH
166	$C_{13}H_{10}$	Fluorene	PAH
168	$C_{13}H_{12}$	Methyl-acenaphthene	MPAH
173	$C_{10}H_7NO_2$	Nitro-naphthalene	NPAH
178	$C_{14}H_{10}$	Anthracene Phenathrene	PAH PAH
180	$C_{14}H_{12}$	Methyl-fluorene	MPAH
180	$C_{13}H_8O$	Fluornone	OPAH
182	$C_{13}H_{10}O$	Acenapthenequinone	OPAH
192	$C_{15}H_{12}$	Methyl-phenanthrene	MPAH
194	$C_{15}H_{14}$	Dimethyl-fluorene	MPAH
199	$C_{12}H_9NO_2$	Nitro-acenaphthlene	NPAH
202	$C_{16}H_{10}$	Pyrene Fluoranthene	PAH
204	$C_{15}H_8O$	Cyclcopenta-phenanthrene-one	OPAH
208	$C_{14}H_8O_2$	Anthraquinone	OPAH
211	$C_{13}H_9NO_2$	Nitro-fluorene	NPAH
216	$C_{17}H_{12}$	Benzofluorene	PAH
216	$C_{17}H_{12}$	Methylpyrene	MPAH
223	$C_{14}H_9NO_2$	Nitro-anthracene Nitro-phenathrene	NPAH
228	$C_{18}H_{12}$	Benz[a]anthracene Triphenylene Chrysene	PAH
242	$C_{19}H_{14}$	methylbenz[a]anthracene	MPAH
247	$C_{16}H_9NO_2$	Nitro-pyrene	NPAH
252	$C_{20}H_{12}$	Benzo[b,j,k]fluoranthene Benzo pyrene	PAH
273	$C_{18}H_{11}NO_2$	Nitrochrysene	NPAH
276	$C_{22}H_{12}$	Indiopyrene Benzoperylene	PAH
278	$C_{22}H_{14}$	Dibenzanthracene	PAH

Target PAH list for HR-AMS analysis

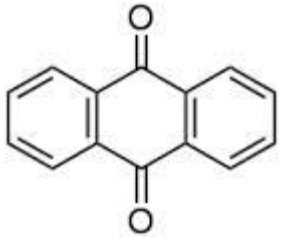
29 PAH compounds (C_{10} to C_{22}) Including:

6 nitro PAHs:



i.e. nitronaphthalene

4 oxygenated PAHs

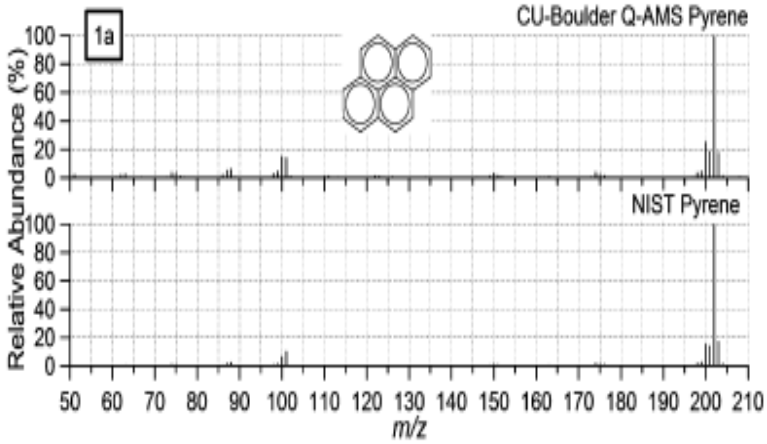


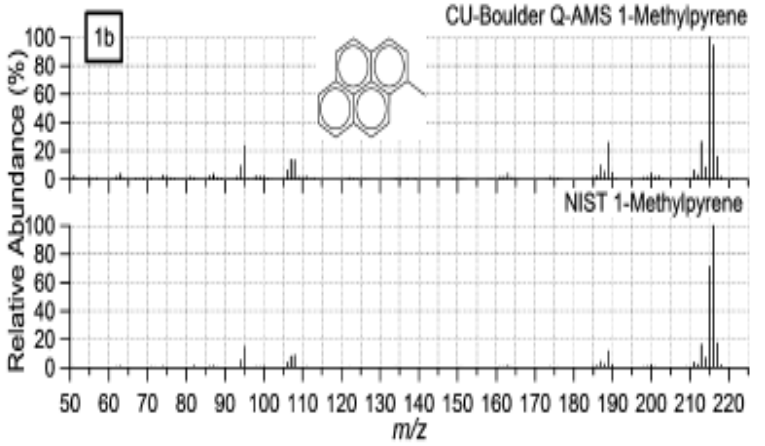
i.e. anthraquinone

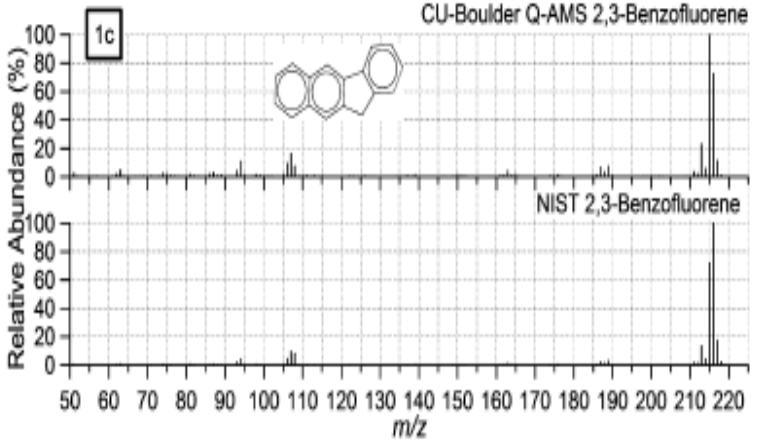
m/z	Chemical Formula	Compounds	% Contribution at Molecular Mass
202	$C_{16}H_{10}$	Pyrene	36.3% [1] 37.2% [2] 27.1% [3]
202	$C_{16}H_{10}$	Fluoranthene	32.1% [2]
216	$C_{17}H_{12}$	Benzofluorene	26.8% [1]
216	$C_{17}H_{12}$	Methylpyrene	19.8% [1]
228	$C_{18}H_{12}$	Triphenylene	25.9% [1]
242	$C_{19}H_{14}$	Methylbenz[a]anthracene	17.0% [1]
252	$C_{20}H_{12}$	Benzo[b,j,k]fluoranthene	26.2% [1] 25.6% [2]
276	$C_{22}H_{12}$	Benzoperylene	24.4% [1]
300	$C_{24}H_{12}$	Coronene	20.6% [1]

Identification by molecular ion but typically this ion is < 30% of total signal.

Currently working with Aerodyne Research Inc. to fine tune PAH analysis to report mass concentration and calculate uncertainties.



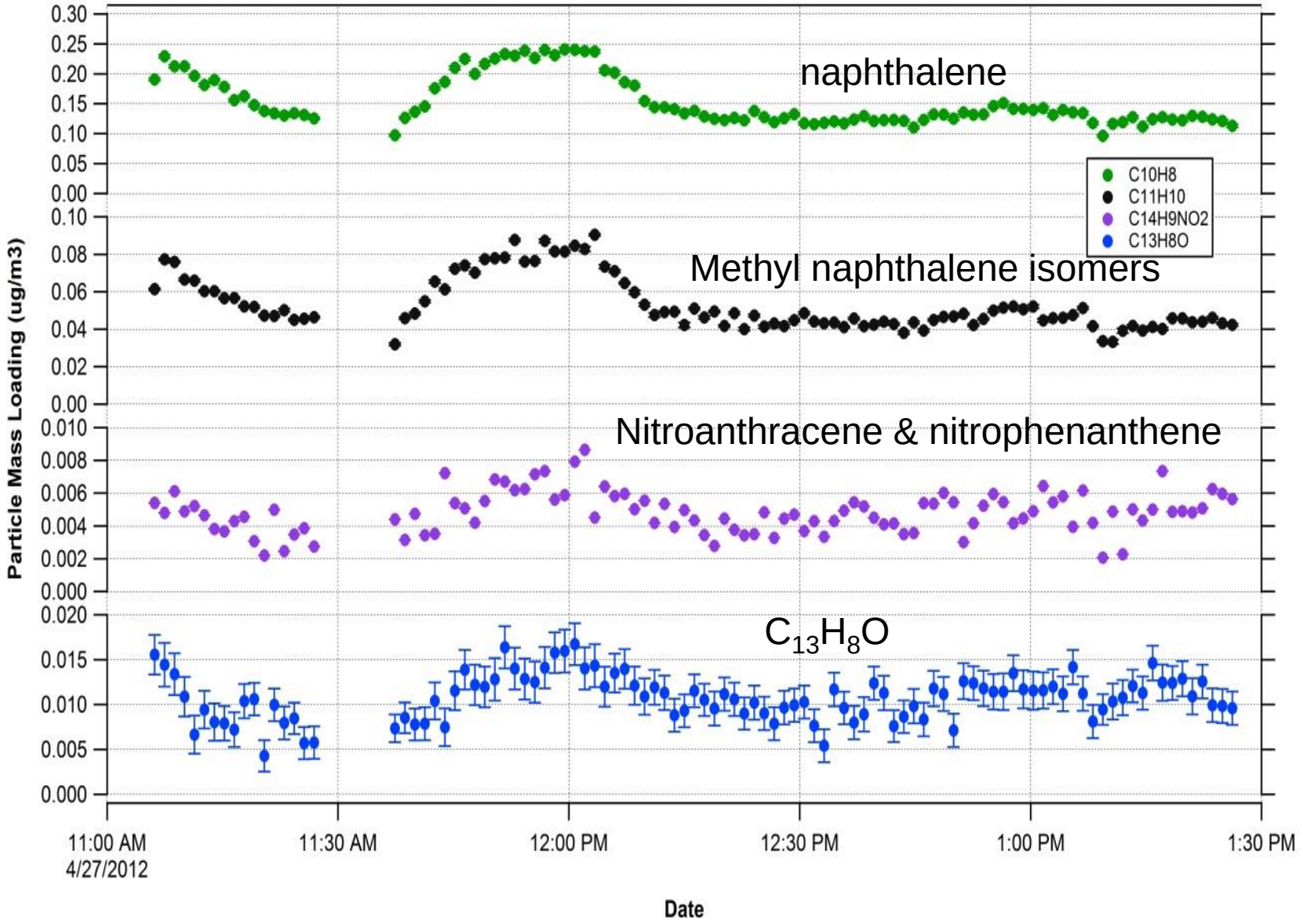




Example PAH data

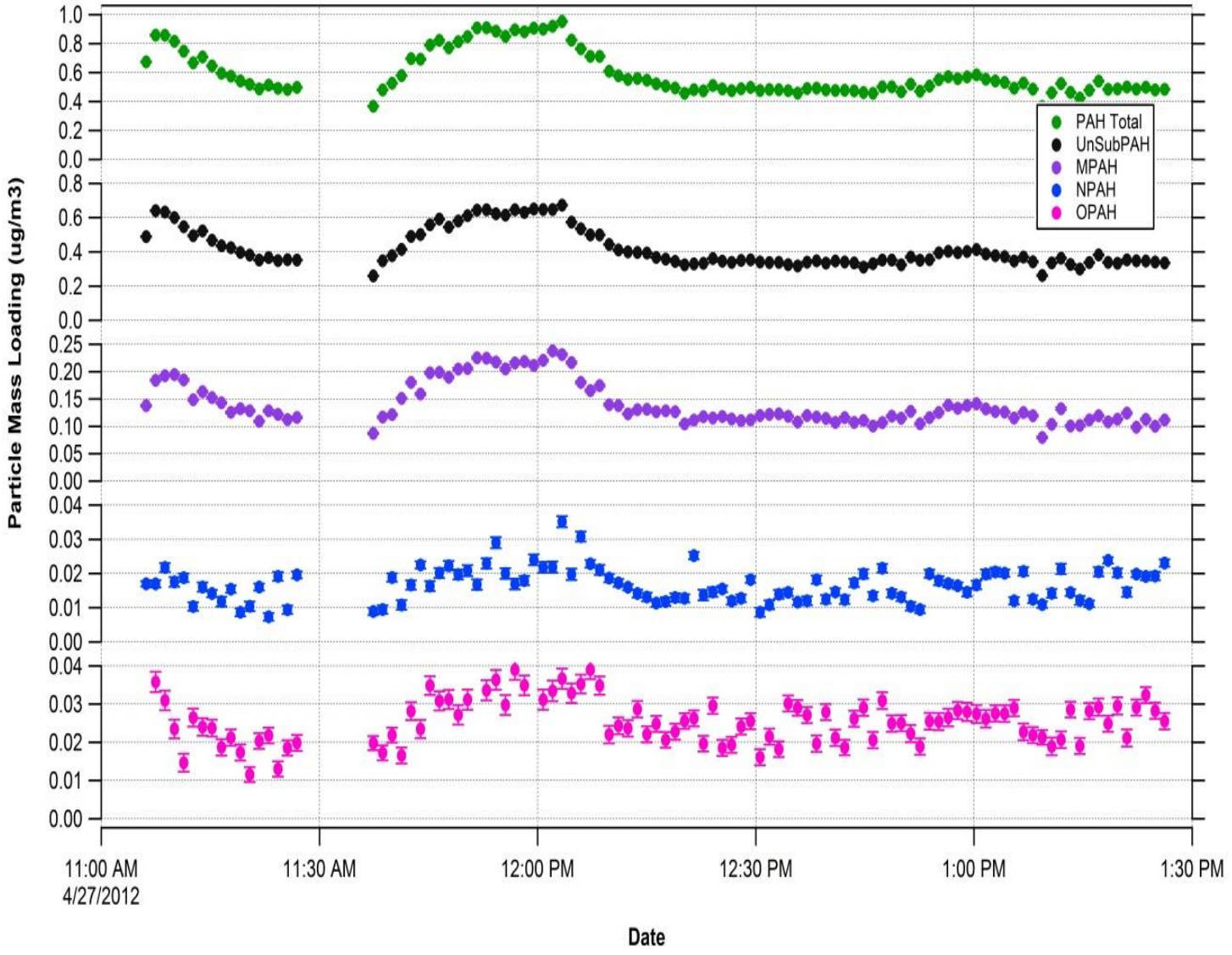
Good signal to noise

LRRI test 4: Diesel exhaust only, 4.5 kW load



Example PAH data

LRRI test 4: Diesel exhaust only, 4.5 kW load



Compounds from list

Σ PAH

Σ unsubstituted PAH

Σ methyl-PAH

Σ nitro-PAH

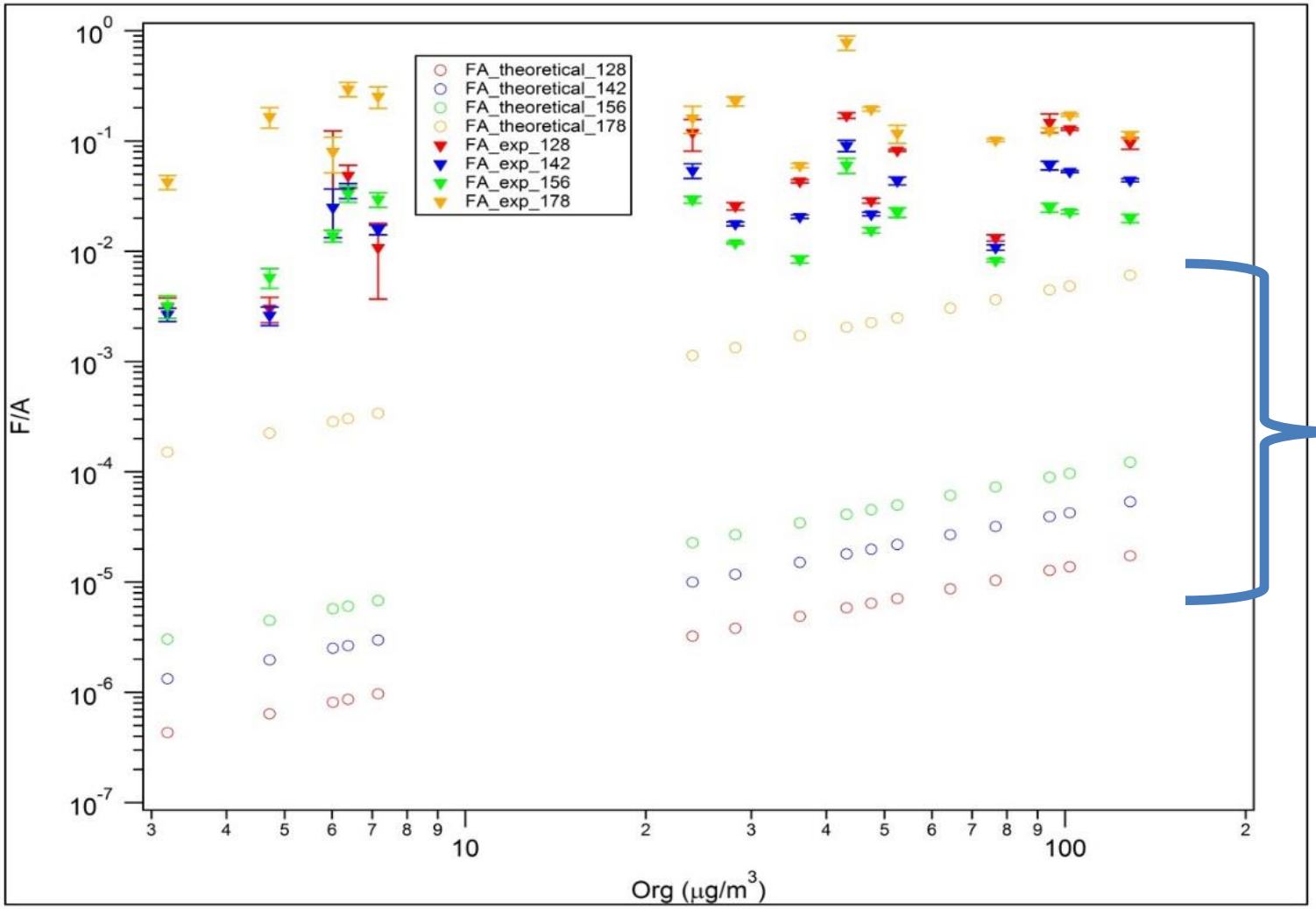
Σ oxy-PAH

Preliminary big finale: Gas-particle equilibrium not achieved.

Enhanced concentration on particles

Naphthalene*, *methylnaphthalene*, *dimethylnaphtahlenes*, *anthracene

Particle concentration / gas phase concentration



CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

Contribution of time in-transit to individual exposure to traffic-related air pollution (Project 5)

Marnie Hazlehurst
06 October 2014

Aims

- Aim 1: To build a multi-pollutant exposure model for traffic-derived air pollutants for use in epidemiological analysis
- Aim 2: To determine the effect of time-in-transit on individual exposure in this cohort
- Aim 3: To estimate the effect of individual-level exposure to traffic-derived air pollution on subclinical cardiovascular disease in MESA Air

Aims

- Aim 1: To build a multi-pollutant exposure model for traffic-derived air pollutants for use in epidemiological analysis
- Aim 2: To determine the effect of time-in-transit on individual exposure in this cohort
- Aim 3: To estimate the effect of individual-level exposure to traffic-derived air pollution on subclinical cardiovascular disease in MESA Air

Multi-Ethnic Study of Atherosclerosis and Air Pollution

- Cohort study of ~7,000 adults in 6 US cities
  - 45-84 years old at recruitment
  - Multi-ethnic sample
    - White
    - Black/African-American
    - Hispanic
    - Chinese

- Spatio-temporal air pollution exposure predictions for ambient air pollutants
  - Project included effort at incorporation of individual-level predictions weighted by time spent indoors versus outdoors
    - MESA Air Questionnaire completed by all participants provides time indoors at home, outdoors, in-vehicle, and at other locations.

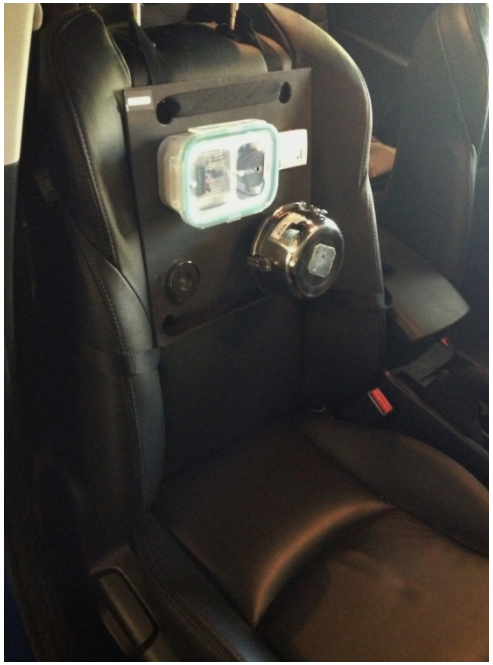
CCAR Project 5 Work: Four Monitoring Campaigns Completed

- Subset of the MESA Air Cohort
  - Selection criteria:
    - English-speaking
    - Non-smoker and not living with a smoker
    - Reported driving at least 30 min/day
    - Not traveling during sampling period

- Four 2-week campaigns
  - Winston-Salem, NC (2013)
    - Winter, n=46
    - Summer, n=47
  - Los Angeles, CA (2014)
    - Winter, n=47
    - Summer, n=46

Passive Time-Integrated Monitoring

- Residential indoor and outdoor, in-vehicle, and personal monitoring
  - Two-week time-integrated passive samplers
  - Measured NO_2, NO_x, SO_2, O_3, pentanes, isoprene, n-nonane, n-decane, n-dodecane, n-undecane, benzene, toluene, m-xylene, o-xylene
  - Novel in-vehicle sampler only sampling while participant is driving
- Location tracking
  - Time-location diary
  - GPS tracking + proximity monitor

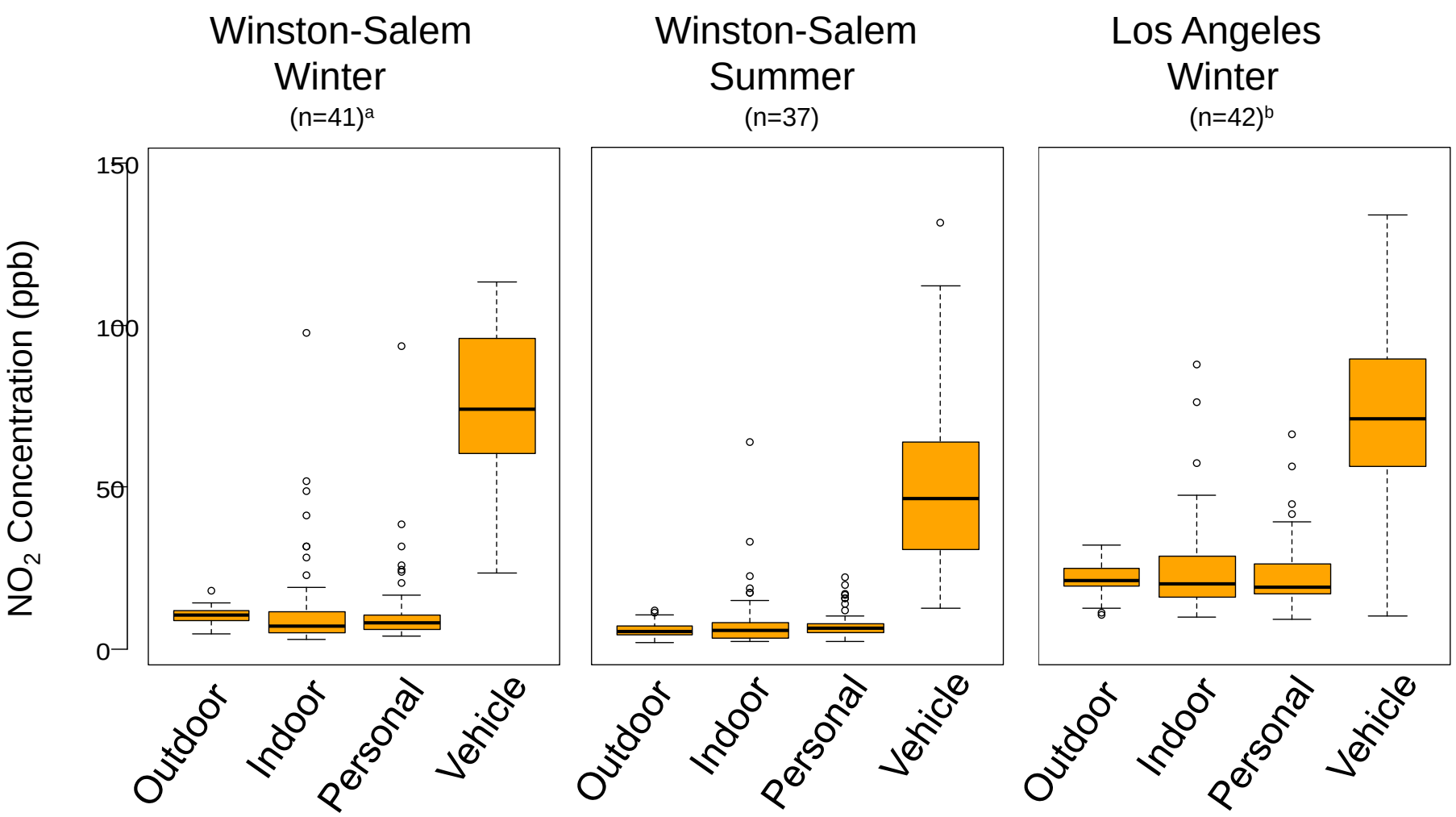


Participant Demographics

	Winston-Salem		Los Angeles	
	Winter (n=46)	Summer (n=47)	Winter (n=47)	Summer (n=46)
	# (%)	# (%)	# (%)	# (%)
Gender				
Male	21 (46)	23 (49)	23 (49)	27 (59)
Female	25 (54)	24 (51)	24 (51)	19 (41)
Race/Ethnicity				
White	20 (43)	21 (45)	14 (30)	15 (33)
Black	26 (57)	26 (55)	9 (19)	7 (15)
Hispanic	0 (0)	0 (0)	21 (45)	19 (41)
Chinese	0 (0)	0 (0)	3 (6)	5 (11)
Age				
Median Age*	72	69	65	67
Age range*	54 - 89	54 - 89	54 - 83	54-90

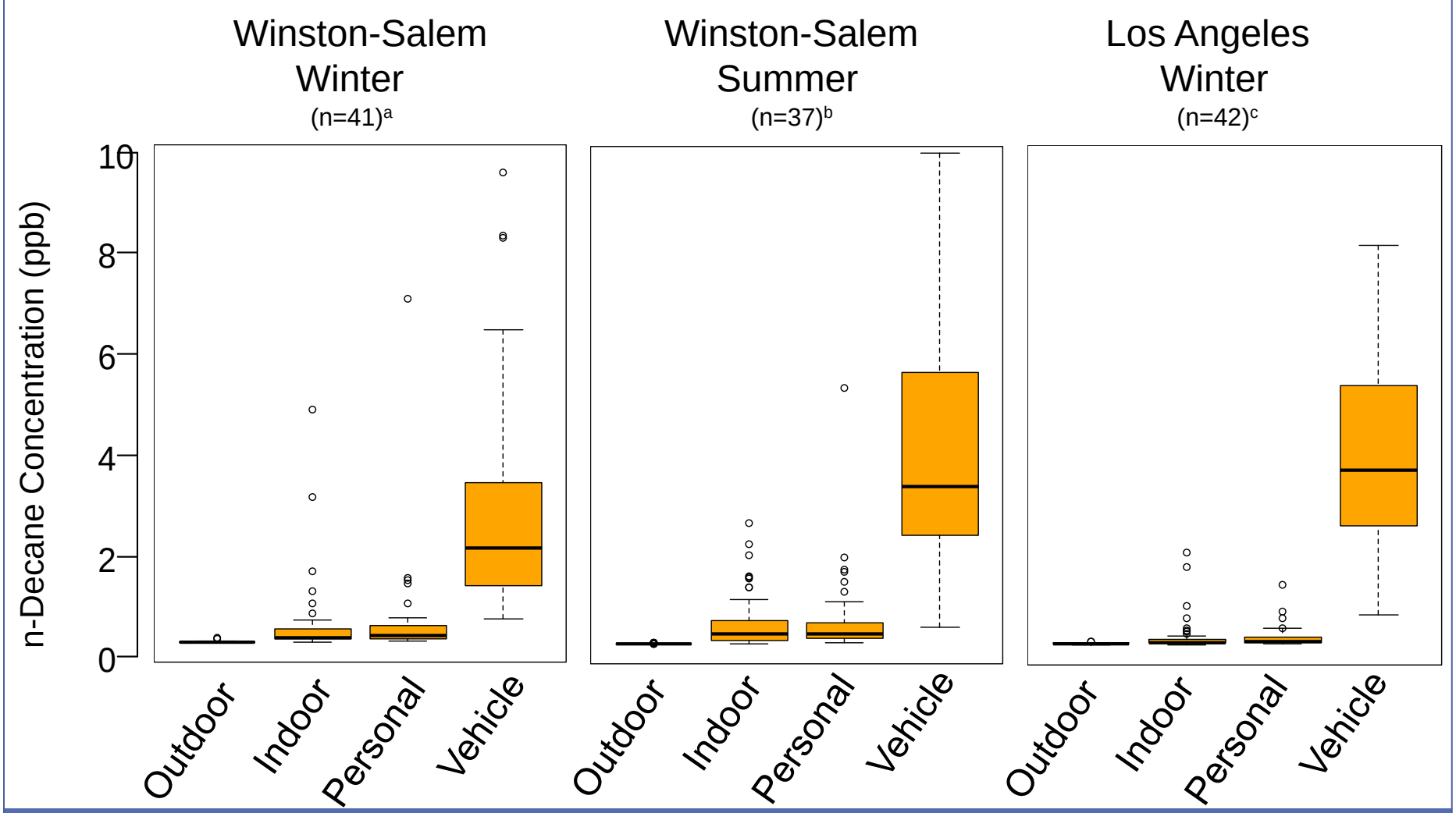
*At MESA Exam 5

Measured NO_2 Concentrations



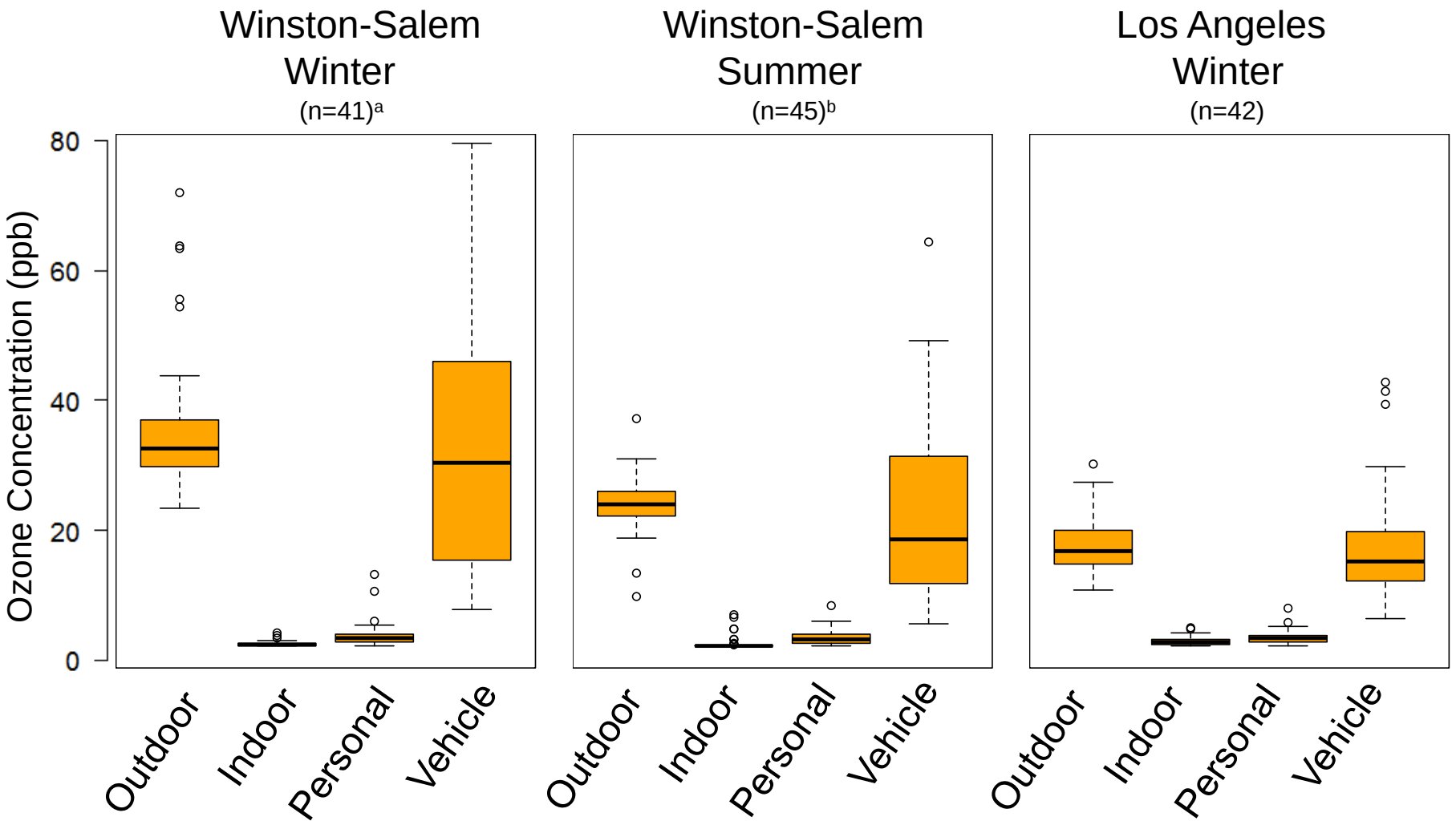
[a] 5 high vehicle outliers excluded for display; [b] 3 high vehicle outliers excluded for display

Measured n-Decane Concentrations



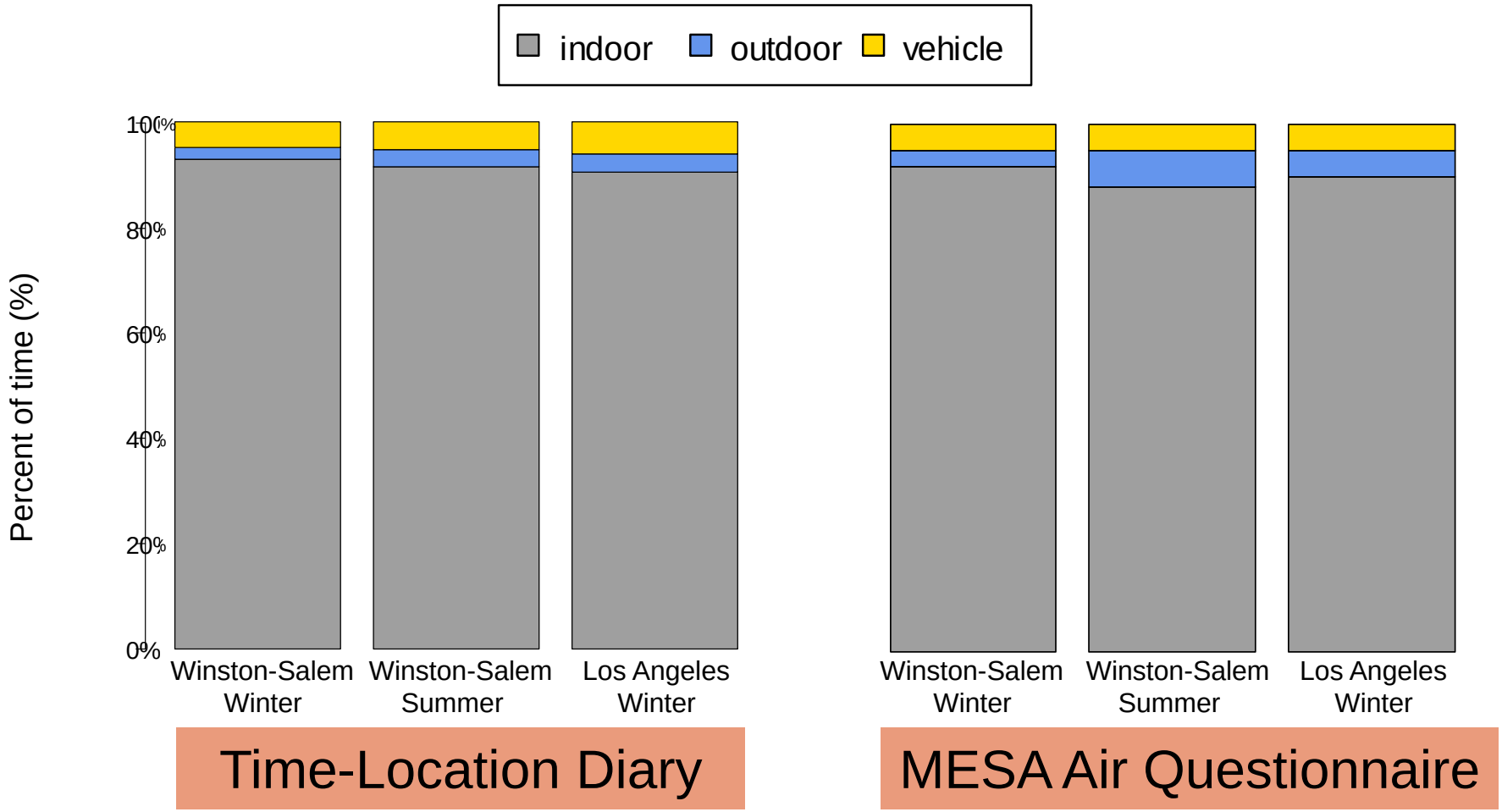
[a] 1 high vehicle outlier excluded for display; [b] 6 high vehicle outliers excluded for display; [c] 2 high vehicle outliers excluded for display

Measured Ozone Concentrations



[a] 6 high vehicle outliers excluded for display, [b] 3 high vehicle outliers excluded for display

Participants spend most of their time indoors



Estimating the contribution of micro-environmental exposures to overall individual exposure to traffic-derived NO_2

- Measured NO_2 is highest in-vehicle, but participants spend only a small percentage of time in the car
- 'Ambient-derived' = 'traffic-derived'
  - Specifically focused on traffic-related air pollution (TRAP).
  - Not interested in indoor sources as we are using NO_2 as a marker of TRAP.
- Focus thus far on nitrogen dioxide (NO_2)
  - Criteria air pollutant
  - Marker of traffic-related air pollution

Time-Weighted Model of Traffic-Derived NO_2 Exposure

$$E_T = E_{Indoor} + E_{Outdoor} + E_{Vehicle}$$

E_T = Total traffic-derived individual exposure to NO_2

E_{Indoor} = infiltration factor * t_{Indoor} * $[NO_2]_{Outdoor}$
$E_{Outdoor}$ = $t_{Outdoor}$ * $[NO_2]_{Outdoor}$
$E_{Vehicle}$ = $t_{Vehicle}$ * $[NO_2]_{Vehicle}$

- Where,
  - t = hours spent indoors, outdoors or in-vehicle as reported via time-location diary
  - $[NO_2]$ = measured NO_2 concentration outdoors or in-vehicle
  - infiltration factor = modeled indoor/outdoor ratio

Infiltration of NO_2

- Calculated infiltration factor (indoor/outdoor ratio) based on the literature[a]

$$\frac{I}{O} = \frac{ACH \times f}{K + ACH}$$

- Three constants (averages):
  - Air exchange rate[b] (ACH)
    - Winston-Salem = 0.50 h^{-1}
    - Los Angeles = 1.01 h^{-1}
  - Penetration factor[c] (*f*) = 1.00
  - Decay rate of NO_2[a,d] (K) = 0.99 h^{-1}

[a]Dimitroupoulou, et al., 2000; [b]Sherman, 2007; [c]Thatcher & Layton, 1995; [d]Yamanaka, 1984

Strengths

- Novel sampling design that captures exposure while driving
- Multi-ethnic cohort in two cities
- Subset of cohort reflects demographic distributions of the entire MESA Air cohort

Limitations

- Literature-derived estimate of infiltration
- Self-reported time-location data
- Older population

Implication for Epidemiologic Analyses

- Even though participants only spend a small amount of time in-vehicle, this time accounts for a large proportion of overall exposure to traffic-related air pollutants.

- Estimates not incorporating information about exposure during time-in-transit may be underestimating individual exposure to traffic-related air pollutants.

Status and Future Work

- Status
  - Successful completion of all four monitoring campaigns
  - In-vehicle exposure is important contributor to individual exposure to traffic-derived NO_2
  - Analysis of GPS and proximity monitor data in progress
- Future work
  - Preparing manuscripts, including: methods, GPS analysis, and NO_2 exposure
  - Consider incorporation of time spent in-transit into individual-level exposure predictions for MESA Air cohort
  - Epidemiologic analyses utilizing exposure predictions

Acknowledgements

- MESA Air participants
- UW CCAR Project 5 investigators, field team, & staff, including:
  - Joel Kaufman, Sverre Vedal, Elizabeth Spalt, Cynnie Curl, Mark Davey, Tyler Nicholas

CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

- Funding
  - UW CCAR – Environmental Protection Agency (RD83479601-0)
  - MESA Air – Environmental Protection Agency (RD831697)

Regression approach to addressing temporal confounding in mobile monitoring

Elena Austin

October 6th 2014

University of Washington CCAR SAC meeting

SCHOOL OF PUBLIC HEALTH · UNIVERSITY of WASHINGTON
excellent science, shared passion, enduring impact

CCAR
Center for Clean Air Research

This research is funded by
U.S. EPA - Science To Achieve Results (STAR) Program
Grant # 83479601

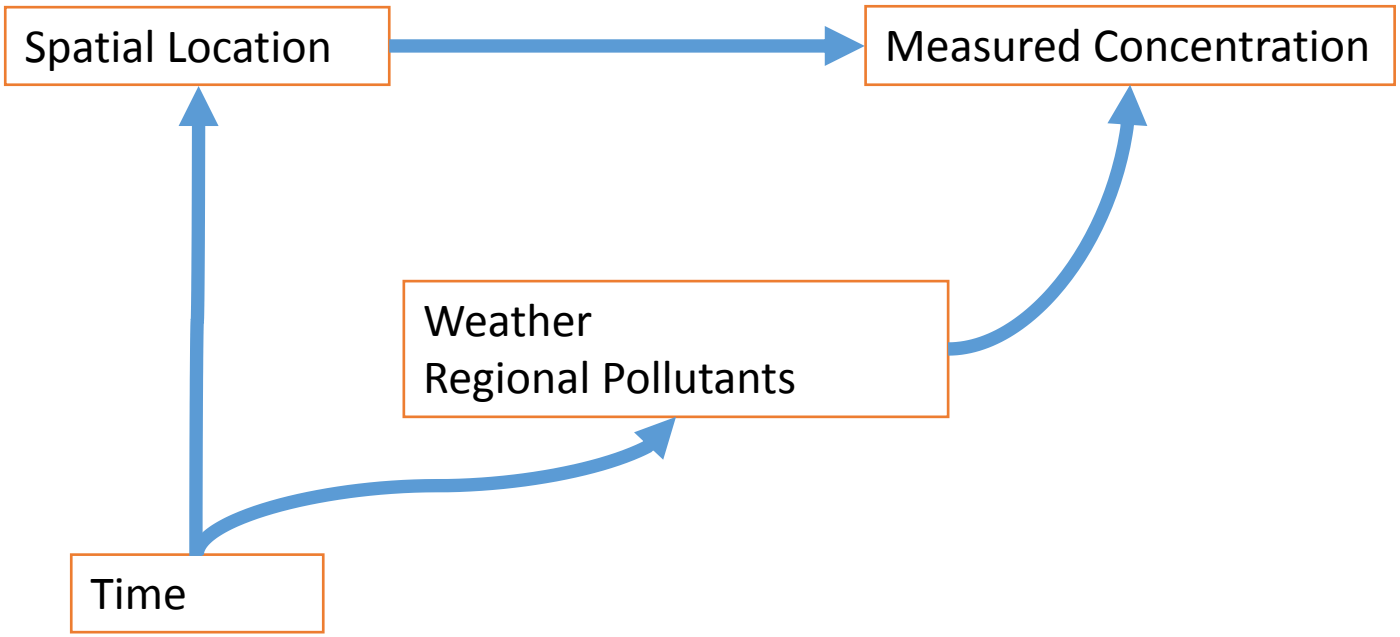
Goals

- Predict single and multipollutant spatial exposures based on mobile monitoring data
  - Describe the importance of confounding by time-varying variables such as weather and regional pollutant concentrations
  - Develop a method to adjust measured values
  - Obtain spatial distributions of single and multipollutant measurements across space

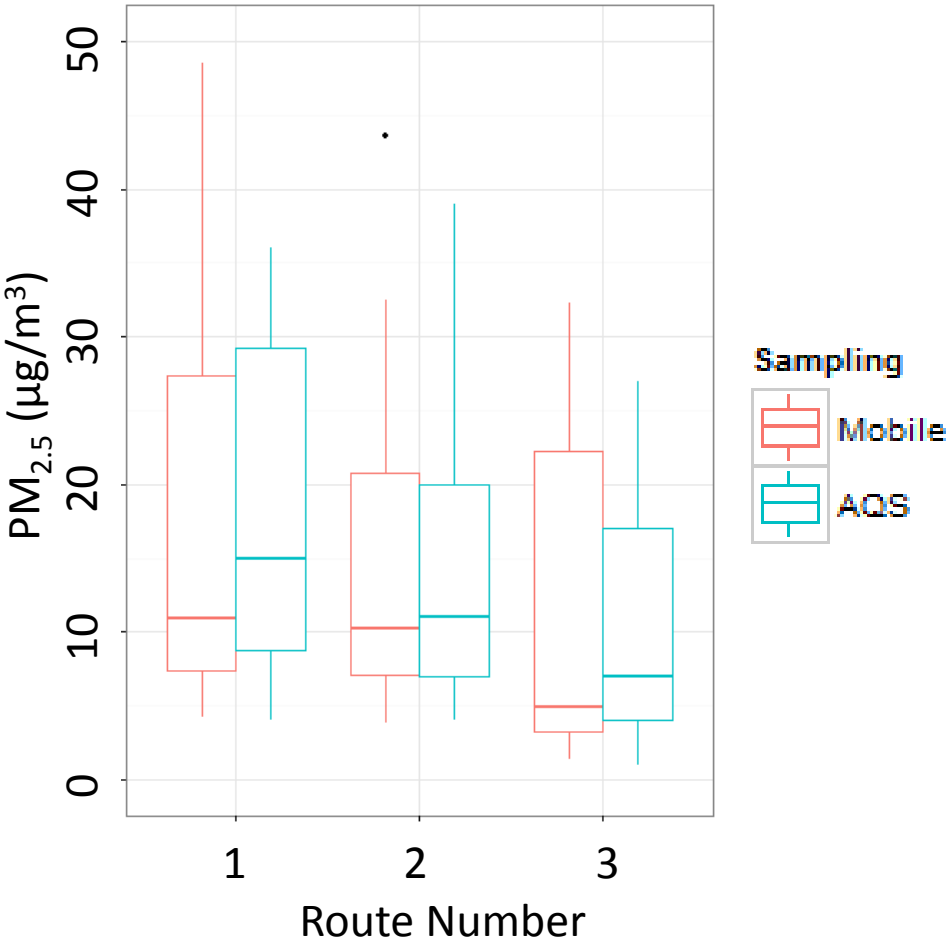
Sampling design

- Each day, one of 3 routes is driven during the afternoon
- This sampling design allows for detailed characteristics to be obtained for 43 locations
  - Medians at different sampling locations
  - Describe distribution around that median
  - Describe differences in exposure based on measured data
- However, there are important
  - Between day differences in pollutant distribution
  - Within day differences in pollutant distribution

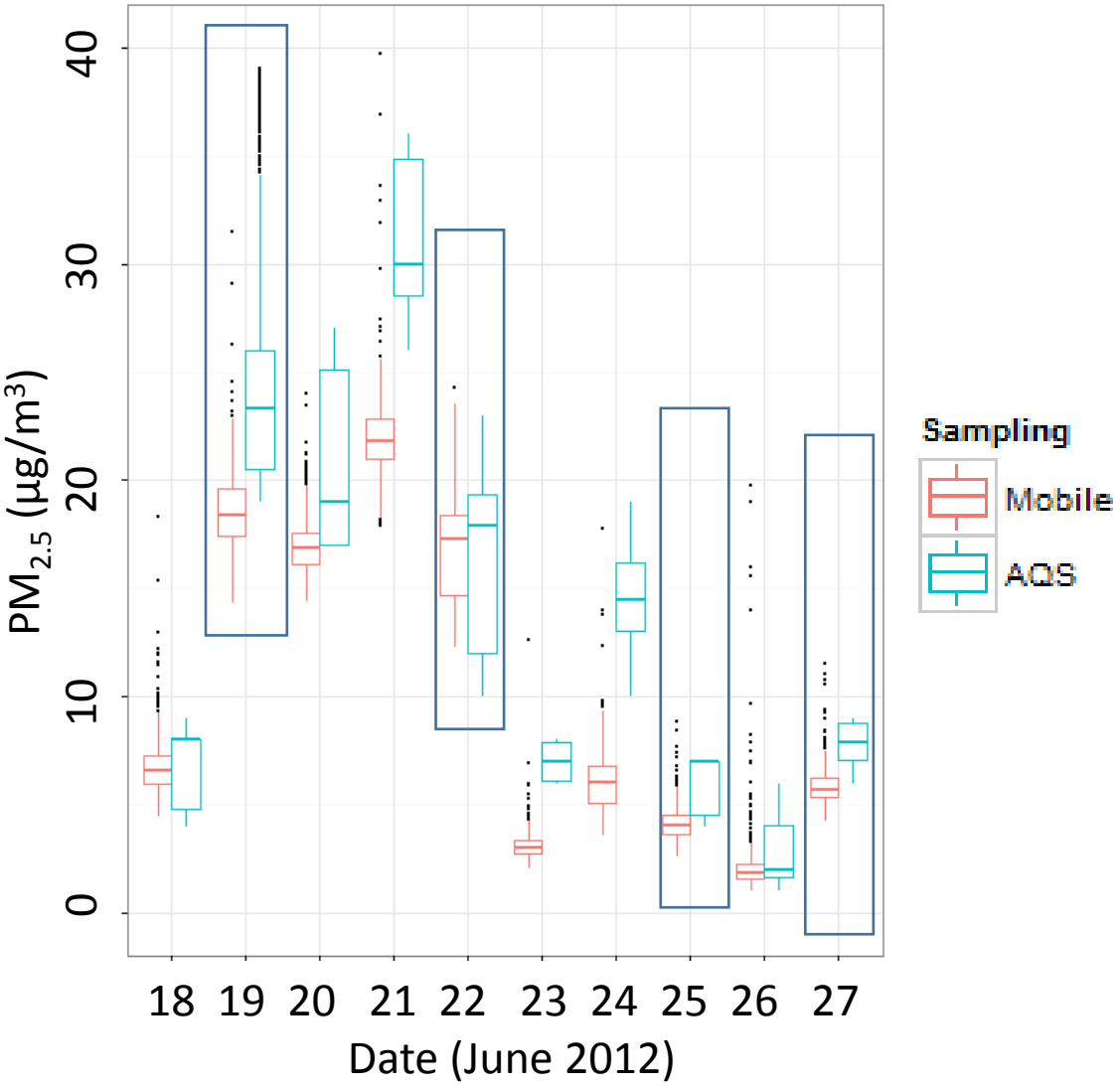
Confounding by time



$PM_{2.5}$ Concentrations by Route, Summer

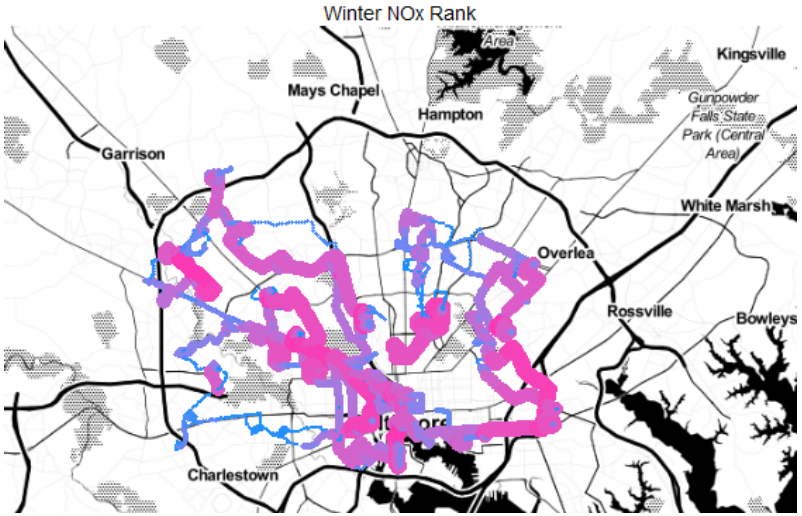


Distributions of $PM_{2.5}$ Concentrations by Day

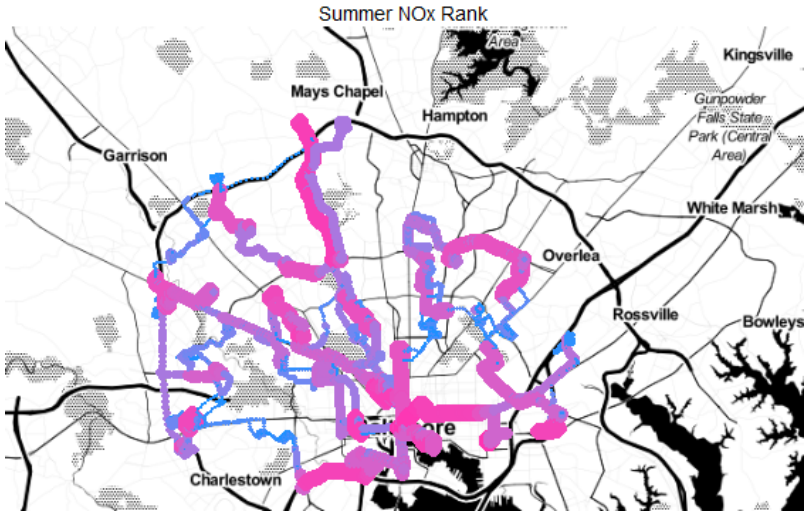


Spatial Pattern in Mobile Monitoring Data

Daily Ranks of NO_x Winter



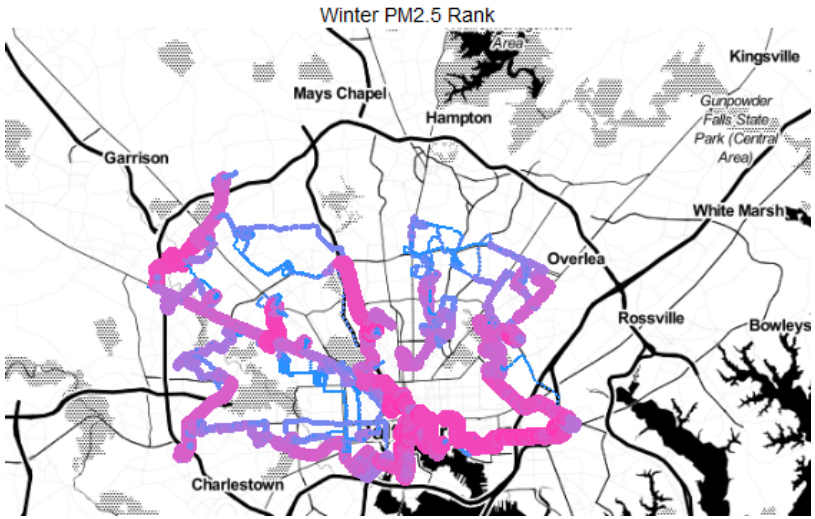
Daily Ranks of NO_x Summer



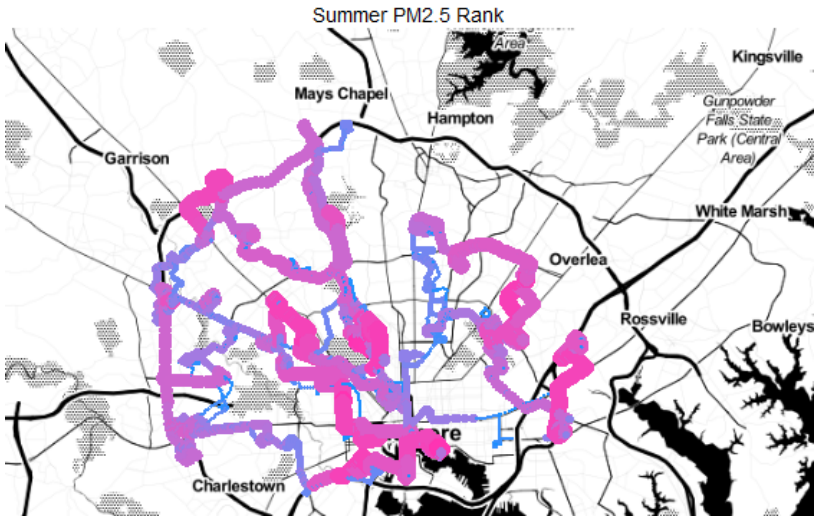
- High Daily Rank in Concentrations
- Low Daily Rank in Concentration

Spatial Pattern in Mobile Monitoring Data

Daily Ranks of $PM_{2.5}$ Winter

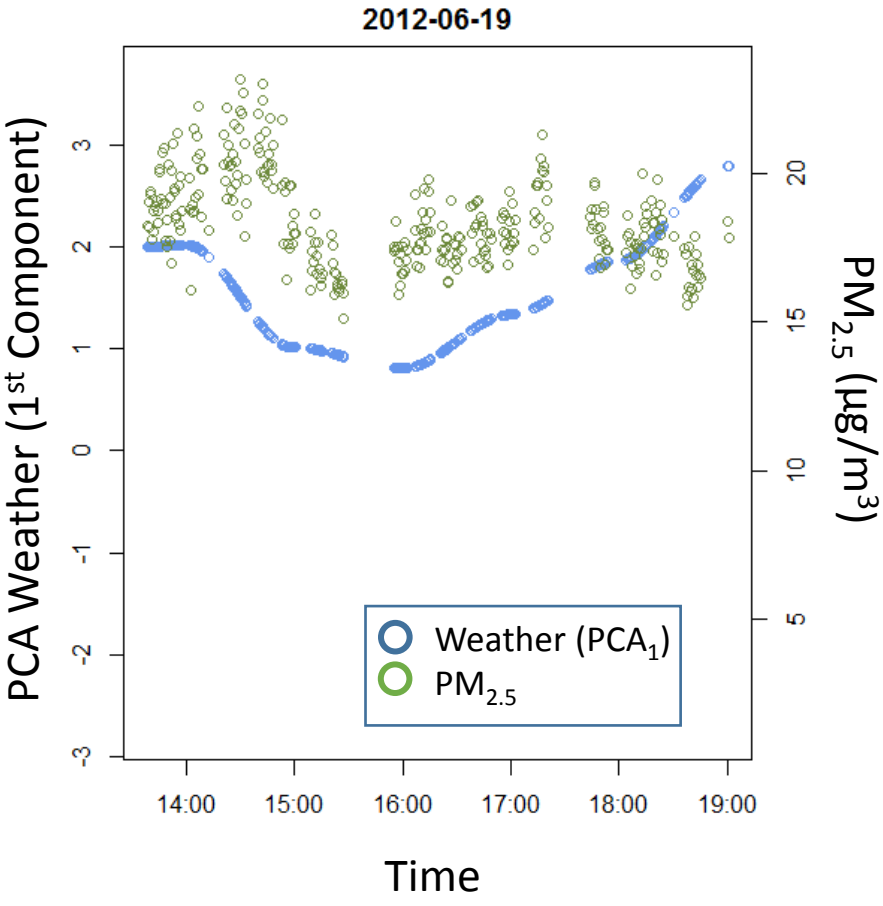


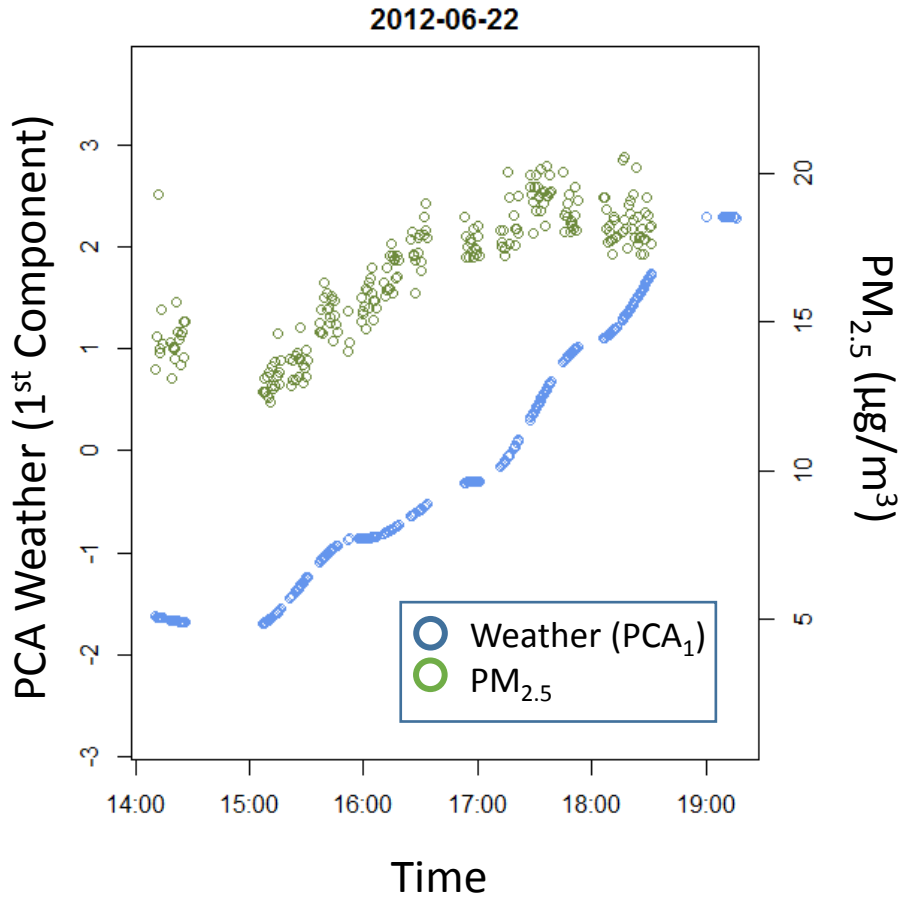
Daily Ranks of $PM_{2.5}$ Summer



- High Daily Rank in Concentrations
- Low Daily Rank in Concentration

Within day differences in concentrations





Correction Approach

- Adjust measurements based on time-varying covariates including
  - Weather (Dry Bulb Temp, RH, Sea Level Pressure, Wind Speed)
  - AQS measurements ($PM_{2.5,}$ Ozone, NO_2)
  - Hour of day (surrogate for city-wide congestion)
  - Solar Radiance (R)
  - Day of sampling
  - Aftersunset
- Produces time adjusted MEDIAN concentrations at fuzzy point locations
- RESIDUALS indicate remaining variability not captured by time varying components

Modeling Approach (separate models for Winter and Summer)

1. Model including all fuzzy point locations
   - Time varying covariates
   - Categorical variable for fuzzy point location

   $$\text{Log[Poll]} = \beta_0 + \beta_{1\text{-}42}\text{FP} + \beta_{43}(\text{R}) + \beta_{44}(\text{aftersunset}) + s(\text{PCA.AQS}_1, \text{PCA.AQS}_2) + s(\text{PCA.W}_1, \text{PCA.W}_2)$$

2. Model including all fuzzy point locations
   - Time varying covariates
   - Spatial smoothing term s(Longitude, Latitude)

3. Analysis stratified by fuzzy point location
   - Time varying covariates
   - Full-interaction model

Verifying Model Output

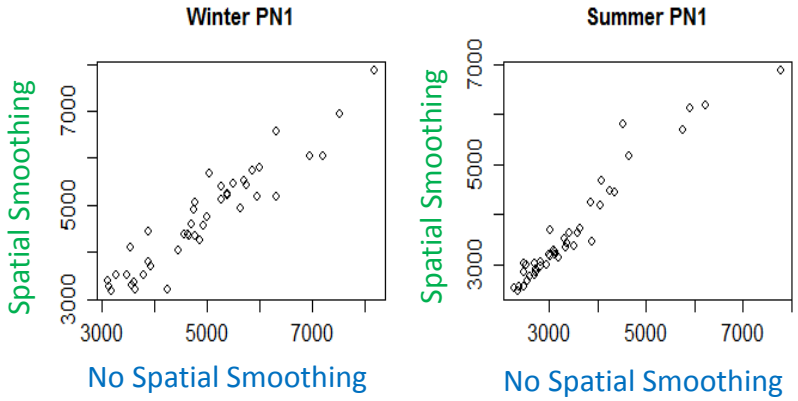
- Leave one out cross validation (requires spatial smoothing term)
- Within sample cross-validation (NO spatial smoothing)
- Comparison of spatial smoothing vs no spatial smoothing models
- Within Fuzzy-Point models
  - Allows for adjustment by all spatially dependent confounders (centered values)
  - Comparison of the intercepts from this model with full model

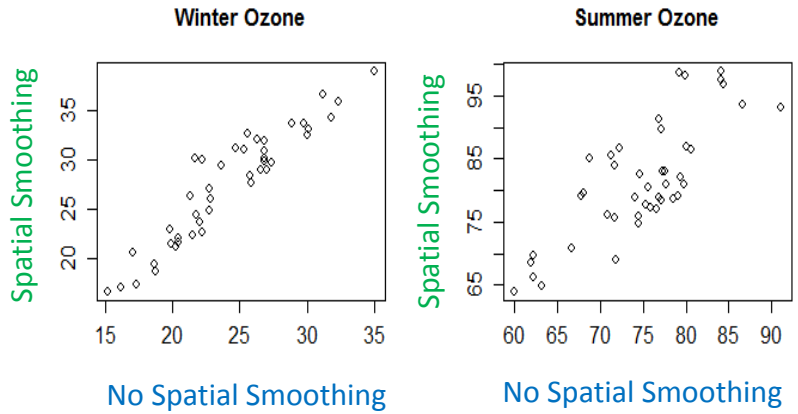
Leave-out one cross validation

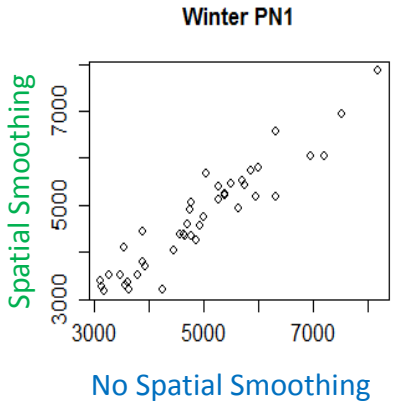
With spatial smoothing term

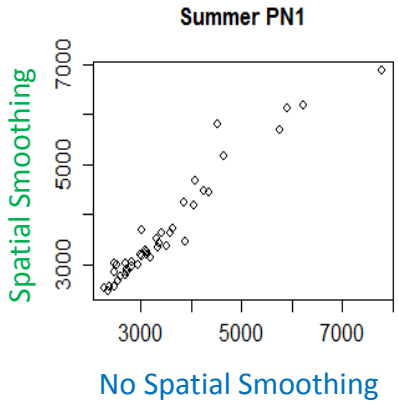
- Performs well ($R_{cv}^2 > 0.7$)
  - Ozone
  - $PM_{2.5}$ (an outlier in winter)
  - PN1
  - VOC (winter has a large outlier)
- Performs adequately (($R_{cv}^2 > 0.6$)
  - BC winter
  - NO_x
- Poor performance
  - NO_2
  - BC summer
- Terrible
  - Particle-Bound PAH

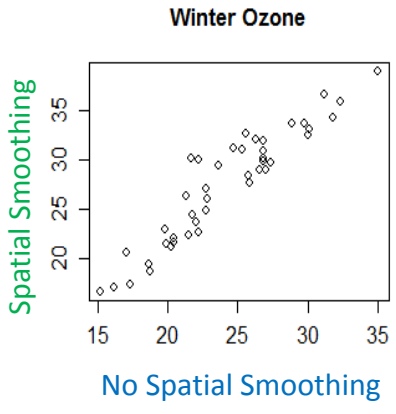
Spatial vs No Spatial Smoothing

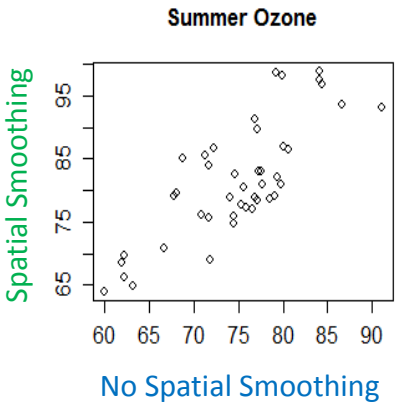










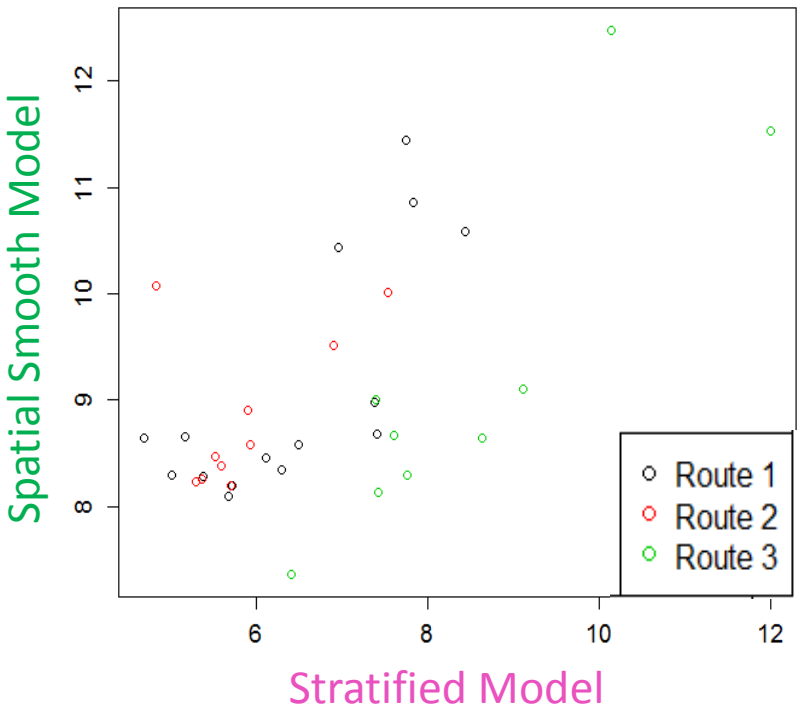


Spatial vs no Spatial Smoothing

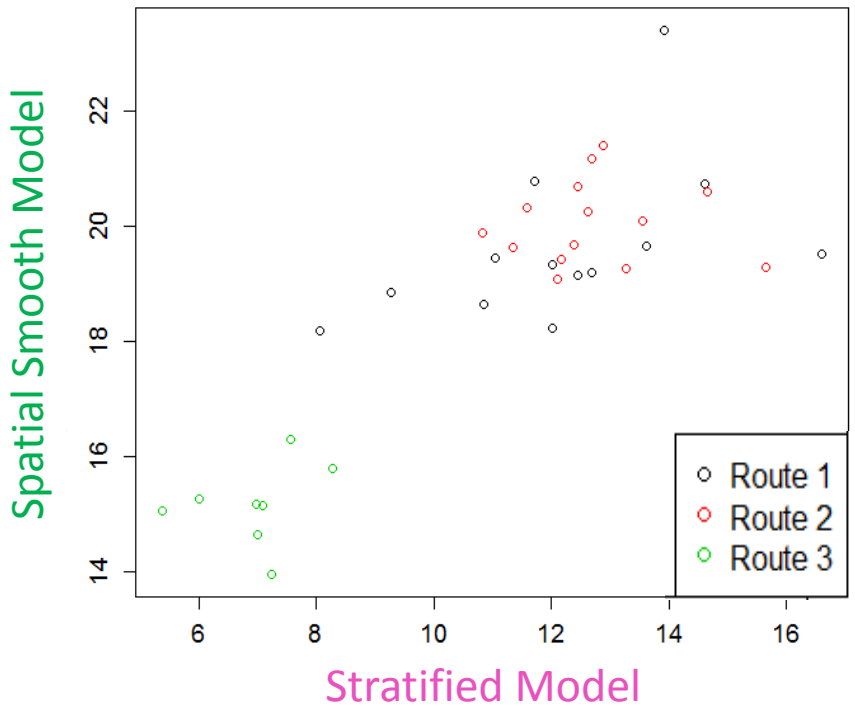
- No meaningful change in predicted concentrations
- Suggests that spatial smoothing is appropriate in this case
- Allows for correction of data not collected at fuzzy point locations
- Allows for spatial predictions, especially if model includes other covariates such as LU variables

Comparing intercept from Stratified Model to Spatial Smooth Model – $PM_{2.5}$

Winter

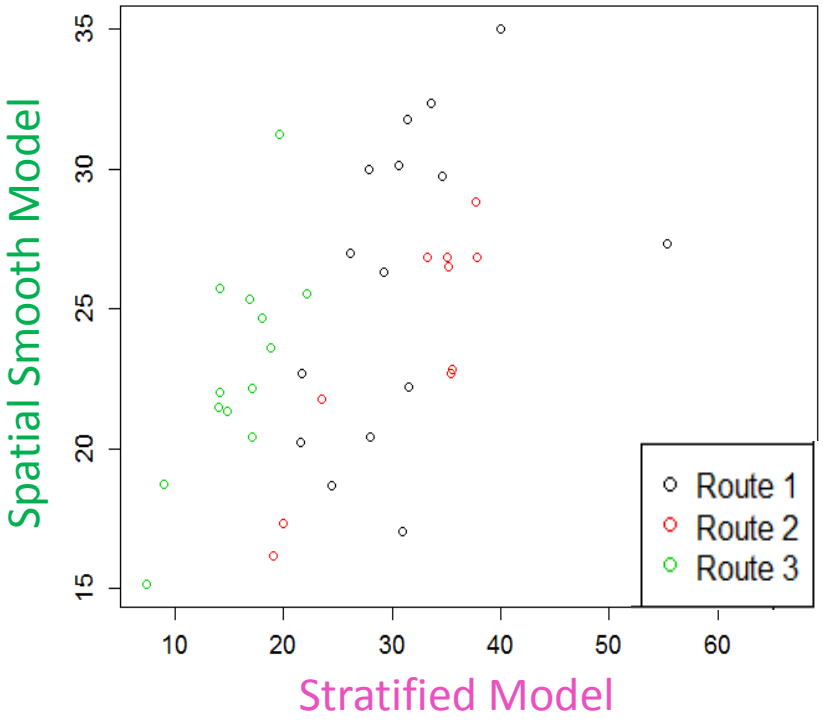


Summer

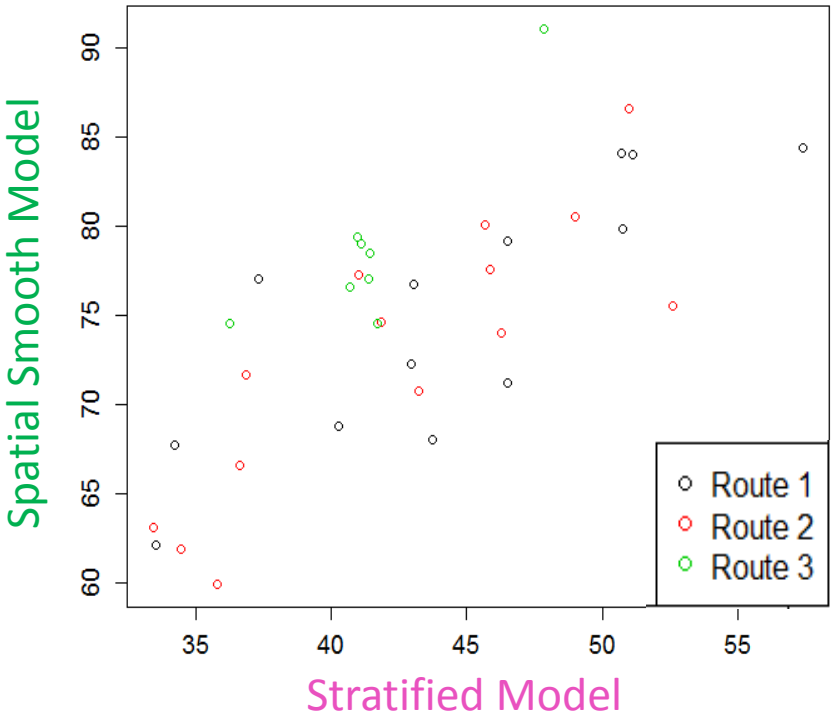


Comparing intercept from Stratified Models to Spatial Smooth Model – Ozone

Winter



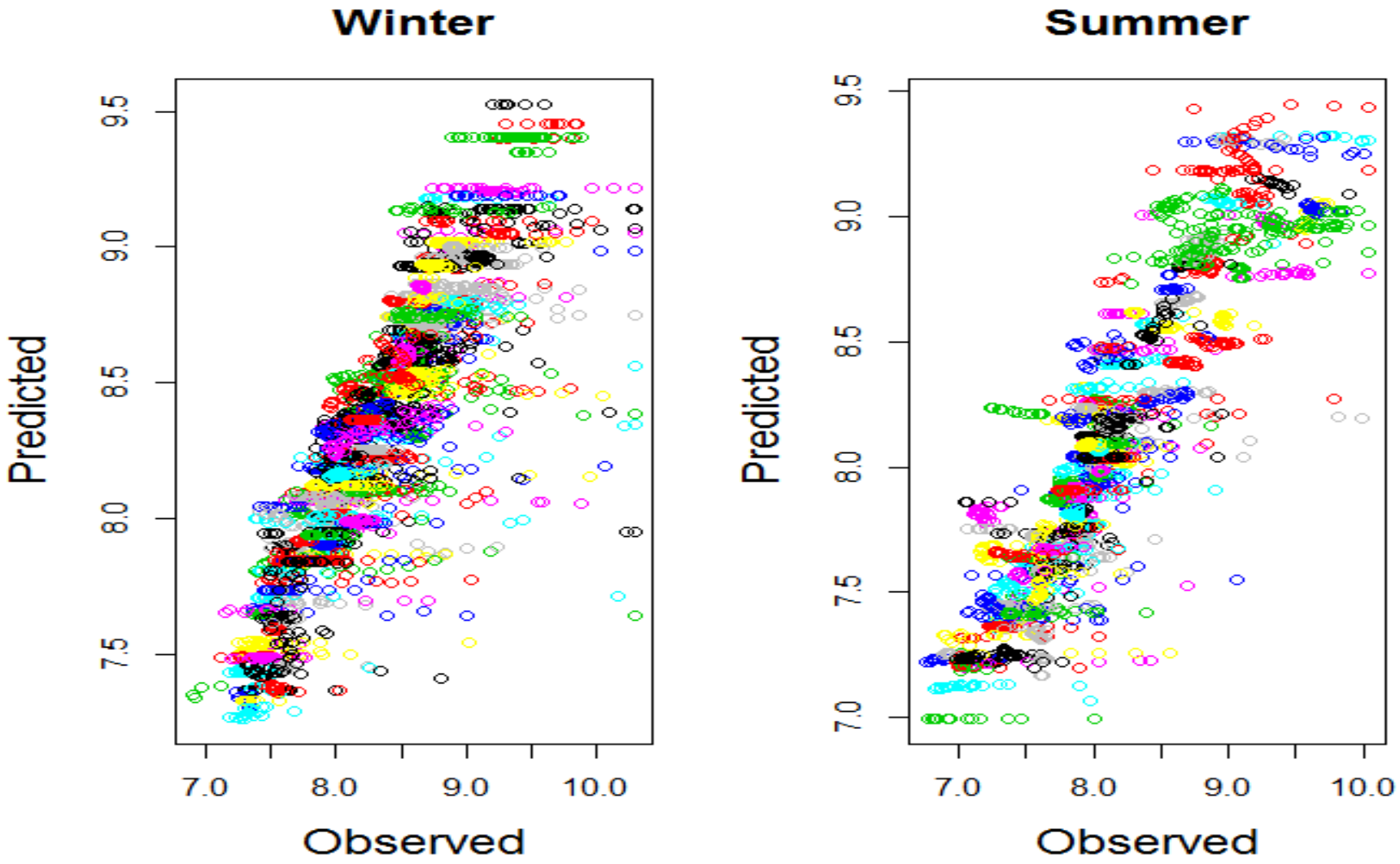
Summer



Adjusting for time varying covariates

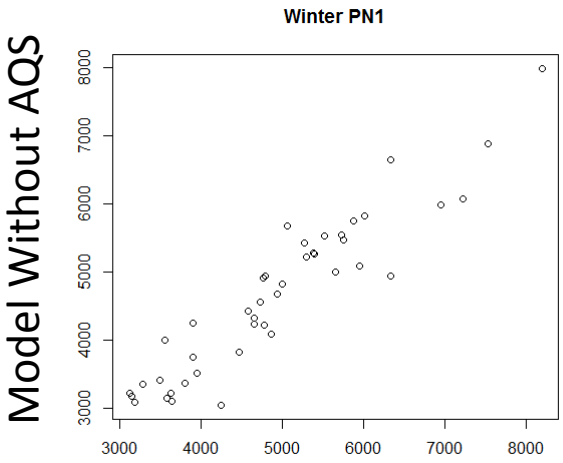
- Results suggest some agreement between stratified analysis and global analysis
- Generally, time varying covariates demonstrate similar relationships across the city
- May not be true in all cases, ex. Ozone in Winter

Within-Sample Cross Validation
Particle Number Winter (Spatial Smoothing Model)



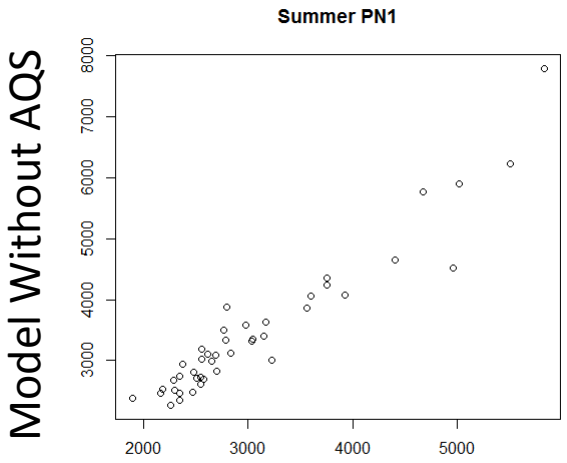
Variable Selection (removing AQS)

Winter PN



Full Model Categorical

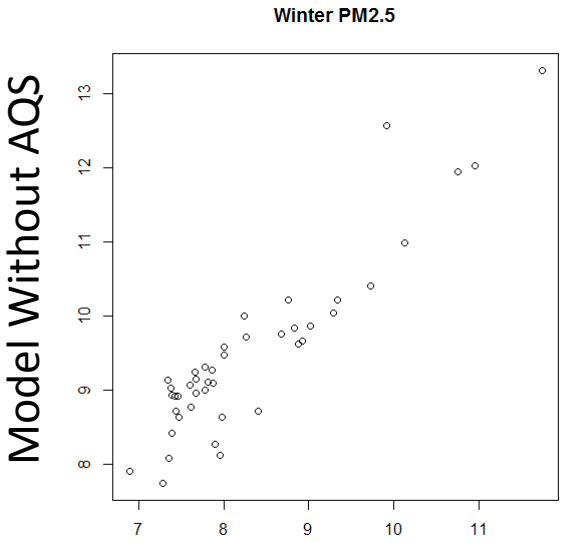
Summer PN



Full Model Categorical

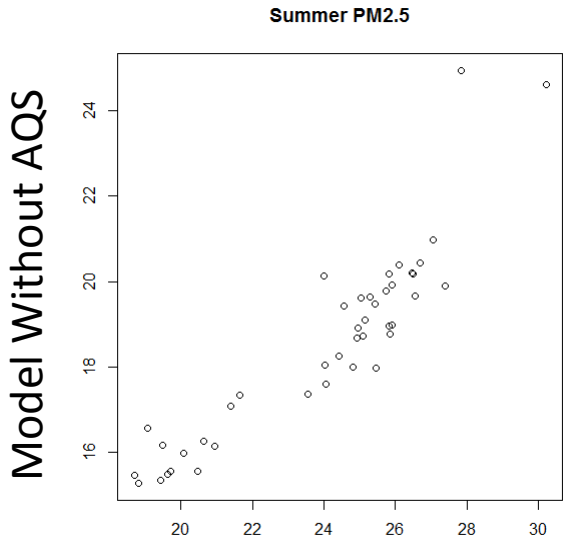
Variable Selection (removing AQS)

Winter $PM_{2.5}$



Full Model Categorical

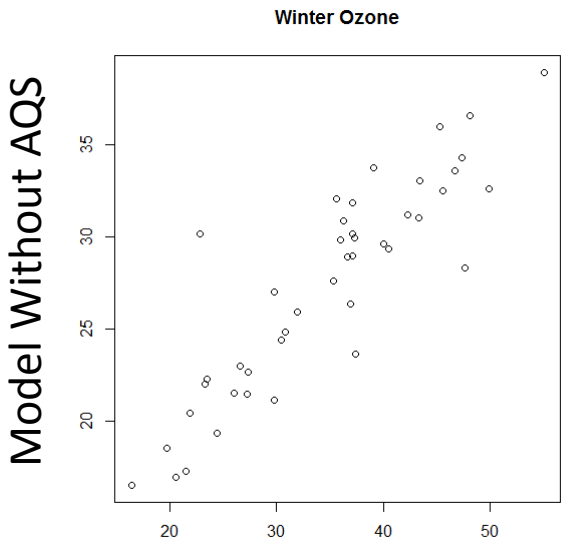
Summer $PM_{2.5}$



Full Model Categorical

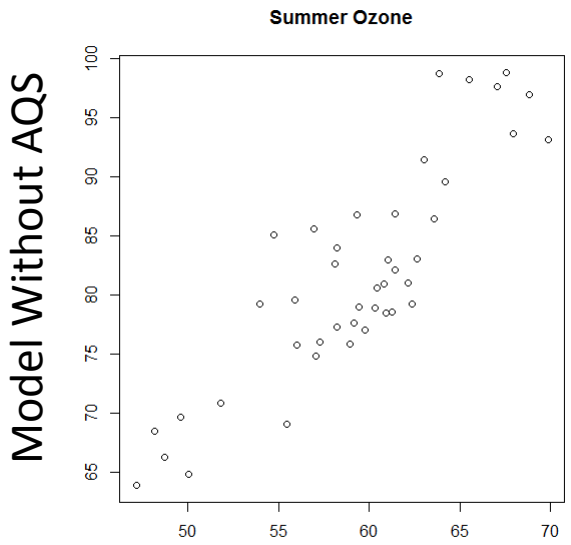
Variable Selection (removing AQS)

Winter O_3



Full Model Categorical

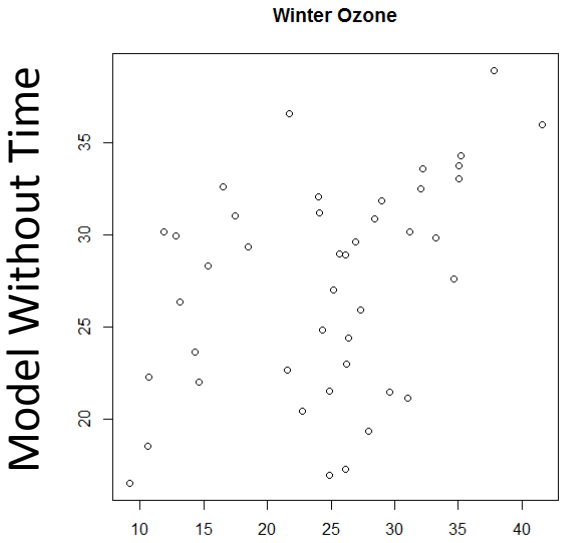
Summer O_3



Full Model Categorical

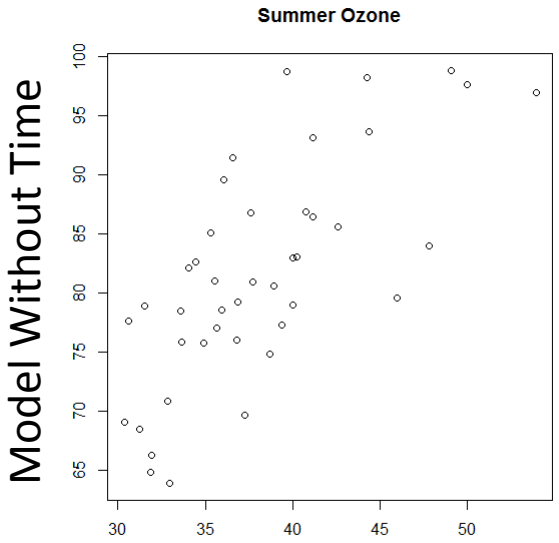
Variable Selection (removing time varying covariates)

Winter Ozone



Full Model Categorical

Summer Ozone



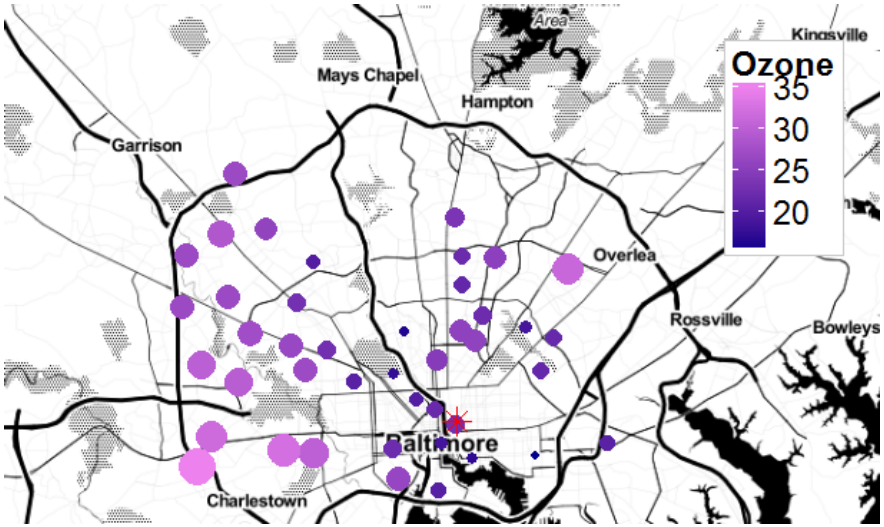
Full Model Categorical

Variable Selection

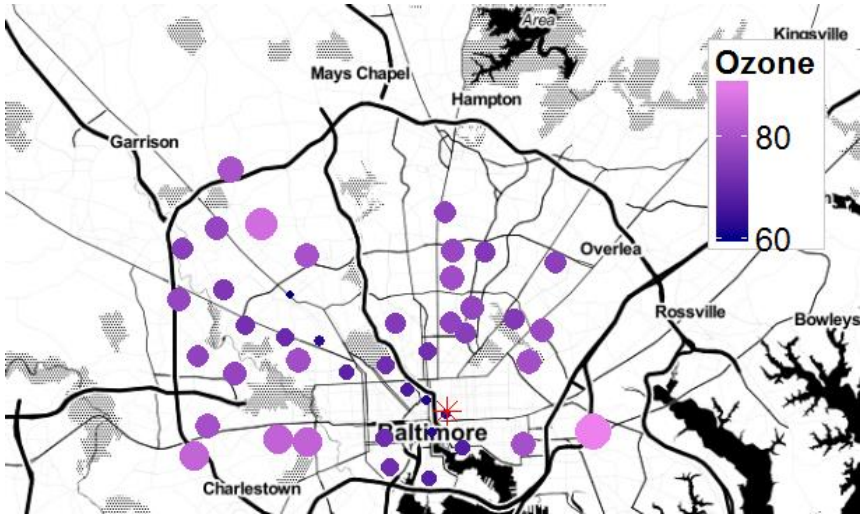
- Model outputs are stable when AQS or Weather variables are excluded
- Excluding both AQS and Weather covariates results in different predictions
  - Confirms the importance of these time varying covariates in obtaining spatial distributions

Spatial Distributions - Ozone

Winter



Summer



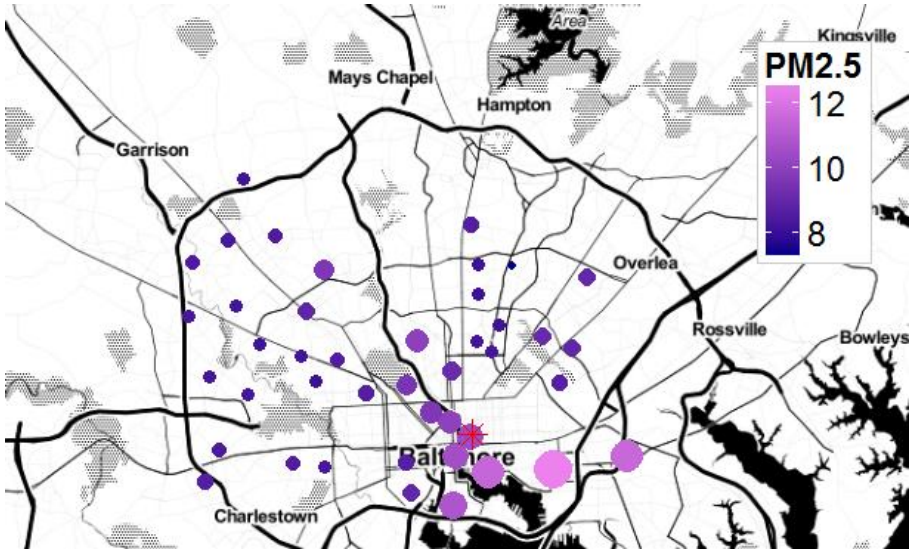
Spatial Distributions - NO_x

Winter

Summer

Spatial Distribution $PM_{2.5}$

Winter



Summer

Conclusions

- An additive model allows for time varying trends to be removed from data set
- Spatial means can be calculated that account for broad regional changes
- Spatial trends are consistent with and without spatial smoothing components
- Relationship between time-varying covariates and spatial locations are reasonably stable

Next Steps:

- Create model without spatial component
- Use residuals to investigate spatial distribution of pollutants, and in particular the association time-adjusted 30 second measurements and Land Use covariates

Acknowledgements

- Erin Riley
- Lianne Sheppard
- Michael Yost
- Timothy Larson
- Adam Szpiro
- Paul Sampson
- Sverre Vedal

- This publication was made possible by USEPA grant **RD-83479601**. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA. Further, USEPA does not endorse the purchase of any commercial products or services mentioned in the publication.
- Research reported in this presentation was supported by the National Institute of Environmental Health Sciences of the National Institutes of Health under award number **T32ES015459**. Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Adjusting for regional scale temporal variability in mobile monitoring data

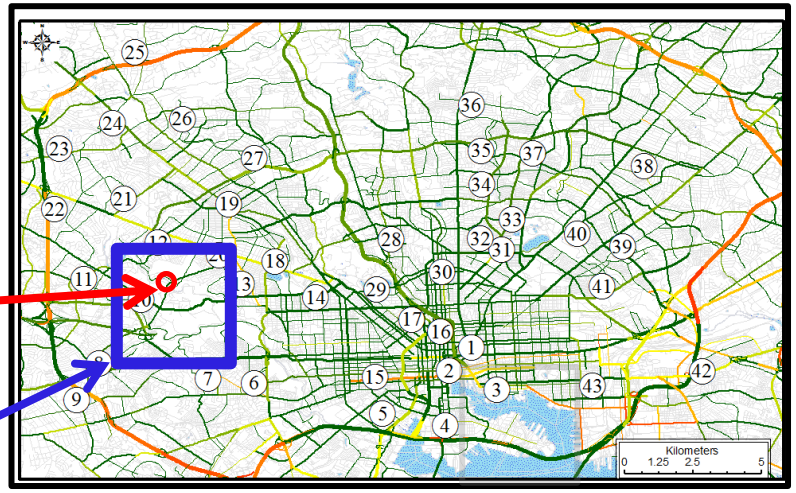
What are the monitoring sites measuring?

Our model (log scale)

$$Y(s,t) = N(s,t) + U(t) + L(s,t) + \epsilon$$

L(s,t): Local
[constant within ~ 500 m]

N(s,t): Neighborhood
[constant within ~ 4 km]*



U(t): Urban [within ~ 30 km]

Data adjustment goals:

- We want to remove variability due to time
- Would *like* an algorithm that can discern between *U(t)* vs. *N(s,t)* *vs. L(s,t).*

* Consistent with EPA monitor definition

Relative importance of U(t) vs N(s,t) varies by pollutant

Urban: Var (U) > Var(N)

- $PM_{2.5}$ metrics (neph, GRIMM)
- Ozone

Neighborhood/ Local: Var(N)> Var (U)

- CO
- NO_x
- Black Carbon
- VOCs
- Nanoparticle (25 nm – 400 nm) count
- PN1 (< 1 micron diam) number count
- PAH
- Coarse particles (> 3 micron)

Our Initial approach: Adjust for temporal variability using measurements at a single fixed site

Assumes fixed site at location s_o provides a continuous measure of temporal variability of a given pollutant across the entire urban area. We adjust for temporal variability assuming it is represented entirely by U(t).

Estimate U(t) from fixed site, s_0, as:

$$U(t) = \overline{Y\ (s_o, t)} \ - \ mean\{Y(s_o, t - 15\ m) \ldots Y(s_o, t + 15\ m)\}$$

Average over the route for all days; accounts for day to day variability

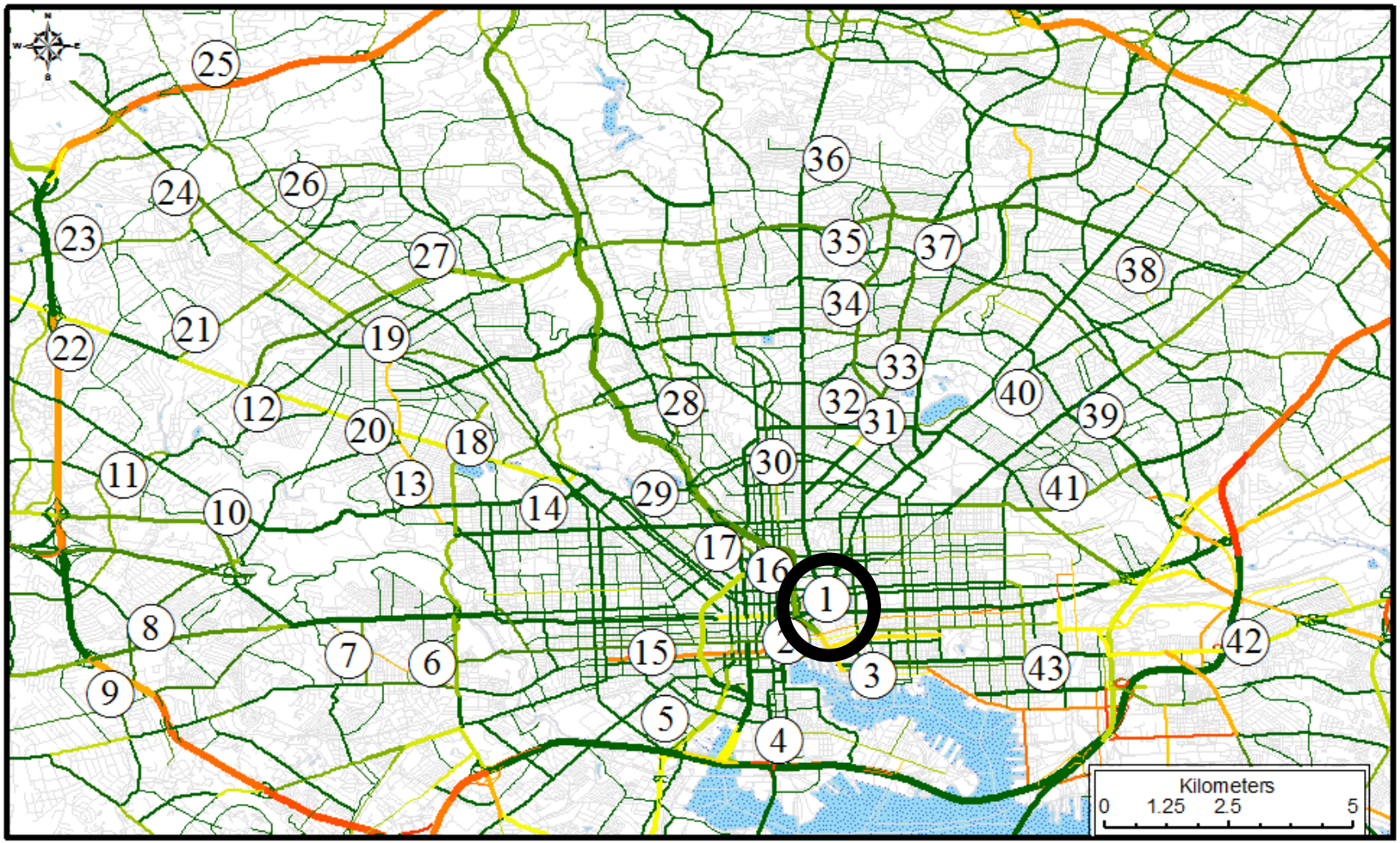
Average over moving 30 minute window; accounts for within day variability

where s_o is away from local influences such that $L(s_o,t) \approx 0$.

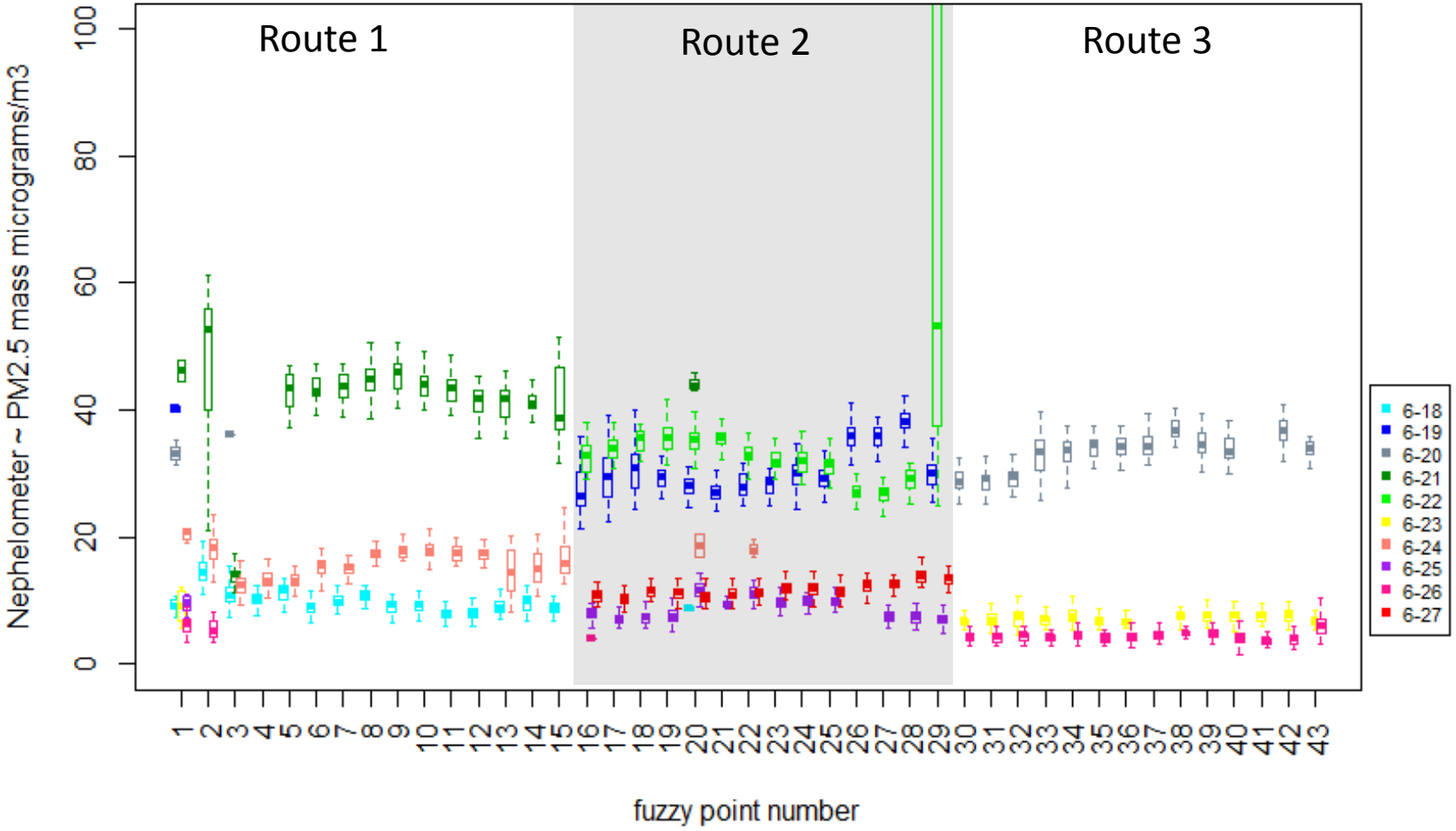
The temporally adjusted mobile measurement at a given location and time is then assumed to only be due to neighborhood and local scale contributions

$$Y^{adj}(s) = Y(s,t) - U(t)$$

Baltimore fixed site and fuzzy point locations. Our fixed site was co-located with an AQS site on the roof of a building downtown.



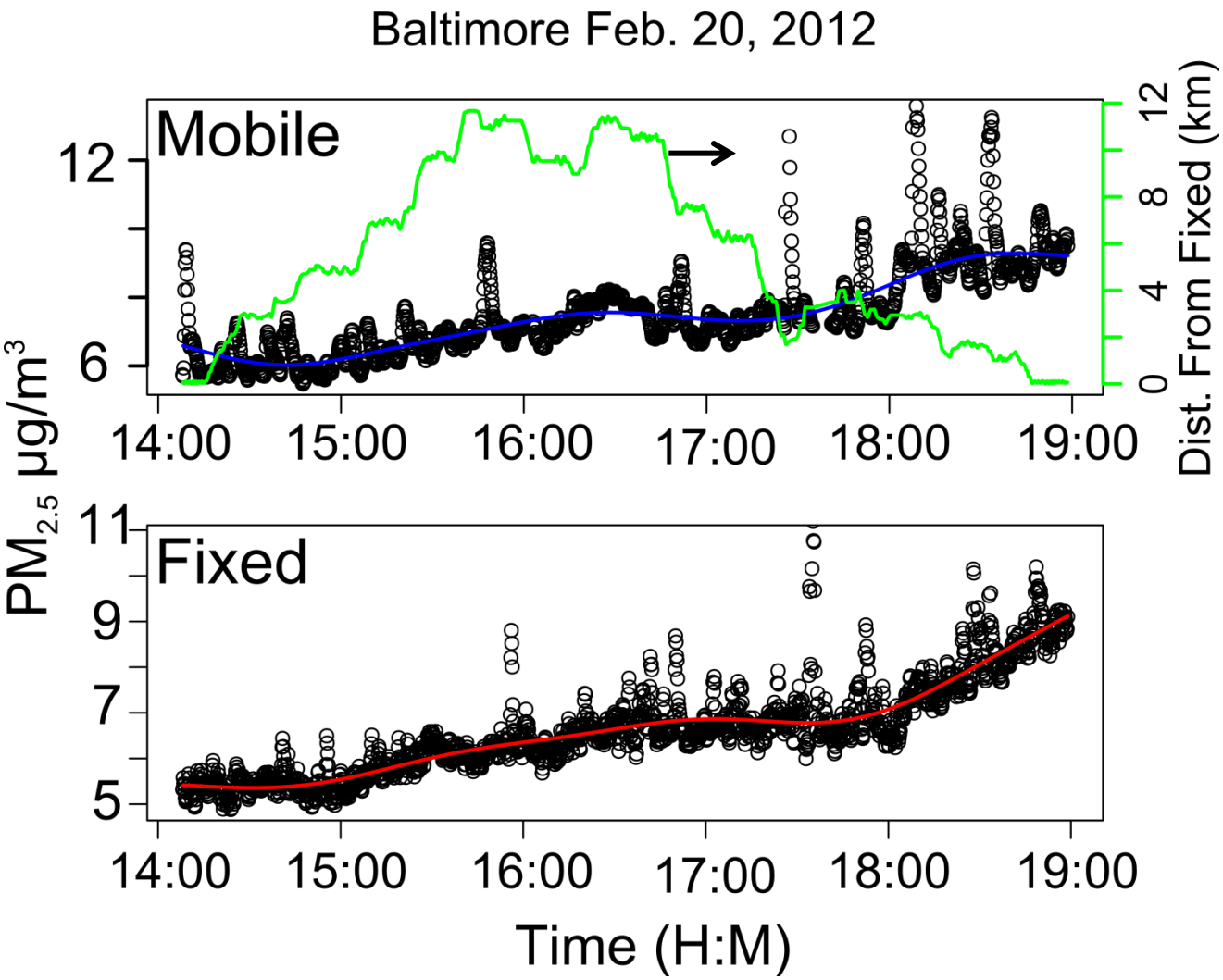
For urban scale pollutants such as $PM_{2.5}$ most of the temporal variability is due to between-day changes



Baltimore summer campaign: Mobile platform 10 s data summarized at each fuzzy point by day

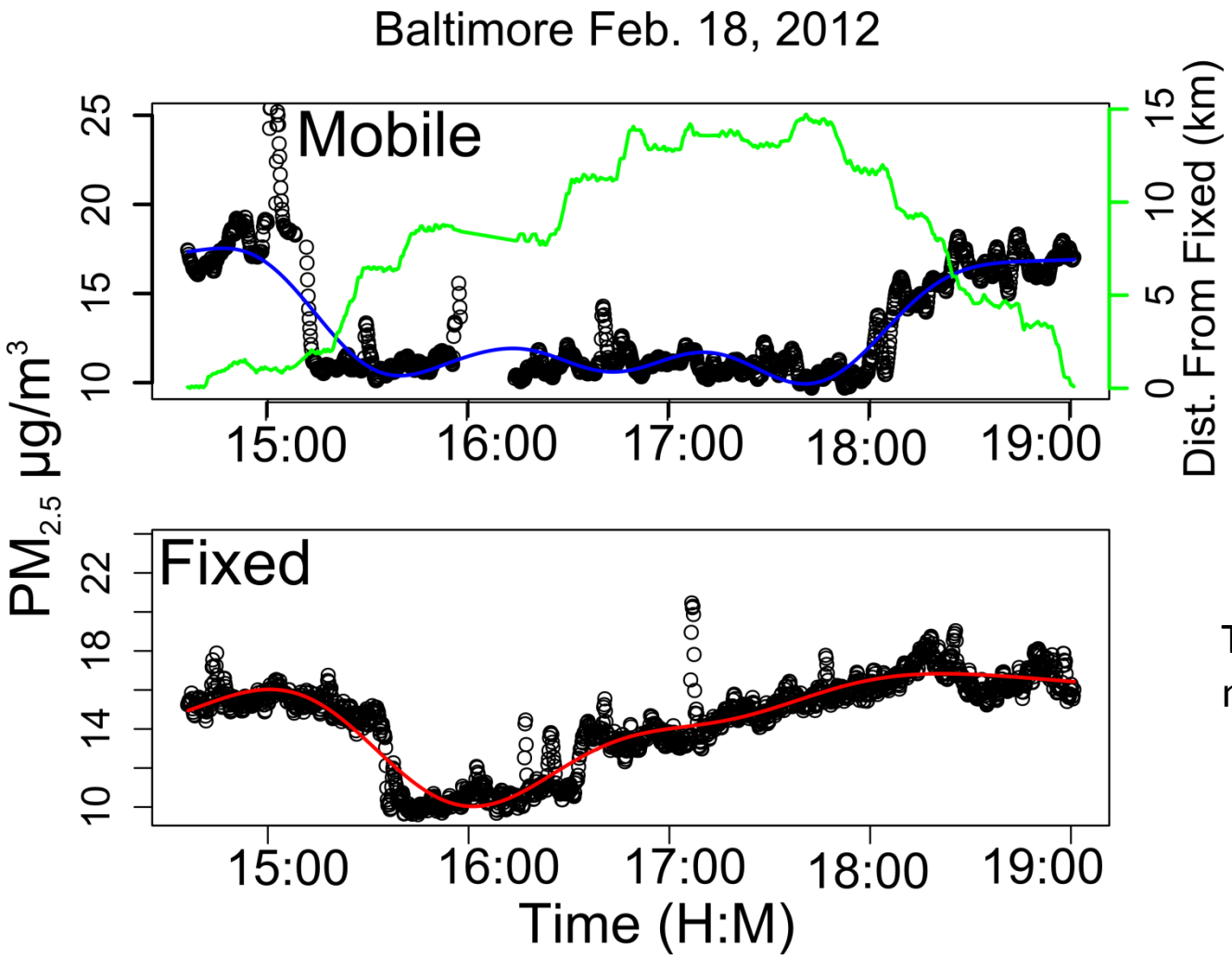
However, $PM_{2.5}$ changes are also significant within a day across all sampled locations

On this day the urban scale variability contributions as observed by the mobile platform (blue trend line) tracks well with the fixed site (red trend line)



Trend lines are modeled means

However, on this day the fixed site trend line does not correspond to that from the mobile data.

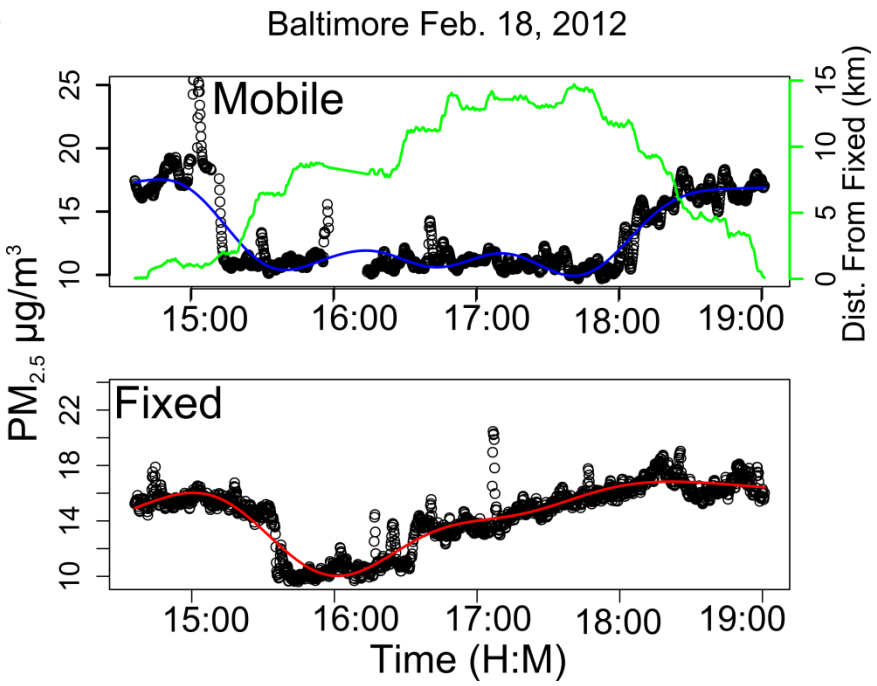


Trend lines are modeled means

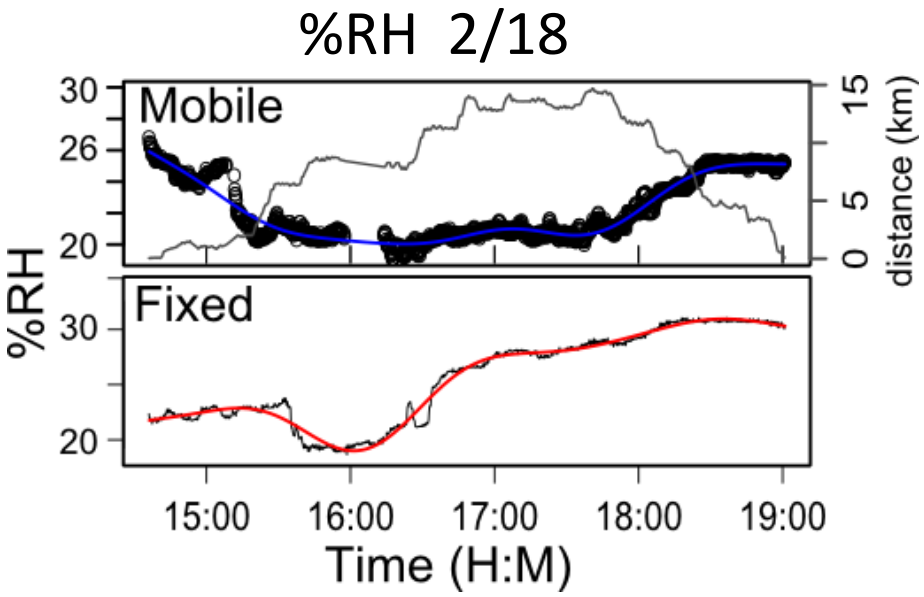
What is causing the differences between platforms?

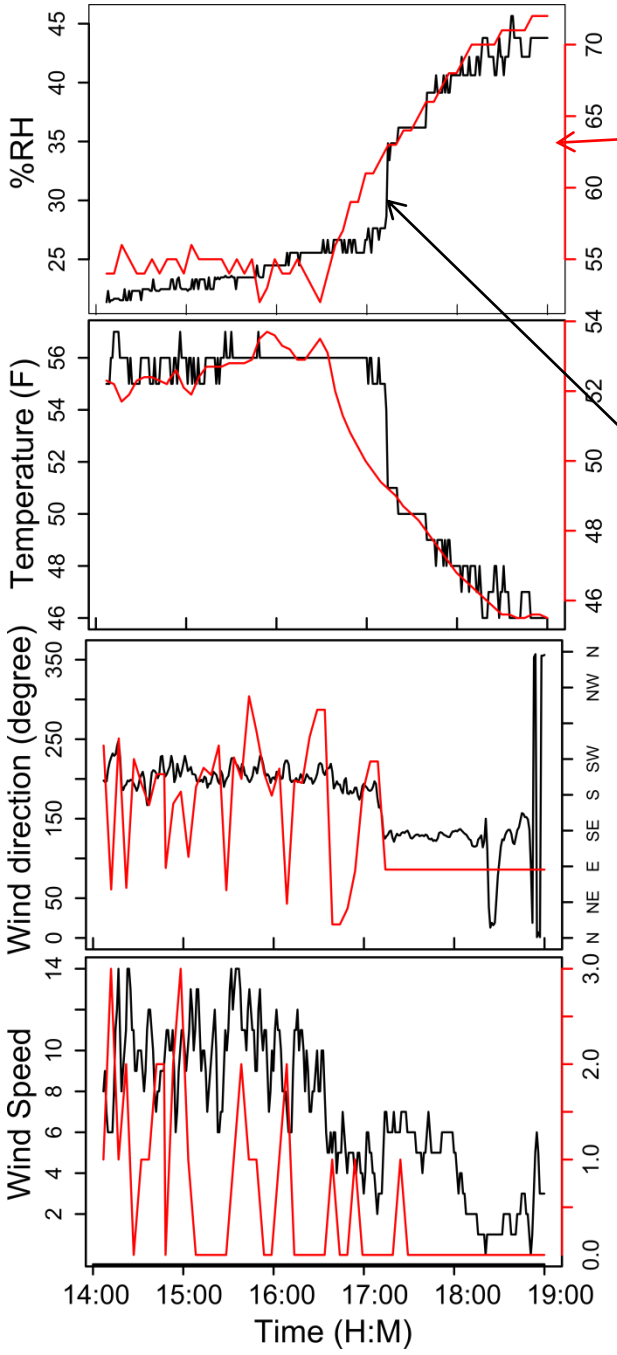
N(s, t) ?

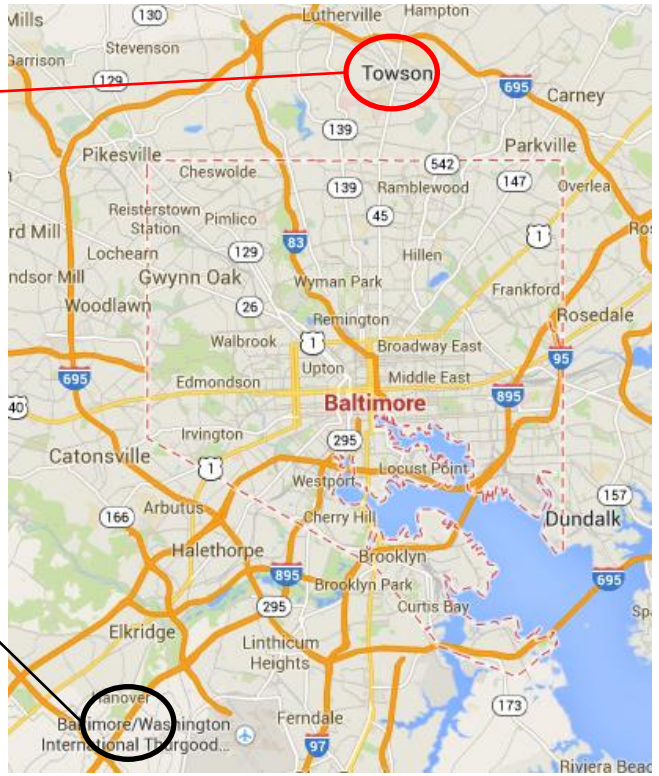
L(s,t) ?



Evidence of a frontal passage during this period suggests it affected spatio-temporal variability across the urban area, i.e., N(s,t)







- Regional pollutants dominated by atmospheric transport
- Mixing changes throughout the day
- Temperature and relative humidity are known predictors of changes in regional air mass (frontal passage?)

Issues with fixed site corrections:

1) Fixed site does not represent temporal variability across all sites:

- weather fronts passing through urban area
- variations in neighborhood scale emission densities across urban area
- Impact of local sources at fixed site

2) Practical: Missing data at fixed site

Can we use the mobile monitoring data for temporal adjustment?

Estimate U(t) + N(s,t) using mobile monitoring data with a "boxcar" filter:

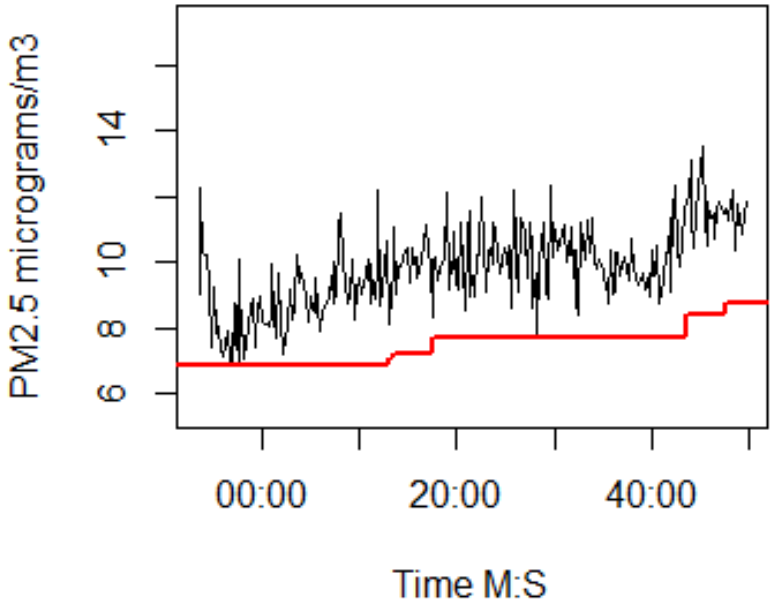
$$\mathrm{U(t)} + N(s,t) = MIN\{Y(\,s,\ t - 15\,m) \ldots Y(s,\ t + 15\,m)\}\ **$$

- Assumes 30 minute minimum represents urban scale variability along route

- Assumes 30 minute window is small enough that changes in background are eventually felt everywhere across route (delayed in time, but magnitude is the same)

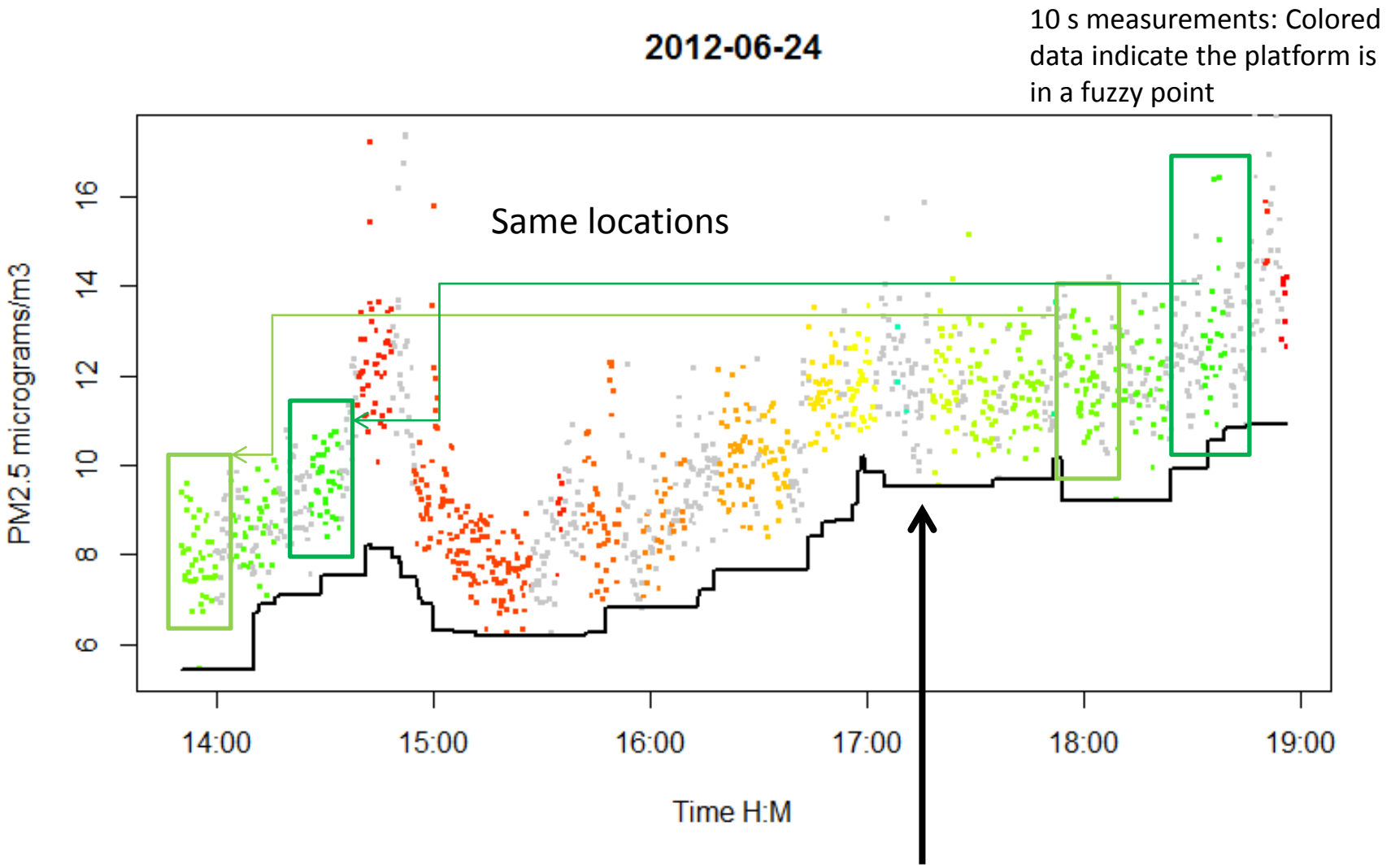
<u>Disadvantage:</u>

Does not *fully* separate *N(s)* from *N(t)*
- problematic when N(s,t) is > U(t)



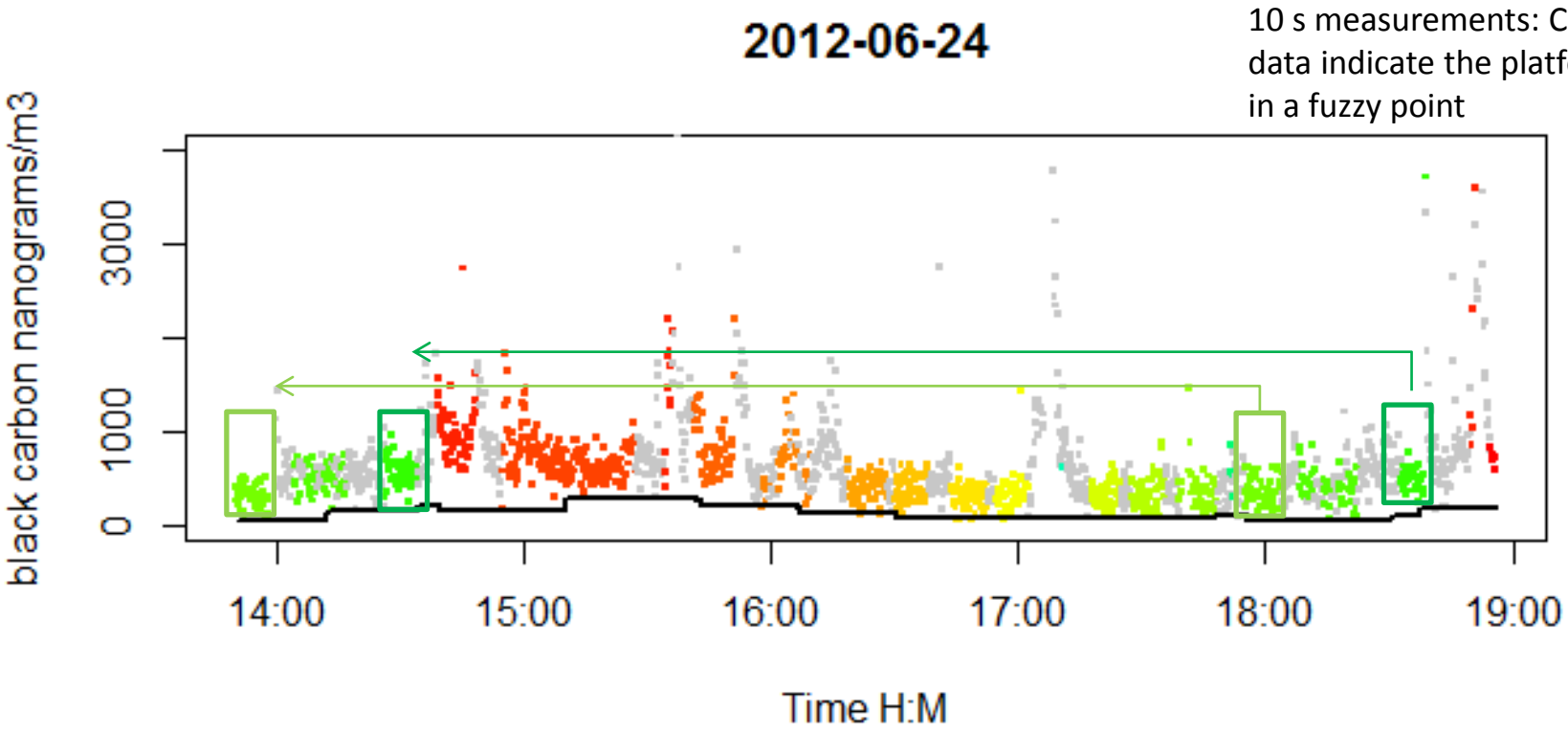
** Ozone use MAX{…}

$PM_{2.5}$ Example: Background varies in time but local/neighborhood impacts at the same site do not



$$U(t) + N(s,t) = MIN\{Y(\,t - 15\,m) \ldots Y(t + 15\,m)\}$$

Black Carbon Example: U(t) + N(s,t) does not vary much over time or space

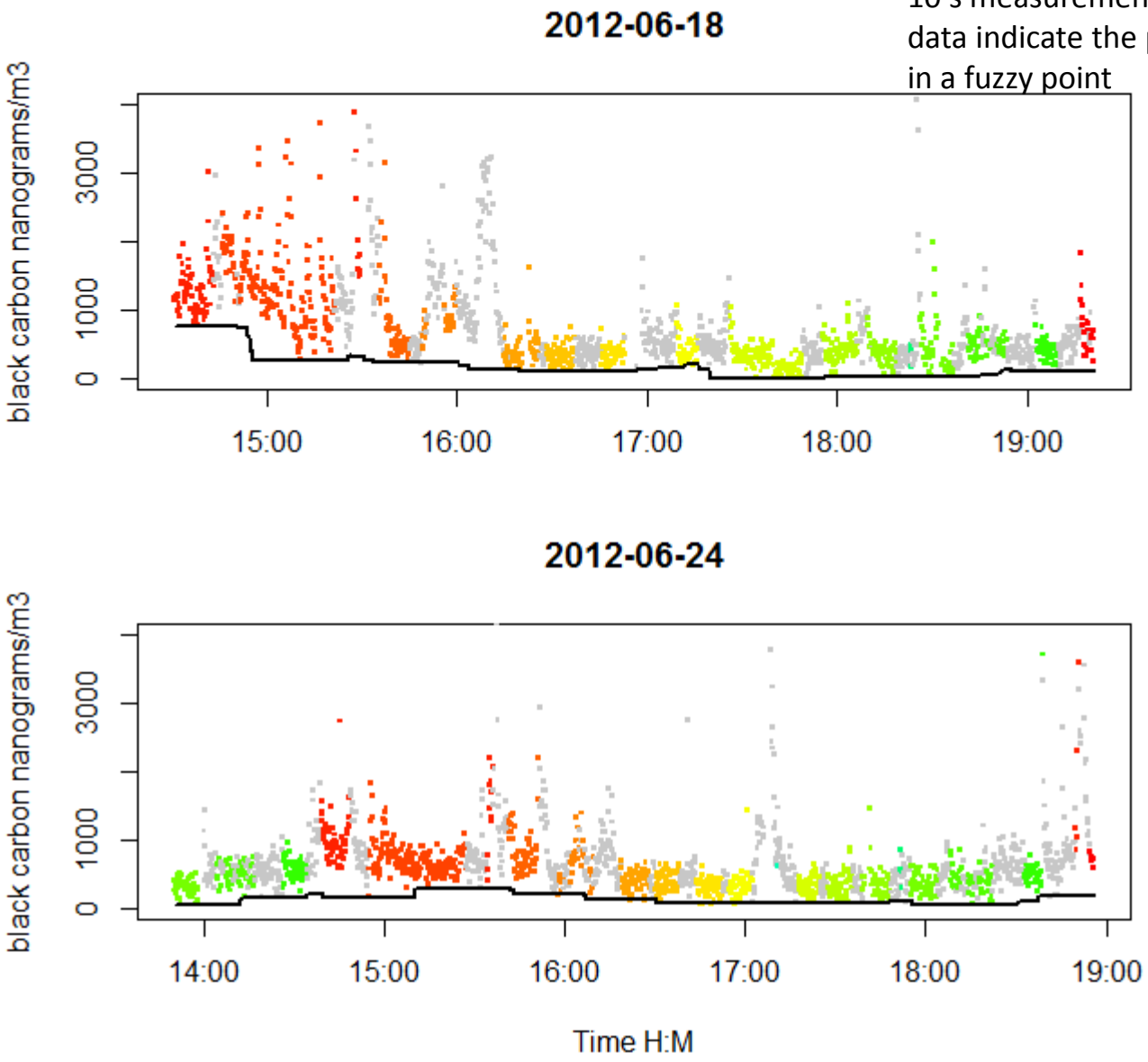


Adjusting for temporal variability using background may over adjust by removing part of both L(s,t) and N(s,t)

Black Carbon Example: Evidence that Boxcar Filter is Removing N(s,t) at some sites

10 s measurements: Colored data indicate the platform is in a fuzzy point

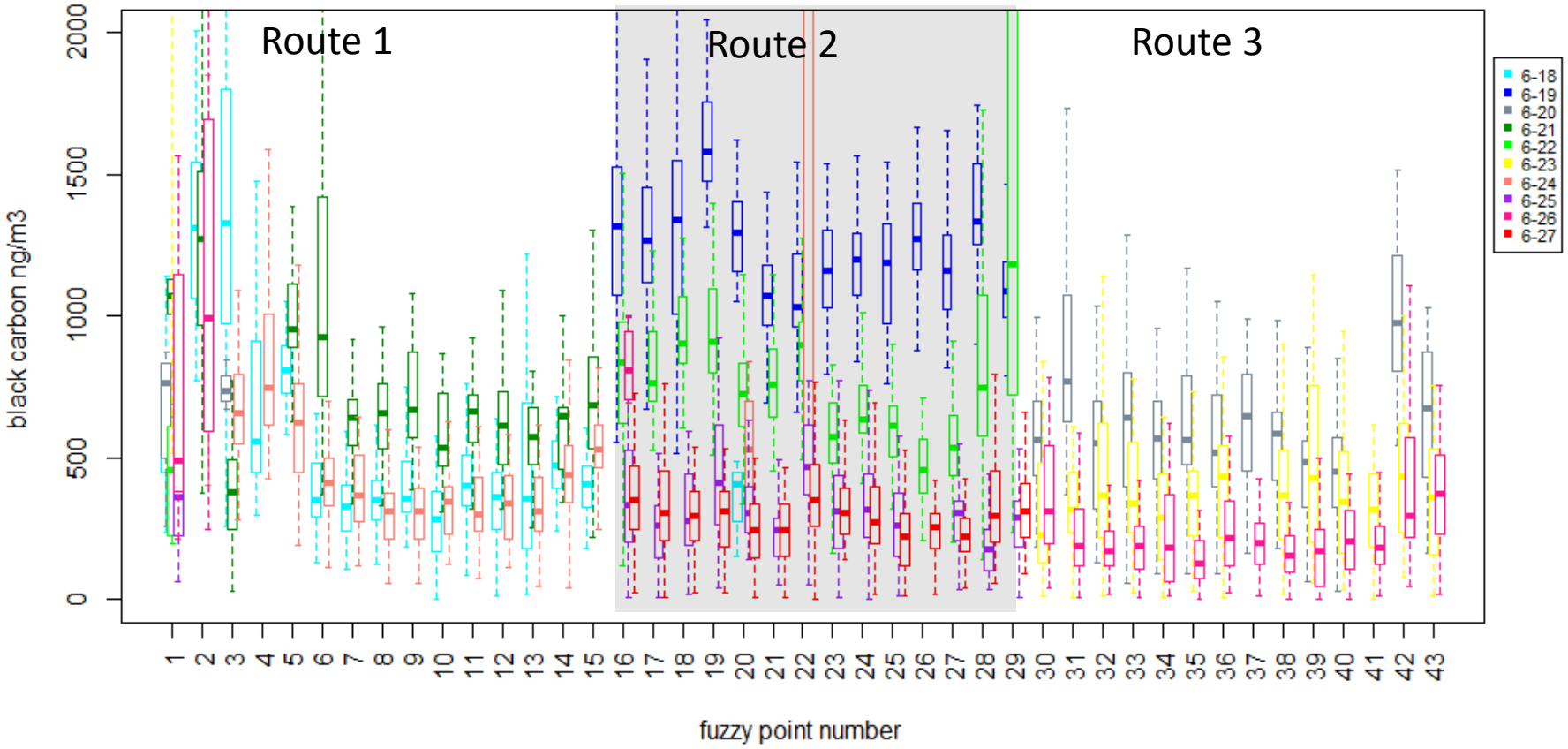
Similar distributions measured on different days



Local/ Neighborhood pollutants – Between day

Between-day variability accounts for less variance than it does for “Urban” pollutants

Baltimore summer campaign: Mobile platform 10 s data summarized at each fuzzy point by day



Is the boxcar filter appropriate for temporal adjustment?

Urban: Var (U) > Var (N)

- $PM_{2.5}$ metrics (neph, GRIMM)
- Ozone

Boxcar adjustment appropriate with caveat: does not insure complete separation of U(t) from N(s,t)

Neighborhood/ Local: Var(N) > Var(U)

- CO
- NO_x
- Black Carbon
- VOCs
- Nanoparticle (25 nm – 400 nm) count
- PN1 (< 1 micron diam) number count
- PAH
- Coarse particles (> 3 micron)

Boxcar adjustment removes both U(t) and N(s,t) – not appropriate

Conclusions:

- Fixed site does not represent temporal variability across all sites, need new method
- For local/neighborhood pollutants, “Boxcar background” does not discern between N(s,t) and U(s,t) – over-adjusts
- Where N(s,t) << U(s,t), as is the case for $PM_{2.5}$ and O_3, “boxcar background” adjustment not as problematic

Another approach:

Regress data on important temporally varying meteorological factors, AQS data, etc.

Discussion by Dr. Austin

Update:
Using the Mobile Monitoring Data for Epidemiology

Outline

- Introduction
  - Value and success of mobile platform
- Nature of the mobile monitoring data
  - Key challenge
- Overview of goals and approach from the original proposal
  - Insights to date
- Overview of planned approaches
  - Remove urban-scale variation
  - Spatial prediction
- Approaches to removing urban-scale variation
  - Overview
  - Break for talks by Erin Riley, Elena Austin
  - Discussion
- Plans for spatial prediction
- Discussion topics and further opportunity for SAC advice

Introduction

- Mobile monitoring is useful tool given limited resources for sampling and data management
  - “Infinite” number of spatial locations – better informs land use regression
  - Complex small-scale variation can be captured for multiple pollutants at the same time
  - Allows deep insights into the nature of on-road exposures
- Key questions:
  1. Do we directly exploit the small-scale variation in these data because it is meaningful or average over it first?
  2. How do we cope with the space-time confounded design?

Introduction

- Examples of successful uses of the mobile monitoring platform data to date:
  - Understand local variation of traffic-related pollutants in time and space
  - Determine single- and multipollutant roadway gradients
  - Better understand on-road exposures
  - Provides information needed to develop an ozone exposure prediction model
  - Allows estimation of airplane-generated UFP spatial distribution (completed in LA and Atlanta)

Nature of the mobile monitoring data

- Timing: Sample every 10 seconds on roads during the afternoon rush hour over a 2-week period
  - Aggregate to every 30 seconds for analysis
  - Up to 12 sampling days
- Locations:
  - 43 fuzzy points sampled along 3 fixed routes with approximately 15 fuzzy points per route
  - 1 fixed site with the same platform, also visited by the mobile platform on every route (FP #1)
  - All locations sampled along the fixed routes, but fewer data points at any single location not in a fuzzy point
- Data: Vectors of 12 pollutants (with ability to add 32 particle size bins)

Key challenge: space-time confounded design

Overview of the goals from the original proposal: Uses of the mobile platform data

- Develop spatial models of selected roadway-source air pollutants to produce city-wide exposure surfaces for traffic-derived air pollution components
  - Apply methods developed by the Biostatistics Core
- Contribute to understanding distributions of traffic-derived air pollutant concentration estimates for various roadway types and traffic conditions

Approach to spatial modeling of the mobile data from the original proposal

1. Remove temporality in the mobile measurements by using data from the fixed site as:

$$C_{15\,\mathrm{sec}}^{adjusted} = C_{15\,\mathrm{sec}}^{mobile} \left\{ \frac{C_{\mathrm{overall}}^{fixed}}{C_{30\,\mathrm{min}}^{fixed}} \right\}$$

2. Summarize the adjusted concentration distribution at fuzzy points
   - E.g. median
   - Data become purely spatial
3. Predict by exploiting the spatial relationship between concentrations and land use covariates
4. Combine predictions from the two seasons to estimate the long-term average

Insights to date

- The temporal structure we want to remove is not well represented by the fixed site on all days and for all pollutants
  - It is not purely a temporal pattern that we wish to remove
  - One location is an imperfect measure of the information we want to remove
- In this dataset, given its space-time confounded design, adequately removing the confounding information while preserving the information of interest is inherently an extremely challenging problem

Overview of plans for estimating spatially-varying long-term averages

1. Spatial prediction from summary measures at the 43 fuzzy points
   - Relies on only 43 locations
2. Spatial prediction from rich space-time data from the full network
   - Allows focus on very fine-scale spatial variation not restricted to fuzzy points
   - Use geocovariate data from the midpoint of a 30-second sample
     - Up to 800 covariates available (before elimination of inappropriate covariates; see slide 17)

Overview of plans for estimating spatially-varying long-term averages

- Data to be modeled:
  - Single pollutants, particularly $PM_{2.5}$, NOx, LAC, O_3
    - Allows direct comparison of results with other predictions at subject homes
  - Multiple pollutants after dimension reduction using
    - Principal component analysis (PCA)
    - K-means clustering
  - Predictive sparse PCA or k-means clustering

General strategy: Conceptualization of data

- View each pollutant as having *urban*, *neighborhood*, and *local* scale structure
  - Terminology loosely corresponds to EPA definitions for regional (> 50 km), urban (up to 50 km), and neighborhood(½ – 4 km) scales
  - Relative variation of each differs according to pollutant due to meteorology, chemistry, sources

General strategy: Modeling approach

- Remove all urban scale features. Justification:
  - Urban scale or greater is only dimly related to traffic sources
  - We believe any urban-source spatial variation will be averaged out if we could collect data over a long time period
  - Operational question: Is this purely temporal variation?
- Prefer not to remove neighborhood scale features
  - Includes some spatial information thought to confound the relationship of interest (e.g. space-time varying weather)
  - Operational question: Can we remove only this confounding while preserving the neighborhood scale information of interest?
- Conduct spatial analyses of adjusted data that represents local plus neighborhood scale features

Removing urban-scale features

- Model for pollutant Y (on log scale) at location s, time t:

$$Y(s,t) = U(s,t) + N(s,t) + L(s,t) + \varepsilon$$

where $U(s,t)$ represents urban scale, $N(s,t)$ neighborhood scale, $L(s,t)$ local scale

- Investigate different approaches for removing urban-scale variation:
  - Correct using fixed site measurements (proposal; noisy and possibly biased estimate of $U(s,t)$; reviewed by Riley)
  - Local adjustment by subtracting the 30-minute minimum (boxcar filter; works well for urban scale pollutants; is less useful for local scale pollutants; discussed by Riley)
  - Pure temporal adjustment using regression (regression estimate of $U(t)$ or $U(s,t)$; utility still under investigation; discussed by Austin)

Break for Riley & Austin presentations

Follow-up discussion on removing urban-scale variation

- Fixed site adjustment no longer being considered
- Boxcar filter approach may be appropriate in some settings
  - Useful for urban scale pollutants
  - Less appropriate for neighborhood and local scale pollutants
- Regression adjustment still under development
- Alternative idea to explore: remove urban-scale variation using a deterministic model
  - Direct connection to downscaling

Plans for spatial analyses

- Use adjusted data for spatial regression
  - Key challenge is how to adjust
  - Apply to single pollutants or multipollutant profiles (refer to slide 10)
- Spatial regression with
  - Land use covariates, simplified using PLS or PCA
  - Possibly also spatial smoothing
- Predict at subject residences

Discussion topics

- Selection of land use covariates
  - Rationale: Use only informative covariates given these data
    - Data are collected on roads
    - Fuzzy geographic locations, particularly fuzzy at fuzzy points
    - Reduce sensitivity to residual temporal confounding
  - Current plans:
    - Drop distance to road covariates
    - Exclude buffers smaller than 100 meters
      - Particularly important for fuzzy point data – want to exclude buffers smaller than the radius of the fuzzy point

Discussion topics

- Decide whether to incorporate spatial smoothing in the prediction modeling
  - Contains useful information we want to preserve?
  - Model will be applied to a larger geographic area than mobile monitoring region
- Determine how to combine the two campaigns to produce long-term averages
  - Before vs. after prediction

Discussion topics

- Given the structure of the data, removing temporal confounding is difficult and may not be entirely successful
- Approaches to determining validity of the results
  - Cross-validation focusing on spatial information
  - Comparison with predictions at subject locations ($PM_{2.5}$, LAC, Nox, O_3)
    - Long-term average estimates
    - Corresponding 2-week averages
  - Sensitivity analyses (e.g. how do results change with different adjusted datasets?)

Thank you!

CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

CCAR PROJECT 4

Vascular Response to Traffic-Derived Inhalation in Humans

Effect of Commute Traffic on Vascular Function

CCAR PROJECT FOUR

Effect of Commute Traffic on Vascular Function

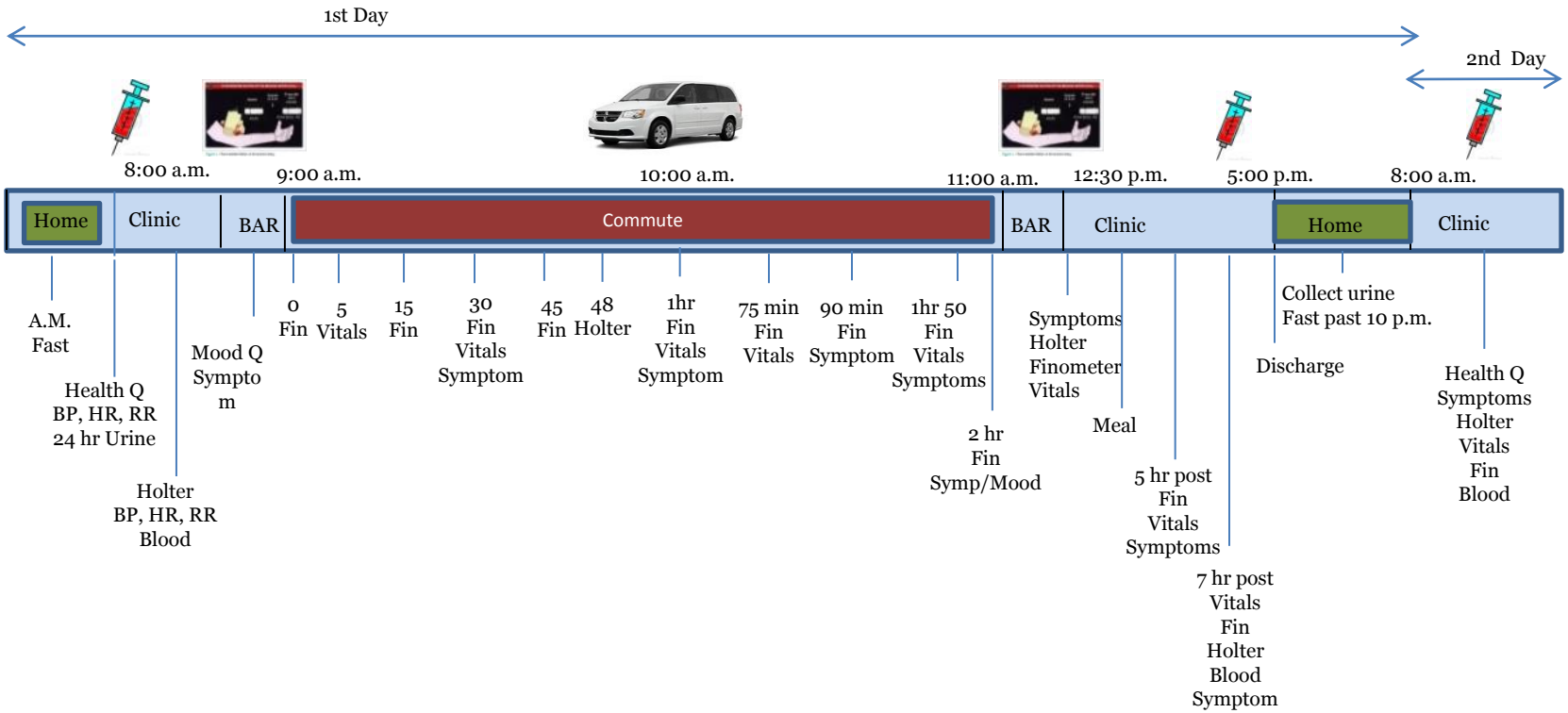
Hypothesis 1 (Specific Aim 1): Acute exposure of human subjects to combustion-derived pollutants in a typical commute will result in brachial artery vasoconstriction, retinal arteriolar narrowing, and increased systolic blood pressure.

Hypothesis 2 (Specific Aim 2): Acute exposure of human subjects to traffic-related combustion- derived pollutants will result in evidence of lipid peroxidation and pro-atherogenic gene transcription and epigenetic changes.

CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

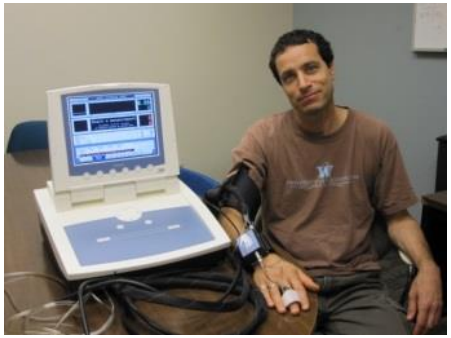
Commute Daily Timeline



CCAR PROJECT 4

Health Measurements

Health Measure	Assessment
Questionnaire	symptoms, mood, exposure perception
Urine	pregnancy, cotinine, 1-Nitropyrene
Vitals	blood pressure, heart rate
24-hr ECG	heart rate variability, SSDN, QT interval
Finometry	caridac output, stroke volume, peripheral resistance
Blood markers	neuropeptides, homocysteine, cytokines, inflammatory markers, CBC, HLD, LDL, gene expression, angiotensin-2, DNA methylation, catecholamines, CRP
Brachial artery ultrasound	diameter, FMD
Retinal photography	arteriolar and venular caliber
SNP (genotype stratification)	TRPV1
Food frequency questionnaire	antioxidants in diet

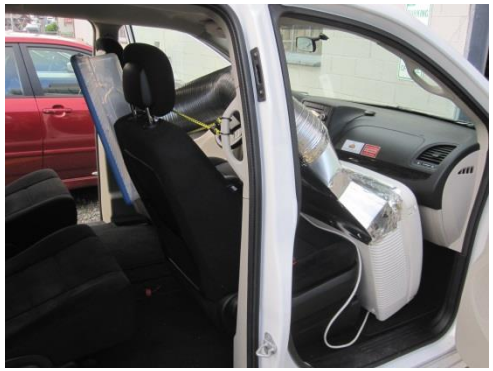


CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

CCAR PROJECT 4 In-Vehicle Monitoring

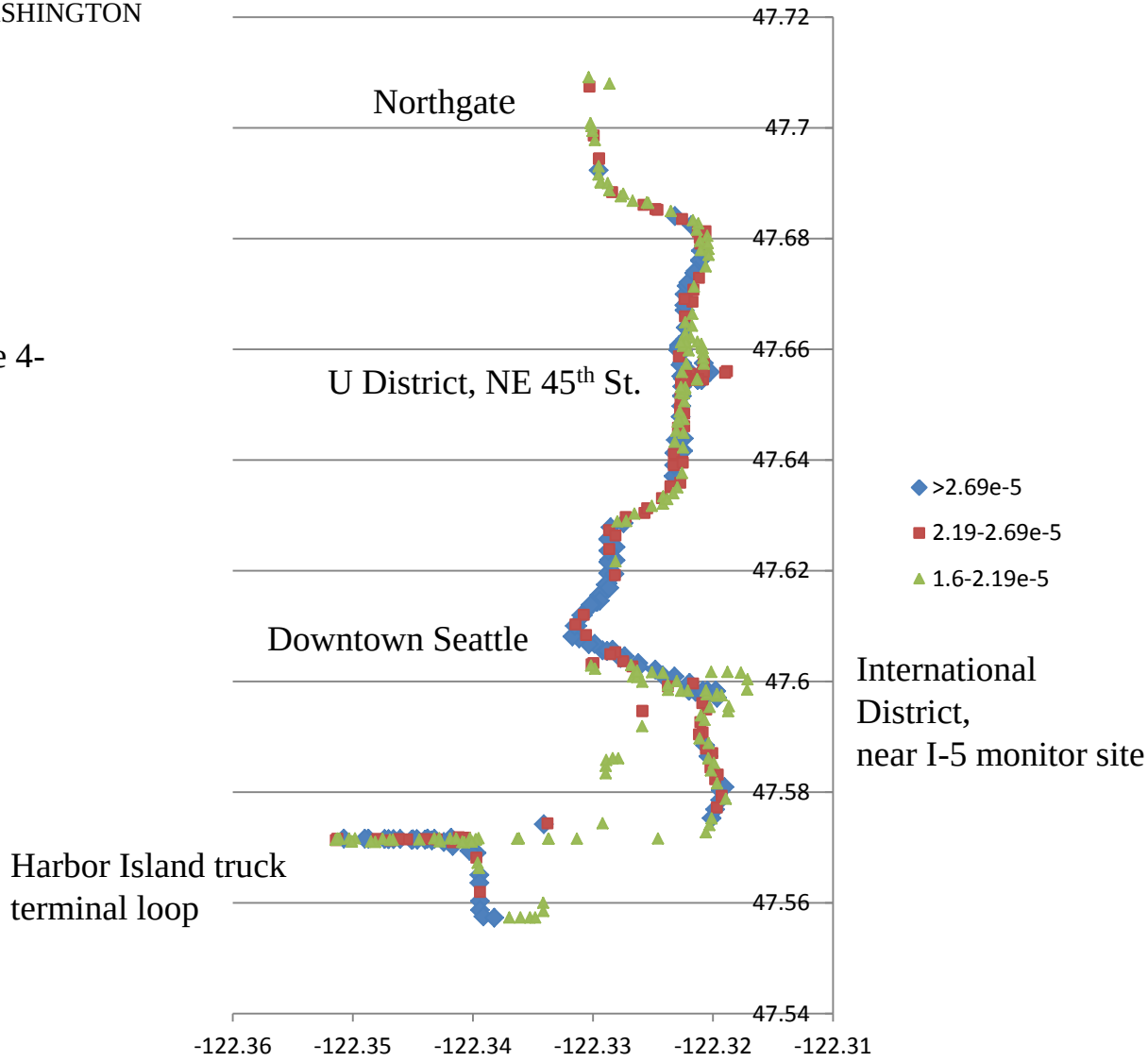
Parameter Assessed	Analyzer	Manufacturer	Measurement range
PM2.5 Light Scatter	Nephelometer	Radiance Research	0 - $1x10^{-3}$ m^{-1}
Ultrafine (0.02-1 um) Particle count	P-Trak 8525	TSI	0 to $5x10^{+5}$ ptcl/cm^3
Black Carbon	micro-Aeth AE51	AethLabs	0-1 mg BC/m^3
PAH, particle-bound	PAS 2000CE	EcoChem	0-1000 ng/m^3
NO2	CAPS	Aerodyne	0 - 2000 ppb
NO	NO Model 410	2B Technologies	0 - 2000 ppb
NOx	Model 410 + converter	2B Technologies	0 - 2000 ppb
Ozone	O3 Analyzer 3.02 P-A	Optec	0 to 255 ppb
CO	T15n Monitor	Langan, Inc.	0 - 200ppm
VOCs	Summa canister, GC-MS		(varies by analyte)
CO2	SenseAir K-30-FS sensor	CO2Meter.com	0 – 5,000 ppm (vol.)
Temperature & Rel. Humidity	Precon HS-2000 sensor	Kele Precision Mfg.	0° - 70°C; 0 - 100%
Location, lat & long	GPS, BU-353	US GlobalSat	5 m accuracy
Traffic source on route	Video LifeCam VX-5000	Microsoft	



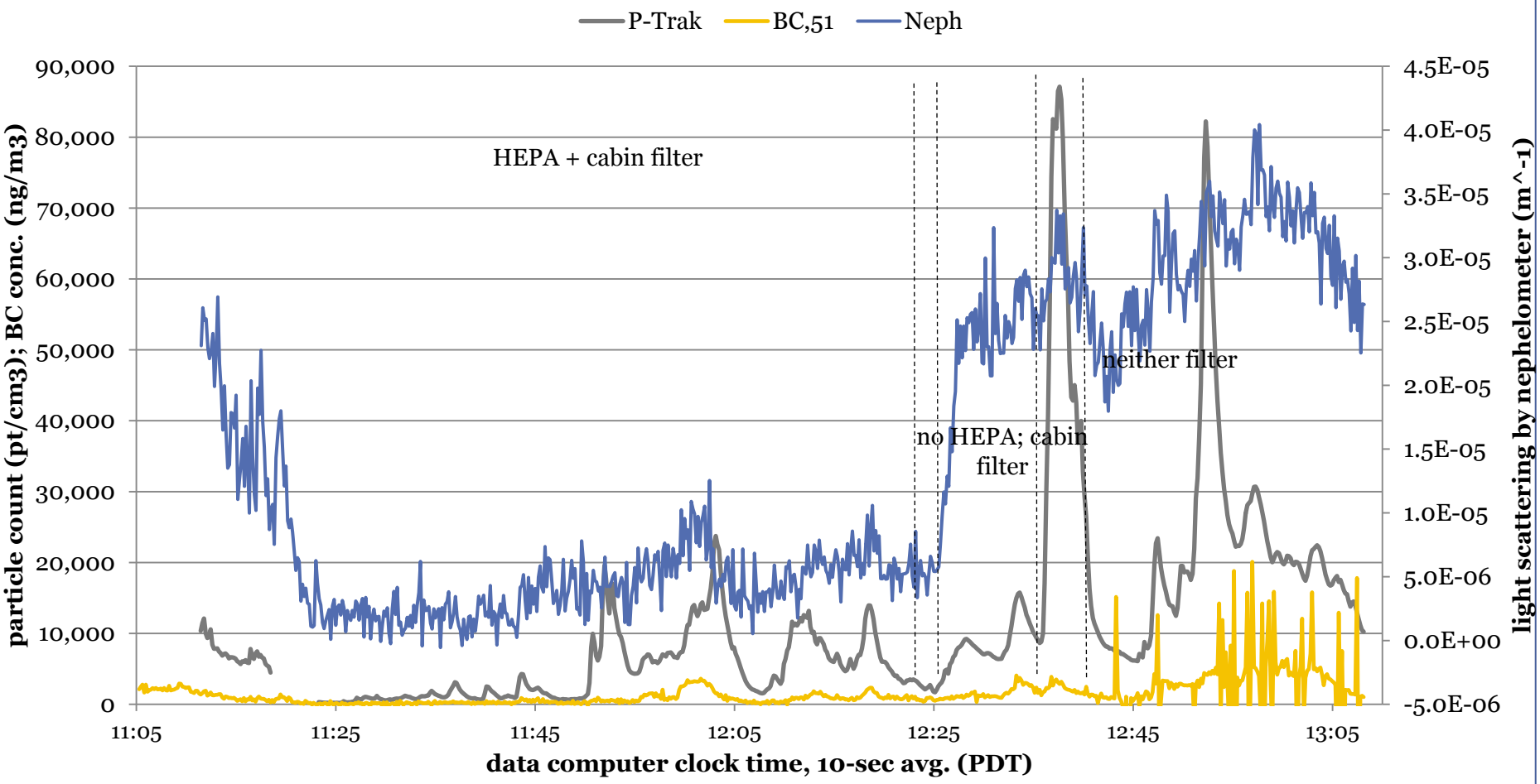
CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

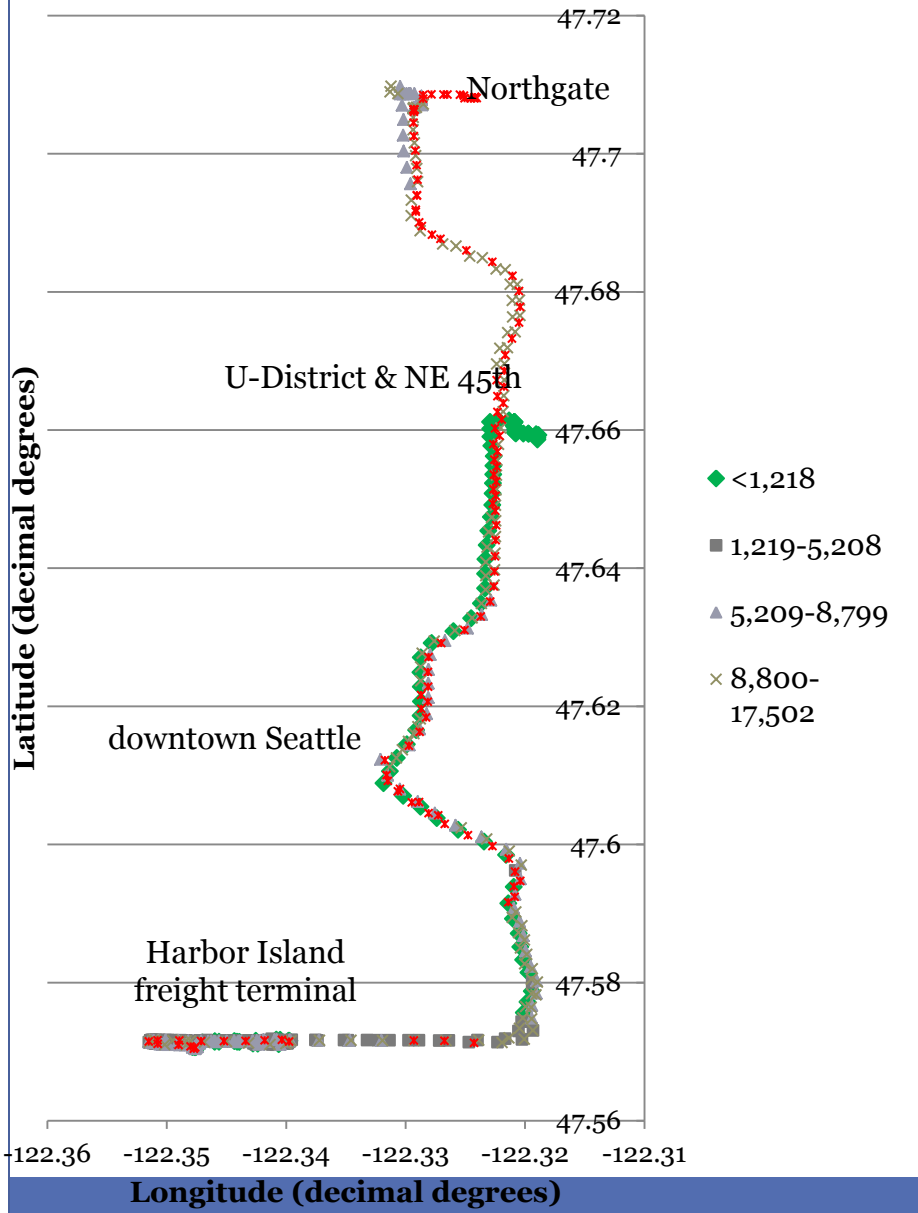
Three Highest quintiles of ambient outdoor PM by Nephelometer plotted geographically, Seattle drive 4-Mar-14



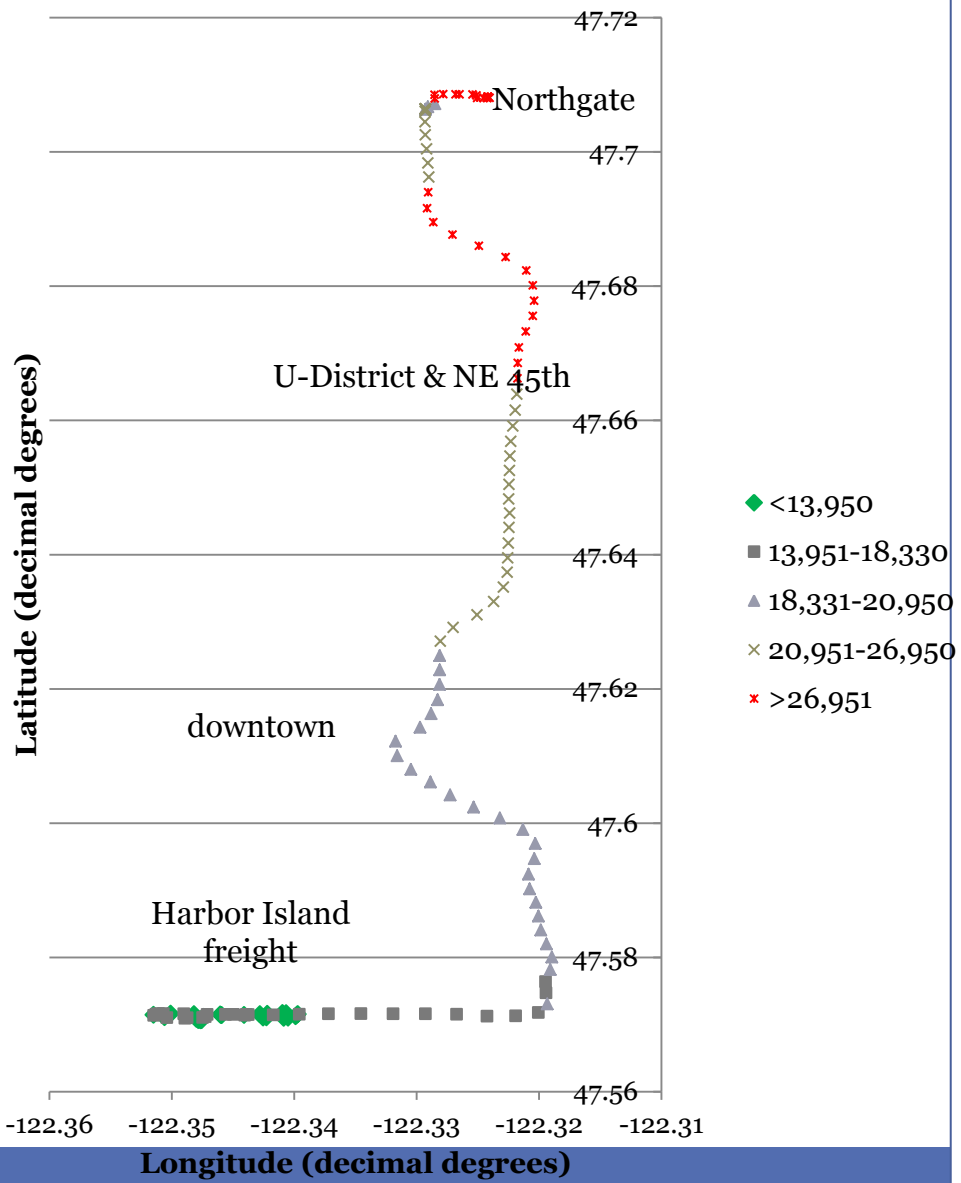
Particle Count, PM2.5, BC in vehicle, I-5 & Harbor I., 1-Oct-14 drive



P-Trak particles (pt/cm³) by quintile for all filter conditions, I-5 & Harbor I. drive



P-Trak particles (pt/cm³) by quintile for no HEPA & no cabin filter, I-5 & Harbor I. drive

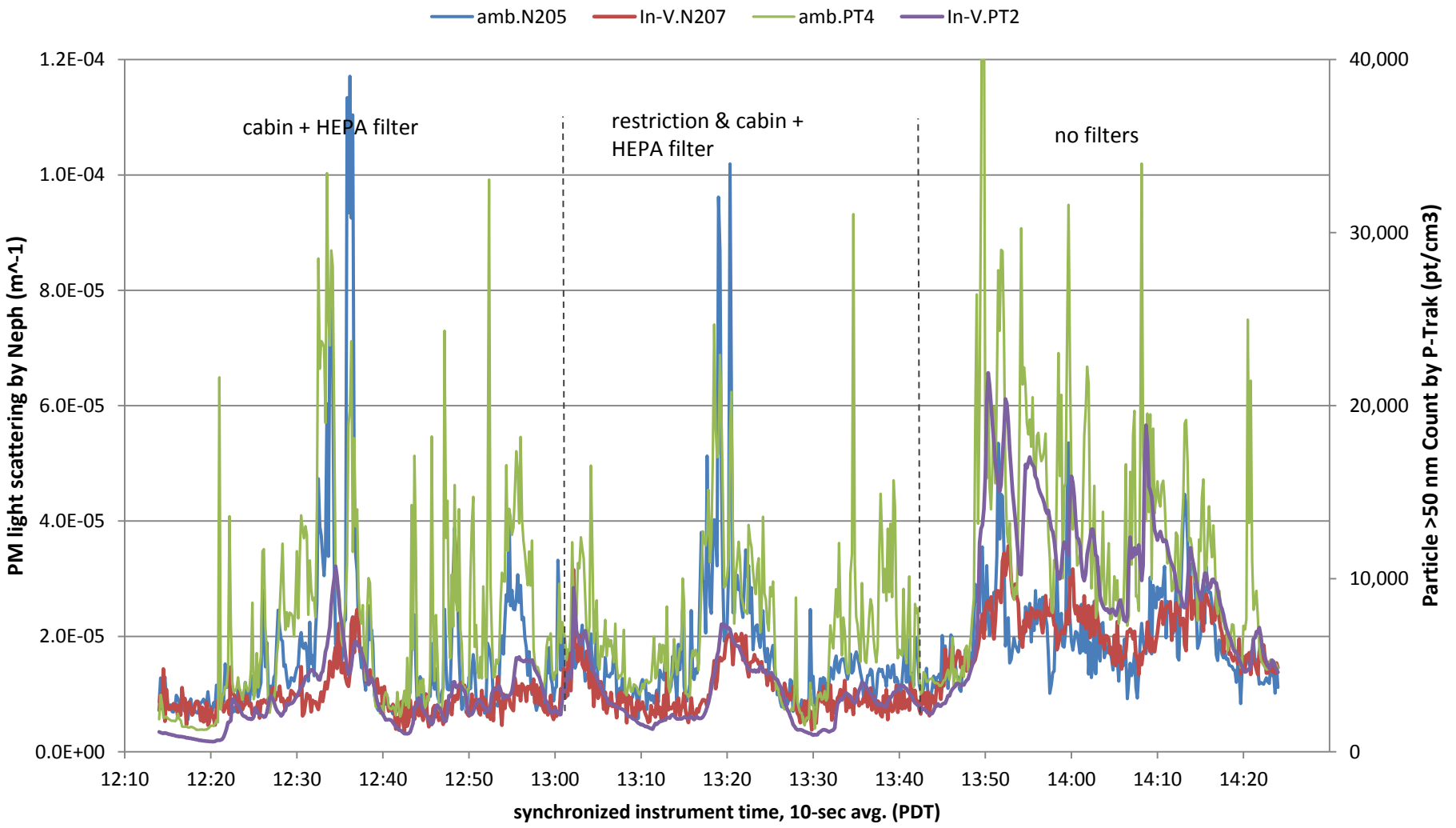


CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

CCAR
Center for Clean Air Research

PM proxy conc. & Count in 3 filter conditions, HEPA & cabin filter test



CENTER FOR CLEAN AIR RESEARCH

UNIVERSITY *of* WASHINGTON

Dual Filtration effectiveness summary– particle reduction

Particle Count, P-Trak (pt/cm3)		fan speed		Reduction at medium fan	
condition	parameter	low	medium	vs. low	vs. medium
HEPA & pre-filter	90th %-ile	285	124		
	mean	223	117	96.5%	95.9%
	10th %-ile	143	109		
no filters in unit	90th %-ile	4,030	2,967		
	mean	3,382	2,842		
	10th %-ile	2,970	2,686		

PM light scatter, Nephelometer		fan speed		Reduction at medium fan	
condition	parameter	low	medium	vs. low	vs. medium
HEPA & pre-filter	90th %-ile	4.98E-06	4.34E-06		
	mean	3.38E-06	2.64E-06	72.4%	73.2%
	10th %-ile	1.9E-06	1.13E-06		
no filters in unit	90th %-ile	1.16E-05	1.18E-05		
	mean	9.59E-06	9.87E-06		
	10th %-ile	7.69E-06	8.12E-06		

Black Carbon, AE51 (ng/m3)		fan speed		Reduction at medium fan	
condition	parameter	low	medium	vs. low	vs. medium
HEPA & pre-filter	90th %-ile	196.8	228.0		
	mean	66.7	74.8	79.9%	71.4%
	10th %-ile	-59.5	-81.3		
no filters in unit	90th %-ile	515.0	414.9		
	mean	372.5	261.3		
	10th %-ile	222.2	100.4		

Long-term outdoor air pollution and DNA methylation in circulating monocytes: MESA Air

Gloria Chi
Center for Clean Air Research
Scientific Advisory Committee Meeting
October 6, 2014

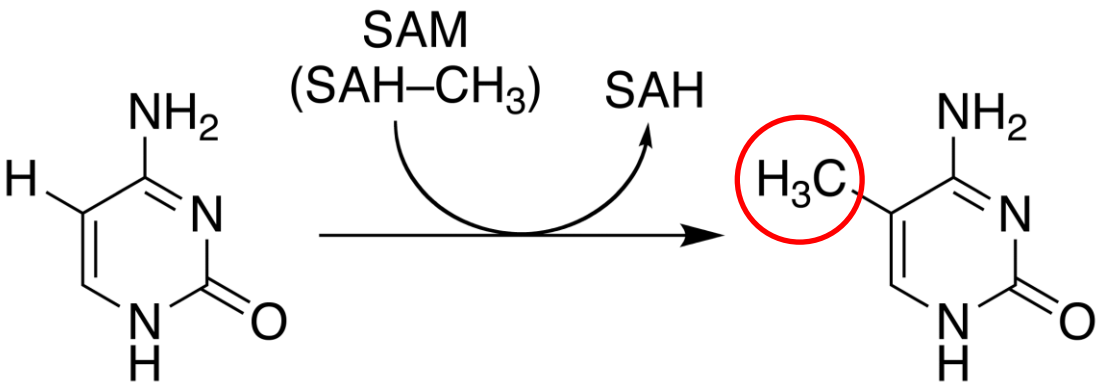
Air Pollution

- Air pollution (AP) is a risk factor for cardiovascular disease
- Biological mechanisms not fully understood

- AP alters DNA methylation (DNAm) in peripheral blood leukocytes

DNA Methylation

- Cytosine guanine dinucleotide (CpG) sites
- Does not change DNA sequence
- Heritable – cell division
- Associated with gene expression
- Cell-specific



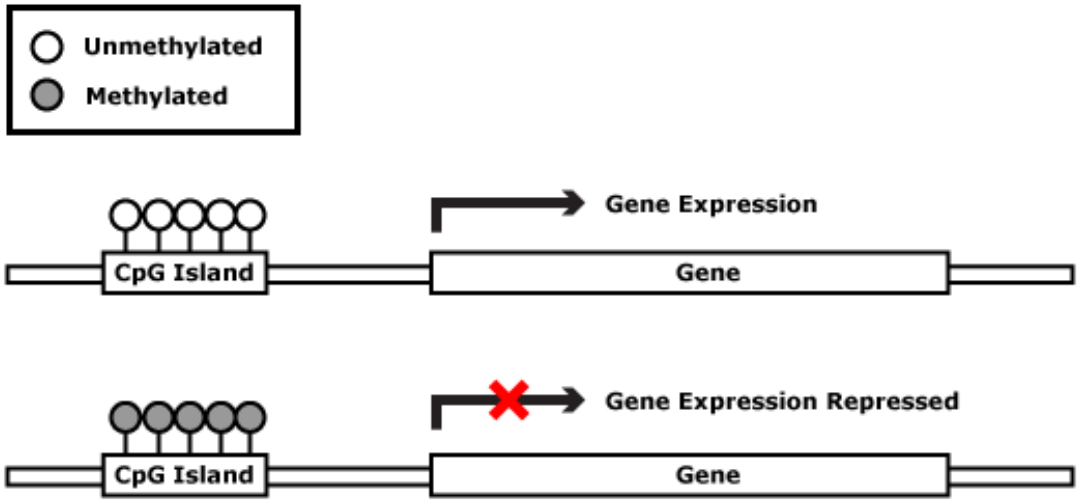
Cytosine

5-Methylcytosine

Source: By DMacks (Own work) [Public domain], via Wikimedia Commons

DNA Methylation

- Cytosine guanine dinucleotide (CpG) sites
- Does not change DNA sequence
- Heritable – cell division
- Associated with gene expression
- Cell-specific



Source: UCSF School of Medicine

Monocytes

- Leukocyte with important role in immune system
- Atherosclerosis
  - Promote chronic inflammation
  - Adhere to vascular endothelium and differentiate into macrophages that accumulate in plaques

Aims

- Test association of ambient AP and DNAm in monocytes
  - Replicate prior studies of global DNAm (Alu and LINE-1) using probe-specific approach
  - Expression-associated methylation sites (eMS)
    - Previous work identified 11,203 eMS in this cohort*
    - Analyze 2,713 top unique eMS in this analysis

*Liu Y et al. Hum Mol Genet. 2013 Dec 15;22(24):5065-74.

Methods

Study Population

- Subset MESA participants (n=1,207)
- AP, methylomics, transcriptomics data

Air Pollution Assessment

- Fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO_X)
- Predicted at participant residences
- Averaged over one year prior to blood draw
- Spatio-temporal models
  - Likelihood-based estimates
  - EPA and cohort-specific monitoring
  - Geographic data such as roadway density and land use

Methods

Tissue Collection

- Blood from 5th exam (4/10-2/12)
- CD14+ monocytes separated on-site
- DNA and RNA extracted

Tissue Analysis

- Illumina Infinium HumanMethylation450 Beadchip
- Illumina HumanHT-12 v4 Expression BeadChip

Methods

Global DNAm

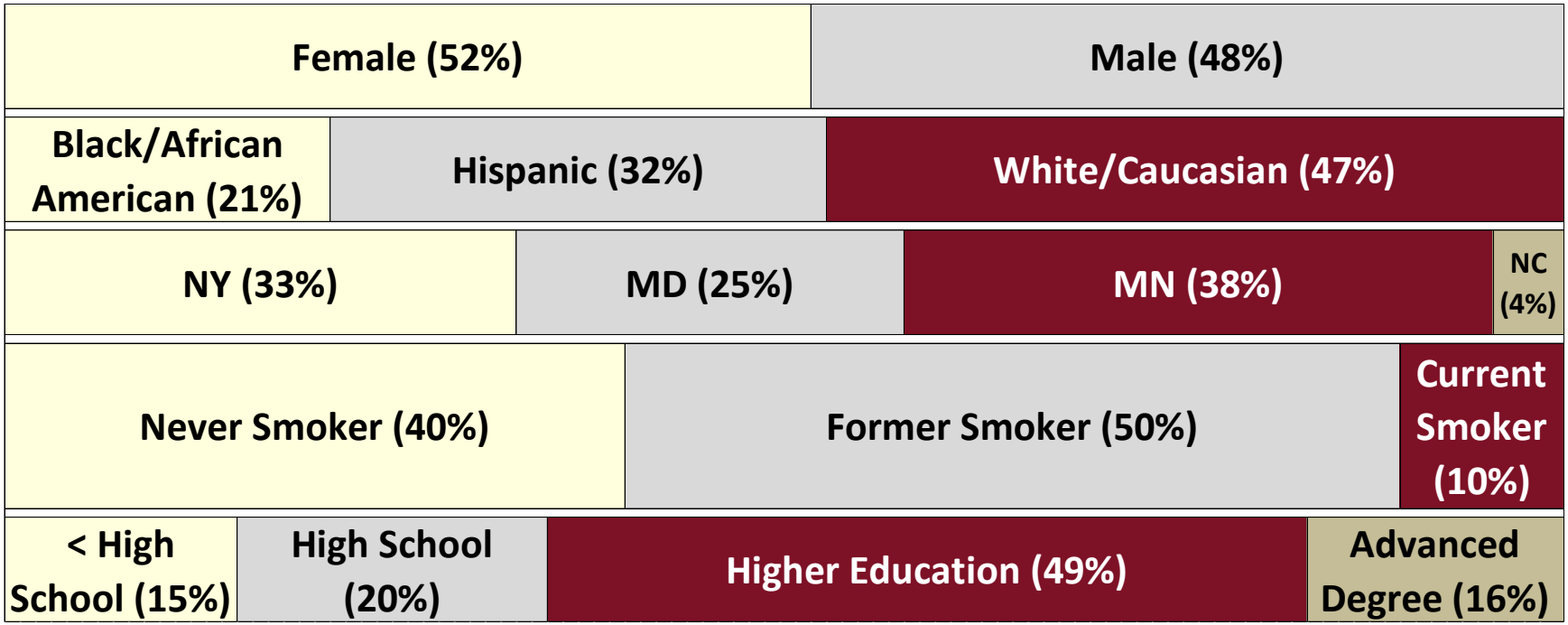
- Exposure: $PM_{2.5}$, NO_X
- Outcome: median methylation of Alu and LINE-1 probes
- Linear regression
- Covariates: demographics, site, smoking, socioeconomic status, body mass index, recent infection, residual cell contamination, chip and row

eMS

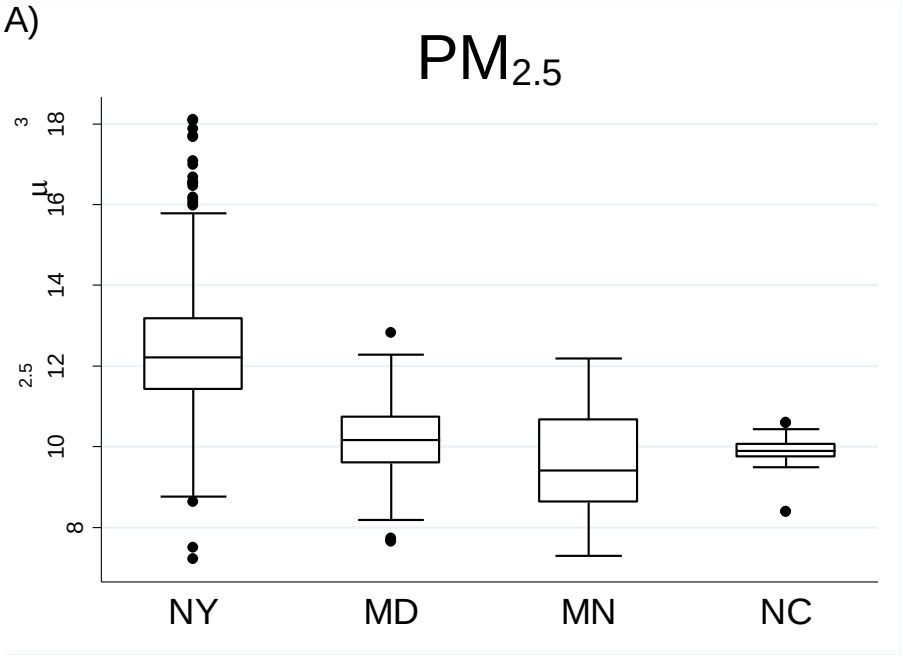
- Exposure: $PM_{2.5}$, NO_X
- Outcome: eMS methylation (n=2,713)
- Linear regression
- False Discovery Rate 0.001
- Same adjustment covariates as above

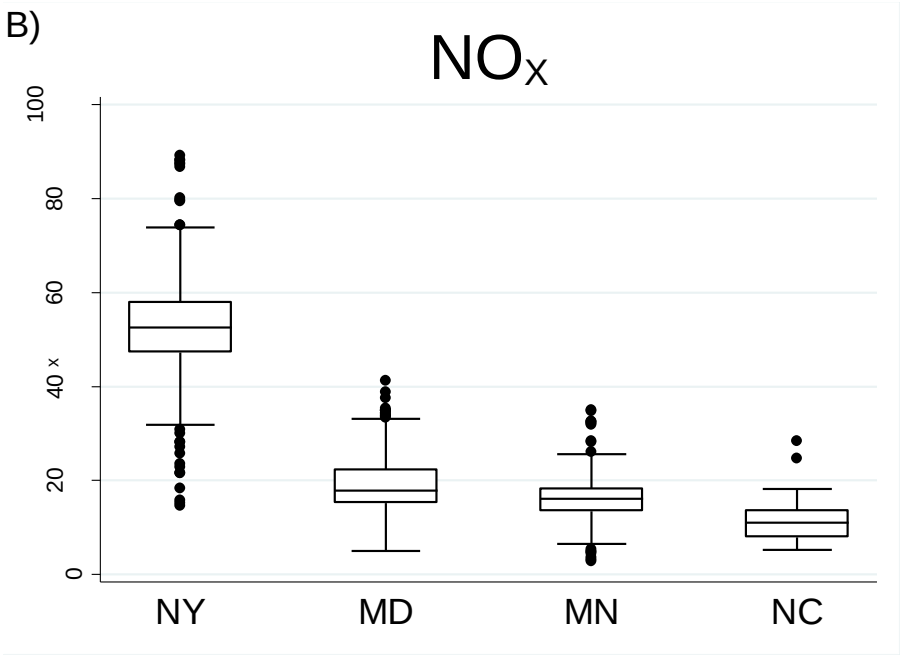
Study Sample (n=1,207)

Age: 69.6 ± 9.4 (mean ± SD)



1-yr Average $PM_{2.5}$ and NO_X Predictions



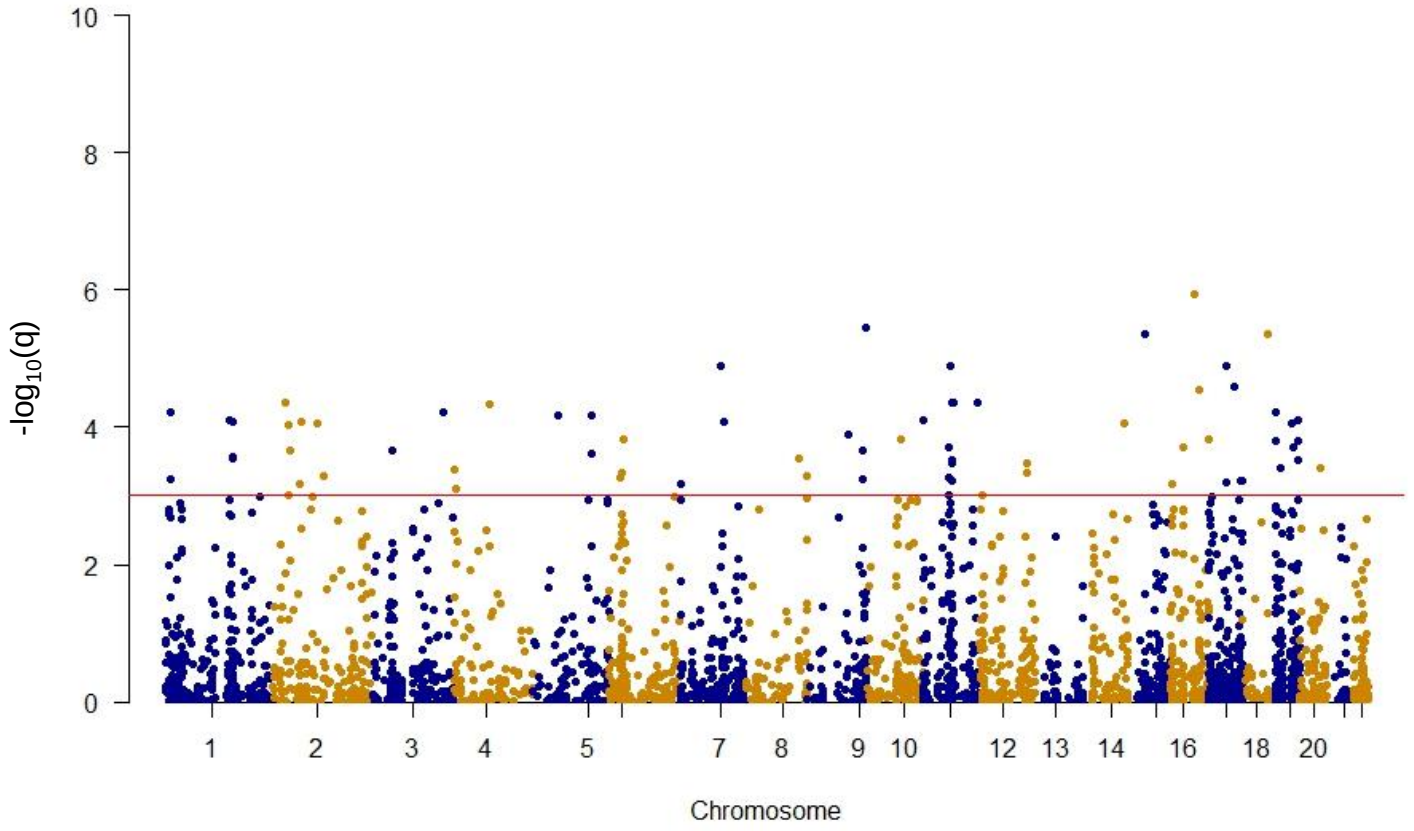


1-yr Average AP and Global DNAm

	$PM_{2.5}$ (per 5 µg/m^3)			NO_X (per 40 ppb)		
	Coef.	**(95% CI)**	**P**	**Coef.**	**(95% CI)**	**P**
Alu						
Model 1	-0.015	(-0.030, -0.0002)	0.047	-0.004	(-0.024, 0.016)	0.695
Model 2	-0.007	(-0.014, 0.001)	0.085	-0.002	(-0.012, 0.008)	0.720
LINE-1						
Model 1	-0.025	(-0.044, -0.006)	0.010	-0.019	(-0.045, 0.006)	0.132
Model 2	-0.003	(-0.012, 0.006)	0.472	0.004	(-0.008, 0.016)	0.532

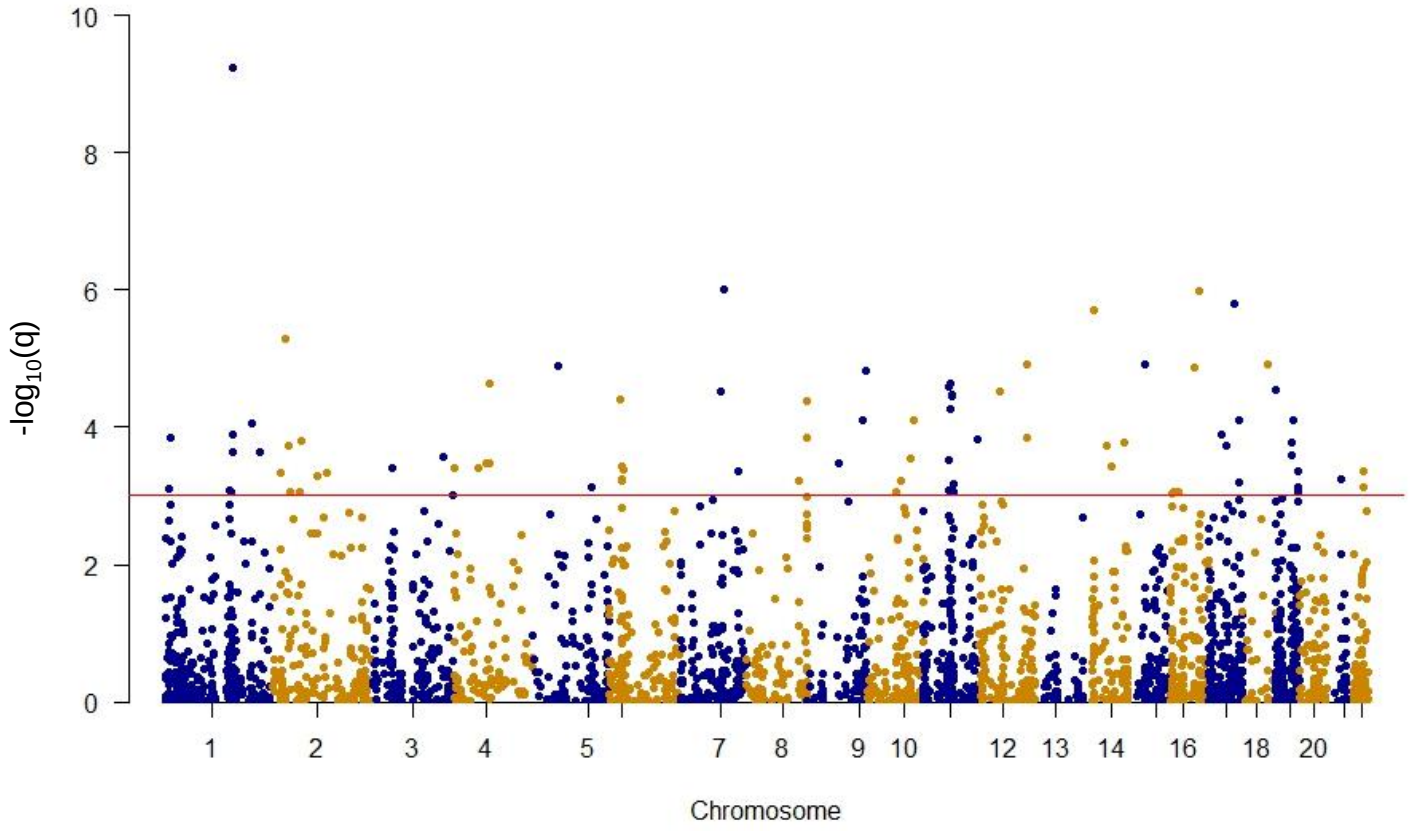
Model 1: age, sex, race/ethnicity, site, smoking, socioeconomic status, body mass index, recent infection, residual cell contamination
Model 2: + chip, position

$PM_{2.5}$ associated with 72 eMS at FDR 0.001*



*Adjusted for age, sex, race/ethnicity, site, smoking, socioeconomic status, body mass index, recent infection, residual cell contamination, chip, position

NO_X associated with 85 eMS at FDR 0.001*



*Adjusted for age, sex, race/ethnicity, site, smoking, socioeconomic status, body mass index, recent infection, residual cell contamination, chip, position

Top 5 $PM_{2.5}$ and NO_X associated eMS

Gene	CpG Site	Chr	Beta	95% CI	P-value
$PM_{2.5}$					
C16ORF57	cg12125117	16	0.18	(0.12, 0.23)	4.29E-10
NOTCH1	cg21252105	9	0.24	(0.16, 0.32)	2.66E-09
PLCB2	cg05059480	15	0.13	(0.08, 0.17)	5.08E-09
BCL2	cg13381110	18	0.18	(0.12, 0.24)	6.52E-09
TRRAP	cg01877450	7	0.15	(0.10, 0.20)	2.91E-08
NO_X					
TAGLN2	cg04922029	1	0.26	(0.19, 0.33)	2.15E-13
Visfatin	cg18989536	7	-0.12	(-0.16, -0.08)	7.04E-10
ATP6V0D1	cg09101151	16	0.17	(0.11, 0.22)	1.12E-09
DDX5	cg21519701	17	0.13	(0.08, 0.17)	2.29E-09
DHRS4L2	cg01878807	14	0.19	(0.13, 0.25)	3.71E-09

Visfatin/Nampt

What is it?

- Proinflammatory cytokine
  - Initially identified as novel adipokine enriched in visceral fat
- Secreted by monocytes, macrophages, adipocytes, endothelial cells

What we know

- Associated with obesity and insulin resistance
- Upregulated in macrophages from human unstable atherosclerotic lesions
- Higher expression in patients with type II diabetes
  - Vascular damage, endothelial dysfunction

What we found

- $PM_{2.5}$ and NO_X negatively associated with DNAm
- DNAm negatively associated with gene expression
- $PM_{2.5}$ and NO_X exposure may be positively associated with visfatin expression

Conclusions

- Weak evidence of association between $PM_{2.5}$ and global DNAm in monocytes
- No evidence of association between NO_X and global DNAm in monocytes
- 72 and 85 eMS associated with $PM_{2.5}$ and NO_X, respectively
- Large number of significant hits in interesting pathways (e.g. Visfatin)

Strengths and Limitations

Strengths

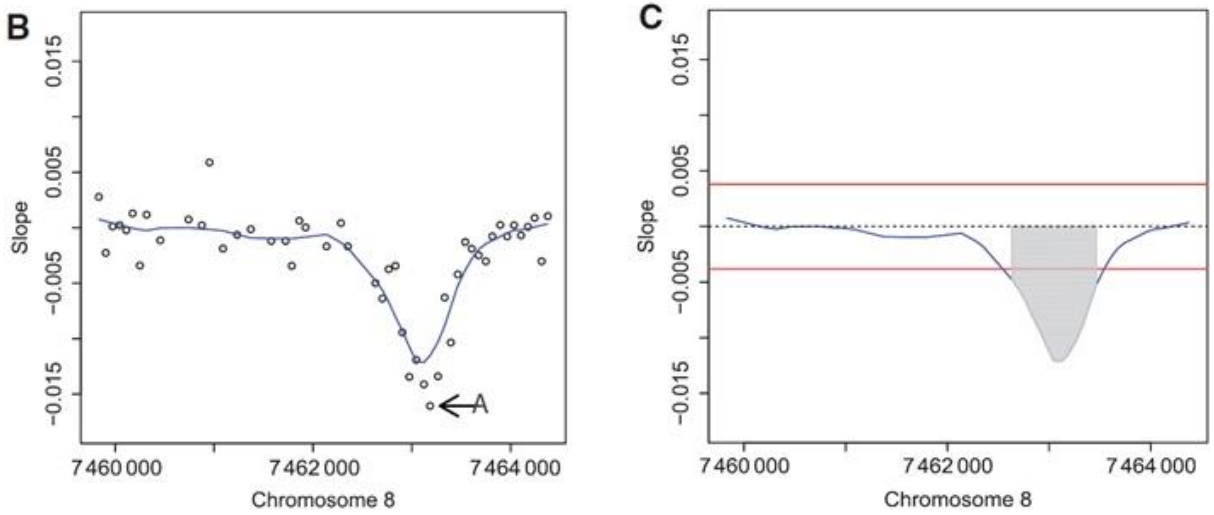
- Sophisticated air pollution assessment
- Purified monocytes
- Methylomic and transcriptomic data

Limitations

- Data from peripheral blood monocytes only allows assessment of systemic differences

Future Directions

- Multi-pollutant exposure model
- Bump hunting algorithm
  - Identify differentially methylated regions
  - Correlated methylation of CpG sites within a region
  - Methylation values of single CpG site are noisy



Jaffe A et al. *Int. J. Epidemiol. (2012) 41 (1): 200-209.*

Future Directions

- Multi-pollutant exposure model
- Bump hunting algorithm
- Pathway analysis
- Test association of air pollution-related methylation signals with subclinical atherosclerosis
- Replication in another cohort

Acknowledgements

- MESA Participants
- MESA & MESA Air Investigators and Staff (Joel Kaufman, Yongmei Liu, R. Graham Barr, Kathleen Donohue, Lifang Hou, Charles McCall, David Siscovick, Amanda Gassett, Cynnie Curl, Casey Olives, Marnie Hazlehurst, Elizabeth Spalt)
- James MacDonald

Funding:

- UW CCAR – Environmental Protection Agency (RD83479601)
- MESA Air – Environmental Protection Agency (RD831697)
- MESA – National Heart, Lung, and Blood Institute (NHLBI; N01 HC-95159 through N01-HC-95169 and RR-024156)
- MESA Epigenomics and Transcriptomics Study – NHLBI (R01HL101250)
- Center for Ecogenetics & Environmental Health – National Institute of Environmental Health Sciences (NIEHS; P30 ES007033)

Project 3 Progress Report

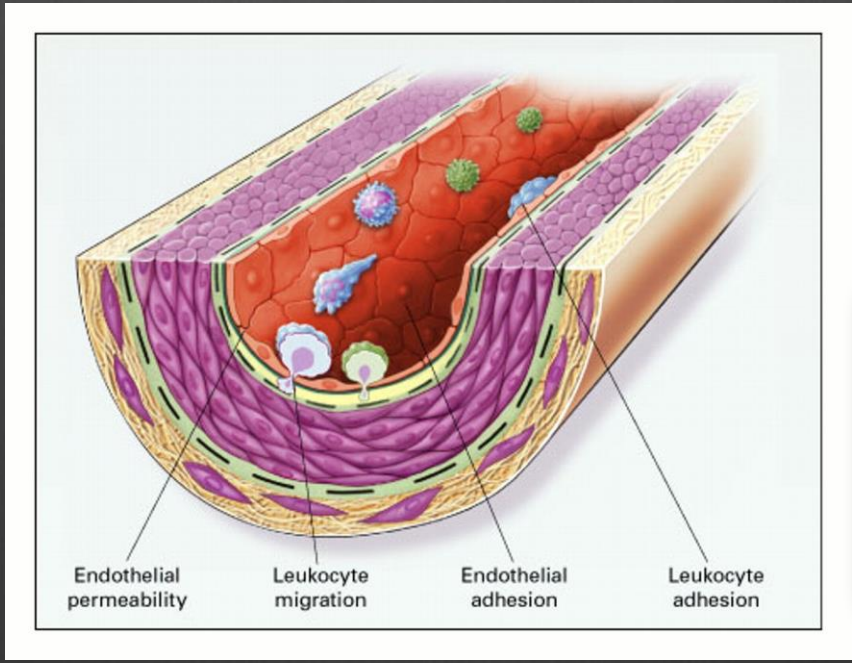
Matthew Campen, Ph.D., M.S.P.H.

UNM | COLLEGE of PHARMACY

Endothelial Activation is a Key Event in the Initiation and Promotion of Vascular Disease

Injury

EC Activation
- ↑ permeability
- ↑adhesion of monocytes
- Relocalization and uncoupling of eNOS



Vessel Wall Accumulates:
- LDL Cholesterol
- Macrophages

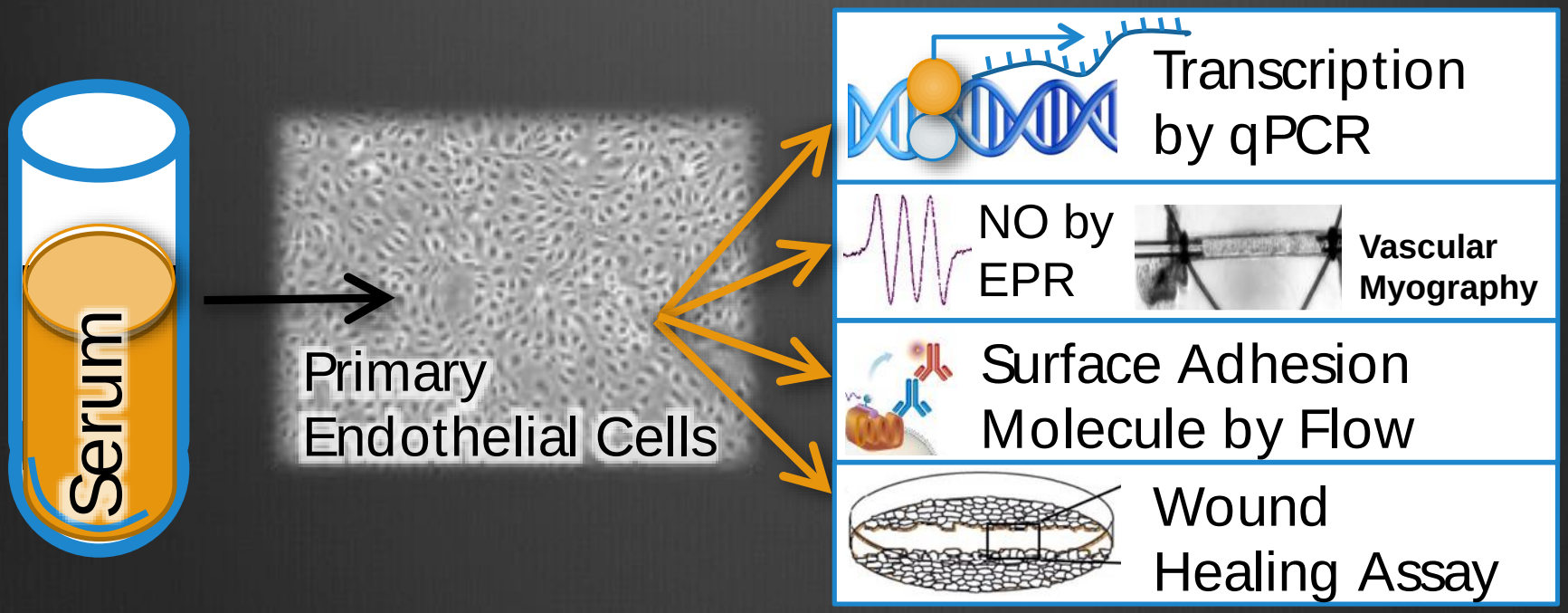
Pro-Inflammatory Transformation:
- Oxidized LDL, lipids
- Foam Cells
- SMC Differentiation

Lesion Growth, Destabilization
- Vessel Occlusion
- Aggregation
- Plaque Rupture

Premise: Serum is a Major Part of the Endothelial Exposome

- Endothelial cells are the target organ for all circulating material
- Serum contains 1000's of factors
  - Protein, lipid, metabolites/small molecules, microvessicles and platelets
- Physical factors – not addressed with current assays
  - Pressure (cyclical)
  - Shear stress and flow
  - Cellular components

General Experimental Theme

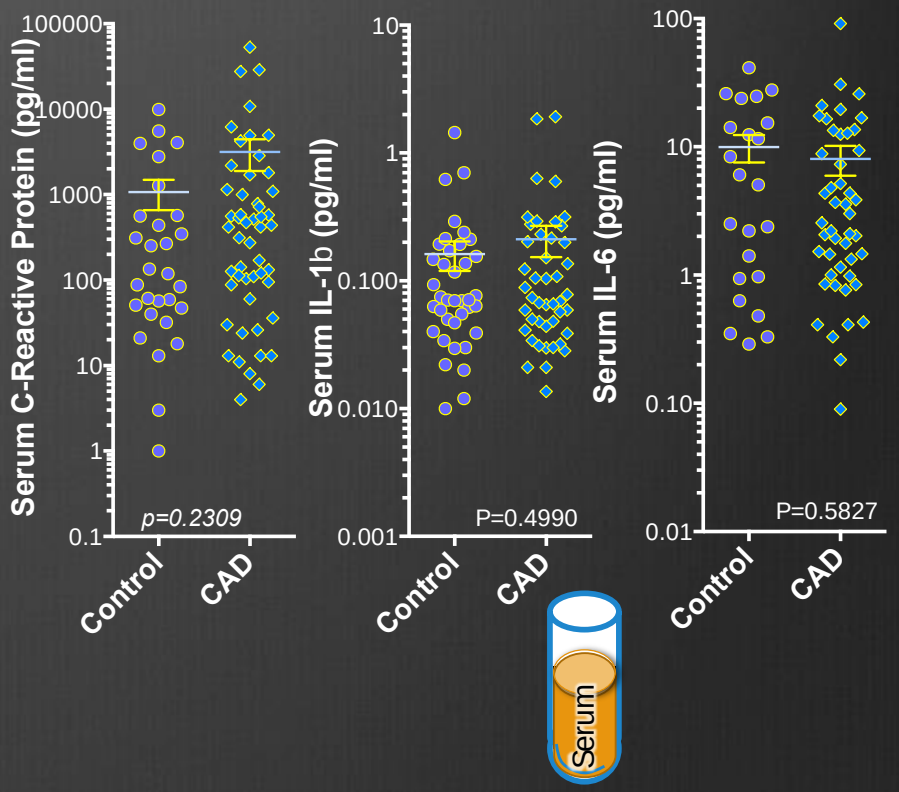


SAC Comments Suggested Further Characterization of Serum Bioactivity Would be of Value

- Update will show progress on:
  - Clinical interpretation
  - Microarray outcomes (ie full transcriptional response)
  - Expanded biological responses
  - Various pollutant effects
  - Metabolomic/proteomic studies of serum post-exposure

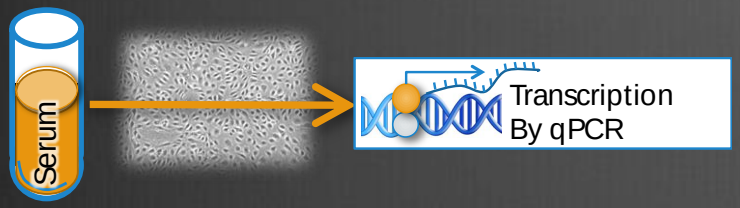
Serum Cytokines Between Coronary Artery Disease (CAD) Patients and Controls: No differences

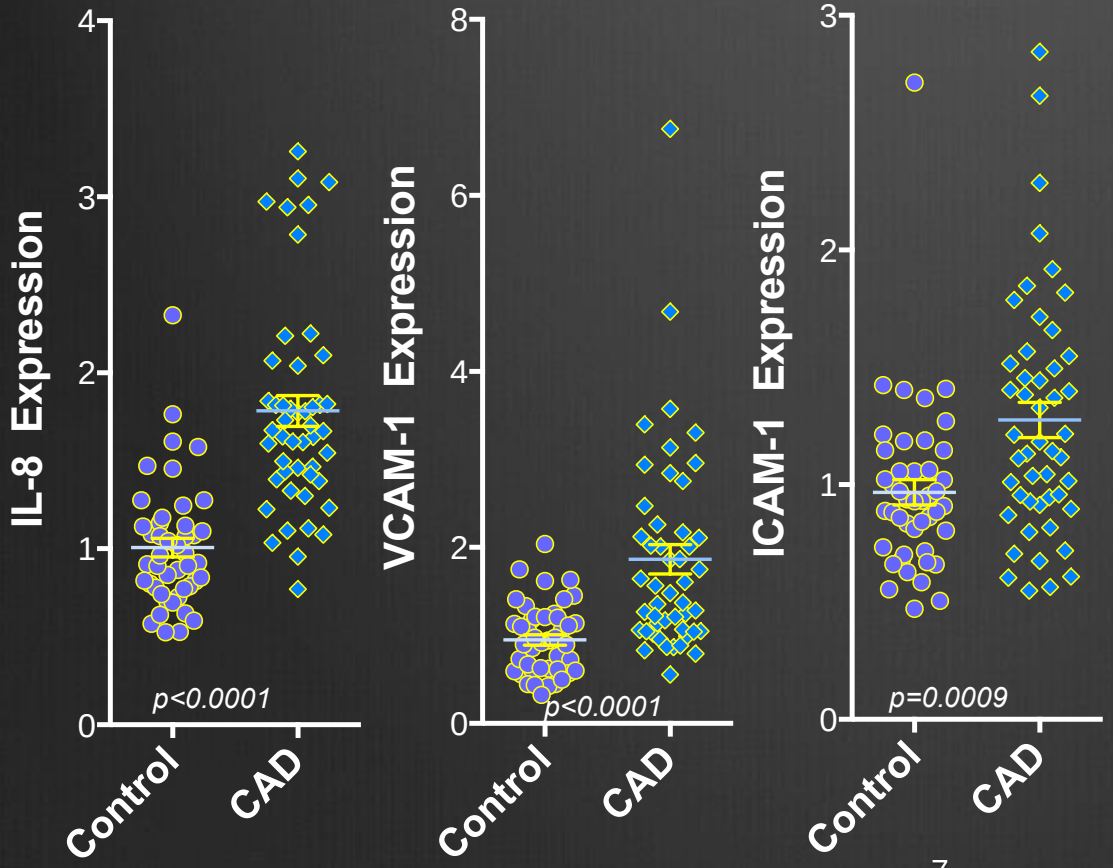
- 48 subjects post-myocardial infarction were referred to a pharmacy clinic where bloods were drawn
- Cardiac event had occurred 1-24 months prior to follow-up serum sample being obtained
- All subjects in stable health, on >7 meds (aspirin, statins, other)
- 45 control samples from a healthy population
  - Much younger
  - Much lower BMI
- Reports note that inflammatory risk factors lose predictive value once subjects have started therapeutics (statins, antiplatelets, etc; Sever et al., JACC, 2013)



Cung et al., submitted, 2014

Comparing IPA in Patients with Diagnosed Coronary Artery Disease with Healthy Controls



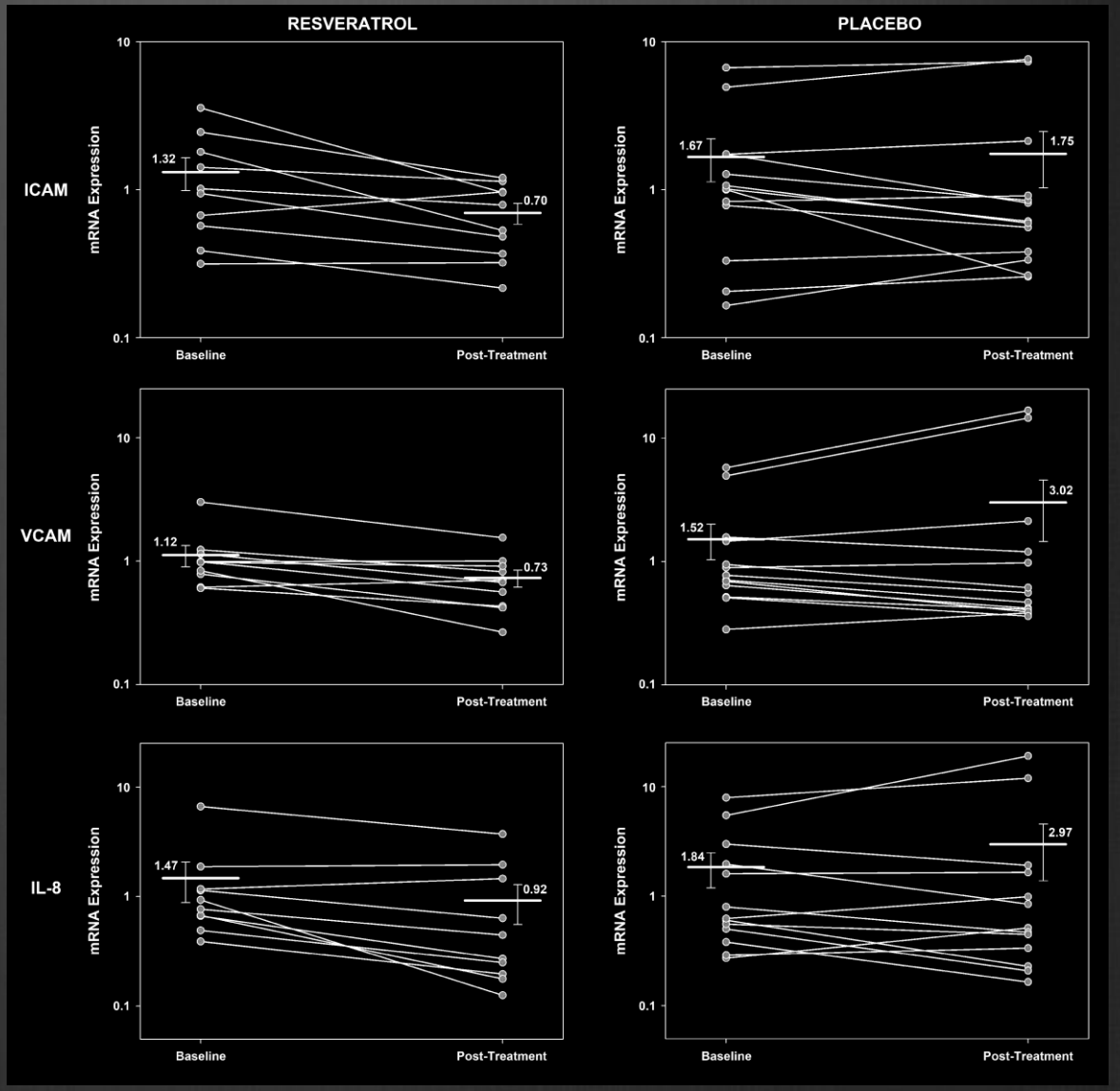


- Endothelial cells incubated with subjects' serum responded in an inflammatory manner
- Endothelial mRNA increases were significantly different between healthy and CAD subjects even after adjusting for age, gender, and BMI

Cung et al., submitted, 2014

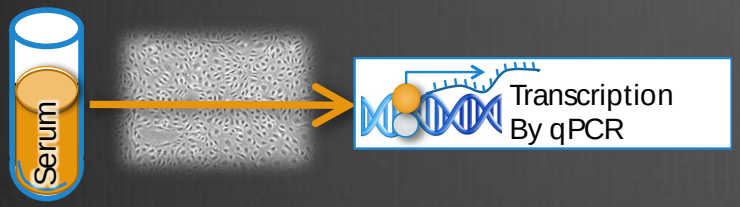
Drug Efficacy Using Endothelial Cells as “Biosensors” - Resveratrol

- Healthy subjects were randomly assigned to placebo or resveratrol treatment for a month
- Blood (serum) was collected before and after.
- IPA was conducted on these serum samples
- Significant reduction in all IPA parameters was observed for resveratrol but not placebo.
- No changes in other metrics, including cytokines, glucose, etc.



Agarwal et al., Intl J Cardiol, 2013

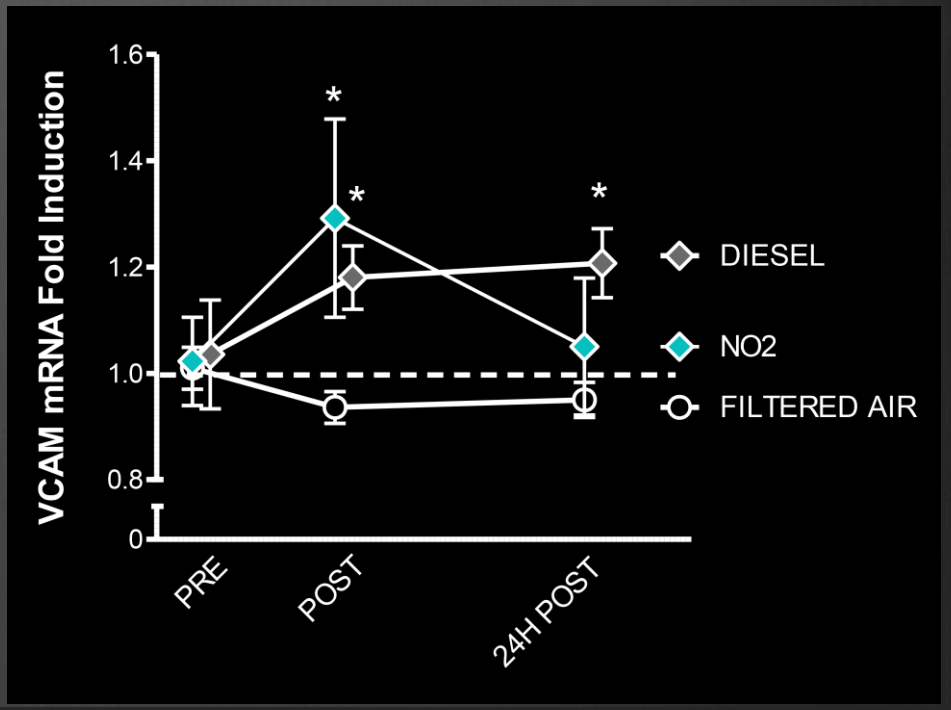
Diesel and NO_2 Induce Serum Factors in Humans that Cause Endothelial Cell Activation



- Treated primary human coronary endothelial cells for 24h with plasma (**10% in media**) from exposed subjects
- Diesel (100 µg/m^3) x 2 h
- NO_2 (500 ppb) x 2 h

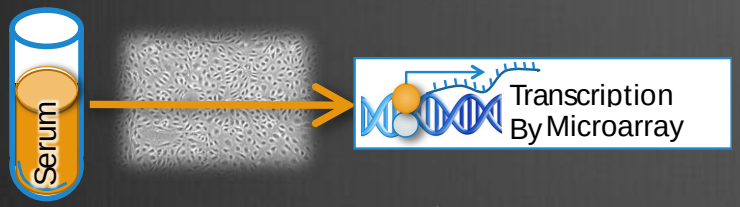
- VCAM, ICAM, p-selectin, IL-8 elevated by post-exposure plasma

Channell et al., Toxicol Sci, 2012

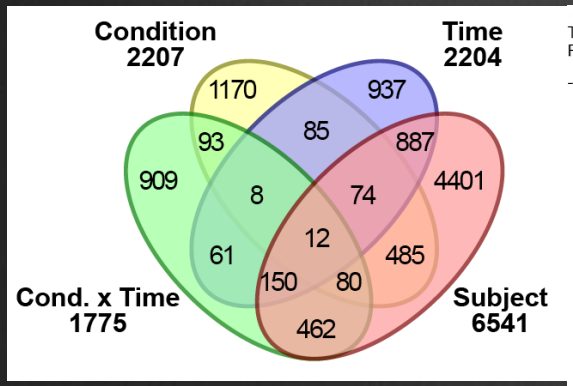


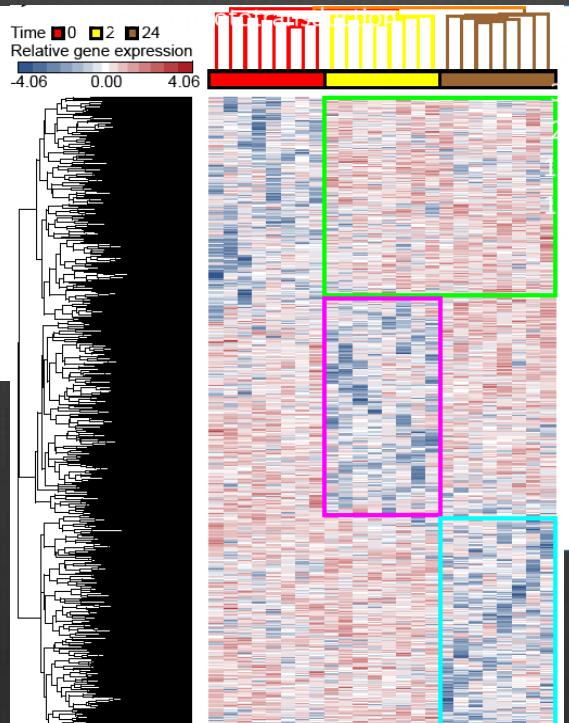
	Diesel Subjects:	NO_2 Subjects
Age	24.9 ± 4.5	25.3 ± 5.5
Gender (M/F)	(3/4)	(5/2)
Wgt (kg)	67.2 ± 9.4	77.3 ± 13.2
BMI (kg/m^2)	23.9 ± 1.9	24.9 ± 2.7
BSA (m^2)	1.8 ± 0.2	1.9 ± 0.2

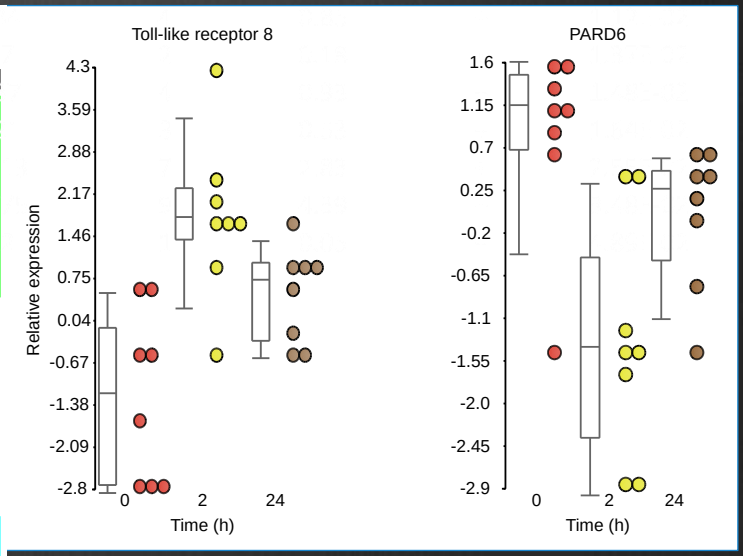
NEW: Genomic Assessment of Human Endothelial Cell Activation by Diesel-Induced Serum Factors



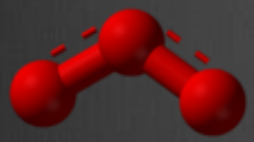
Pathways	#	#	expected	+/-	P value
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	134	10	3.36	+	2.41E-03
PDGF signaling pathway	132	9	3.31	+	6.88E-03
Inflammation mediated by chemokine and cytokine signaling pathway	233	13	5.85	+	6.90E-03
FAS signaling pathway	31	4	0.78	+	8.19E-03
Wnt signaling pathway	297	15	7.45	+	9.24E-03
2-arachidonoylglycerol biosynthesis	6	2	0.15	+	1.02E-02



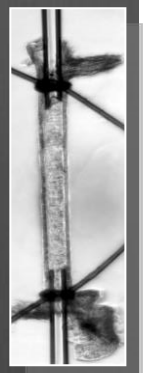


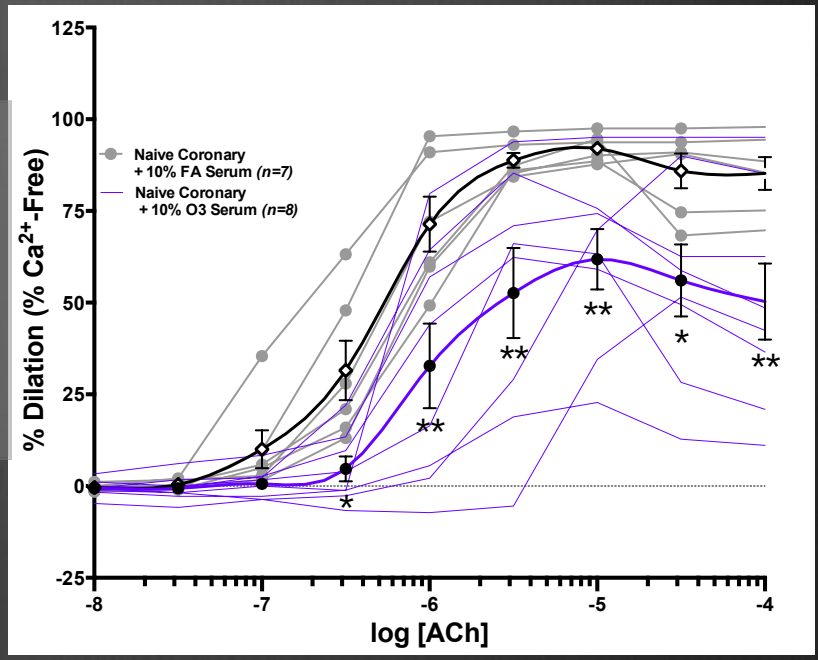


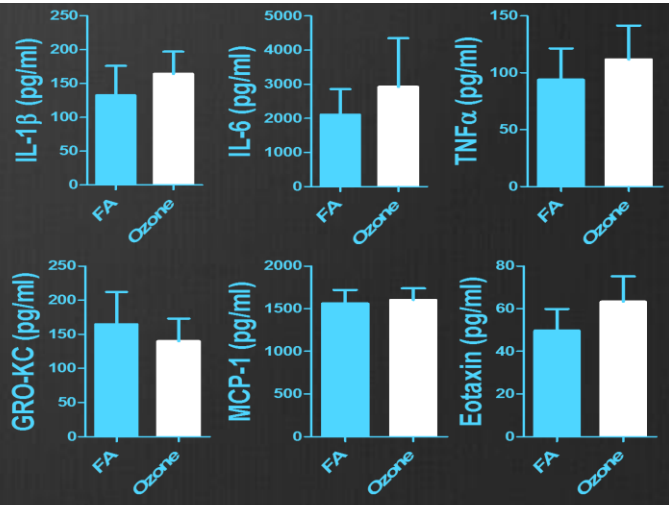
Serum from Ozone-Exposed Rats Impairs Vasodilation *Ex Vivo*



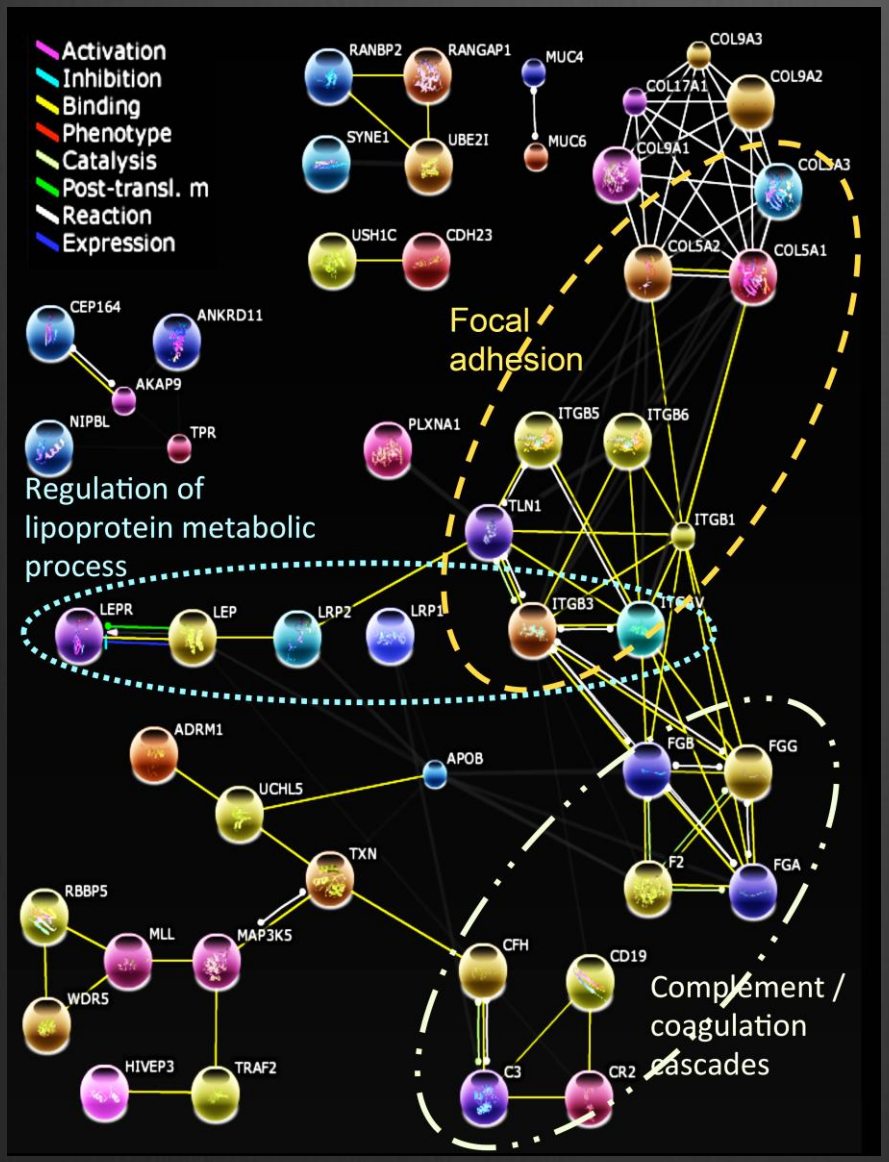
- Sprague-Dawley rats, male, ~10 weeks old
- Single 1ppm x 4 h O_3
- Infusion of a dilute (10%) serum in the lumen of isolated coronary arteries leads to impaired vasodilation and ⇧tone
- No effect of serum from air-exposed rats (control)
- No Δ in serum cytokines
- >1,400 findings by proteomics

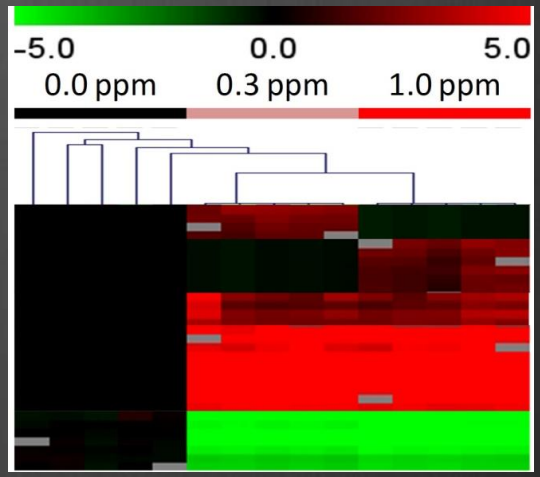


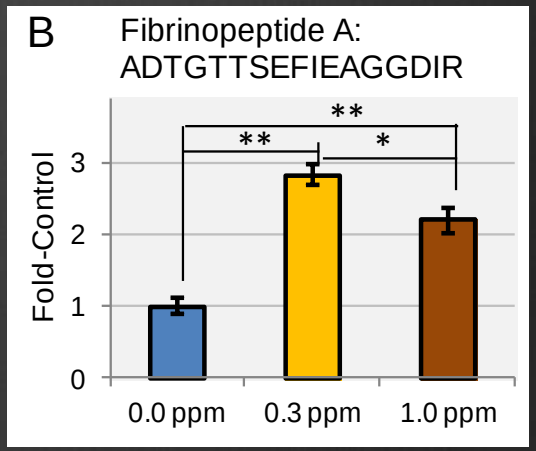




Ozone Serum Proteomics



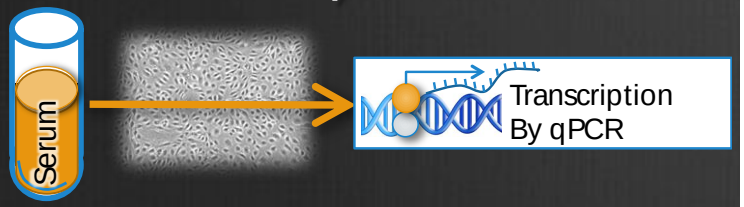




Courtesy of Andrew Ottens, PhD
Virginia Commonwealth University

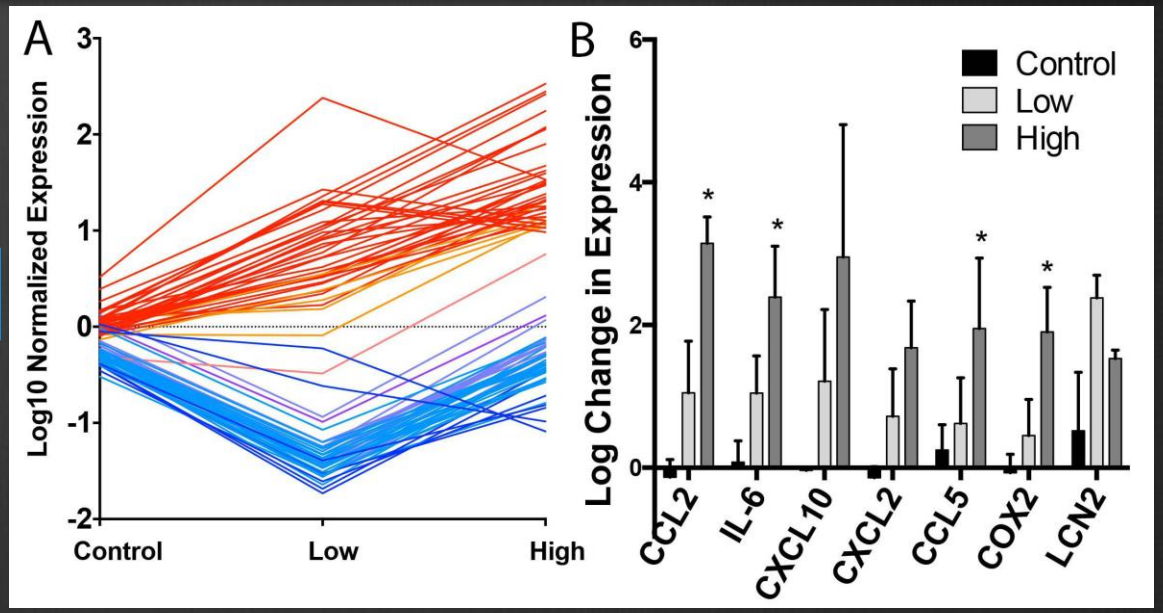
MWCNT Microarray and Gene Ontology: Inflammatory Response

- Serum from instillations of 10 and 40 µg MWCNTs in C57BL/6 mice
- On murine cerebrovascular ECs
- Affymetrix microarray data (KUGR, G.Pickett)



Upregulated Pathways

Gene Ontology	#Genes	P value	Bayes Factor
response to biotic stimulus	15	< 0.0001	20
immune response	12	< 0.0001	15
defense response	12	< 0.0001	14
response to stimulus	15	< 0.0001	10
inflammatory response	6	< 0.0001	10
response to stress	10	< 0.0001	9
response to wounding	6	< 0.0001	7
response to pest, pathogen or parasite	7	< 0.0001	7

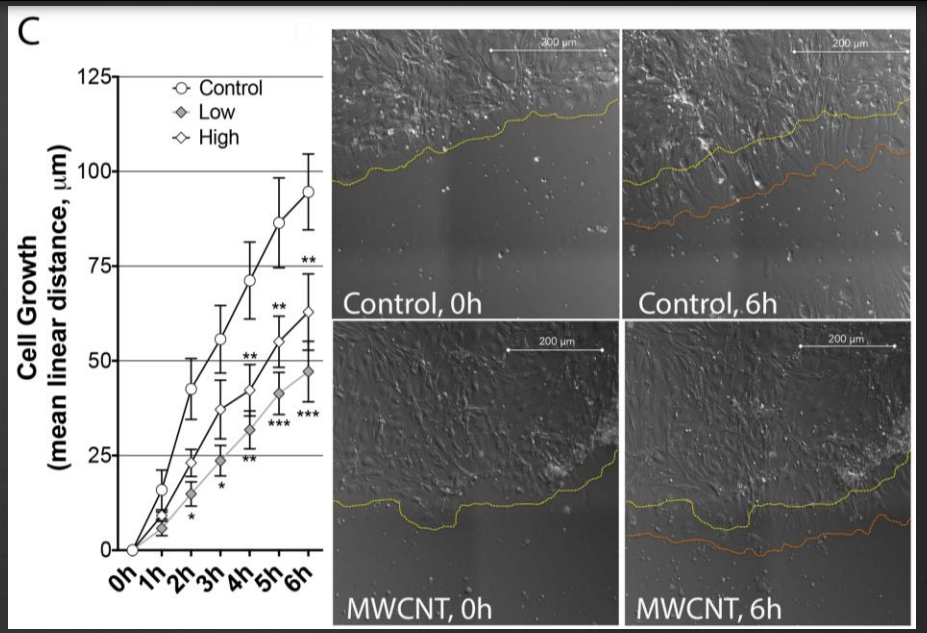


MWCNT Microarray and Gene Ontology: Kinetic Response

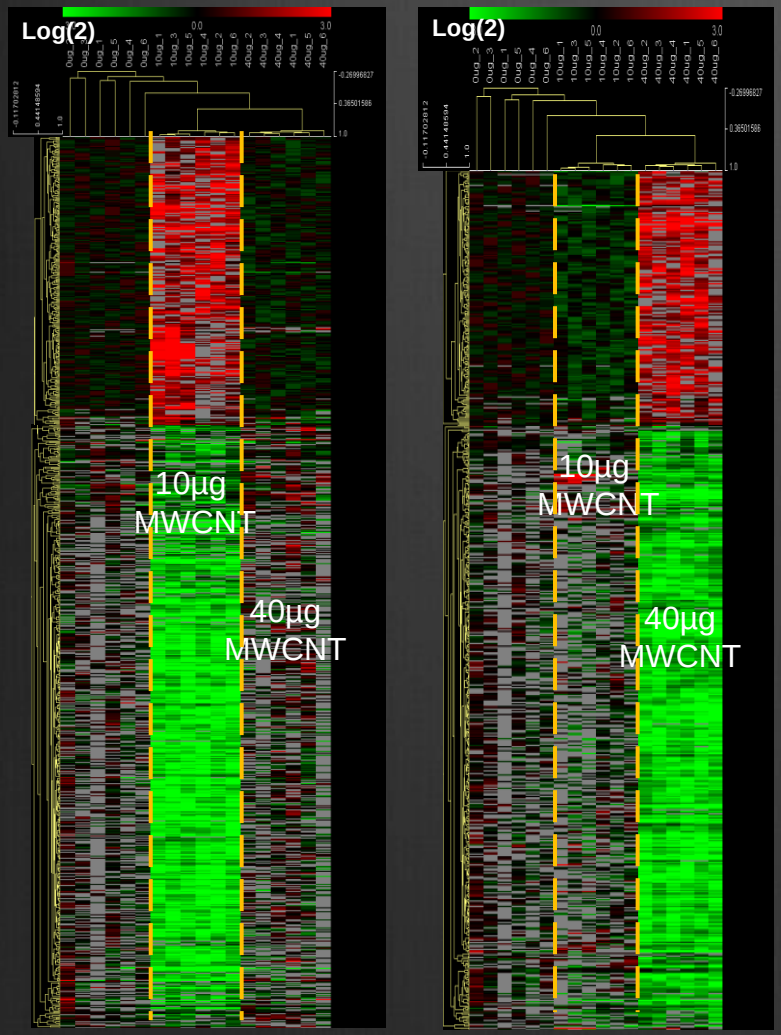
- Confluent mouse cerebrovascular ECs were ‘scratched’ with a sterile pipette tip
- Regrowth captured via live-cell imaging system for a 6 h window
- Cells incubated with serum from 0, 10, 40 ug MWCNT-instilled mice

Downregulated Pathways

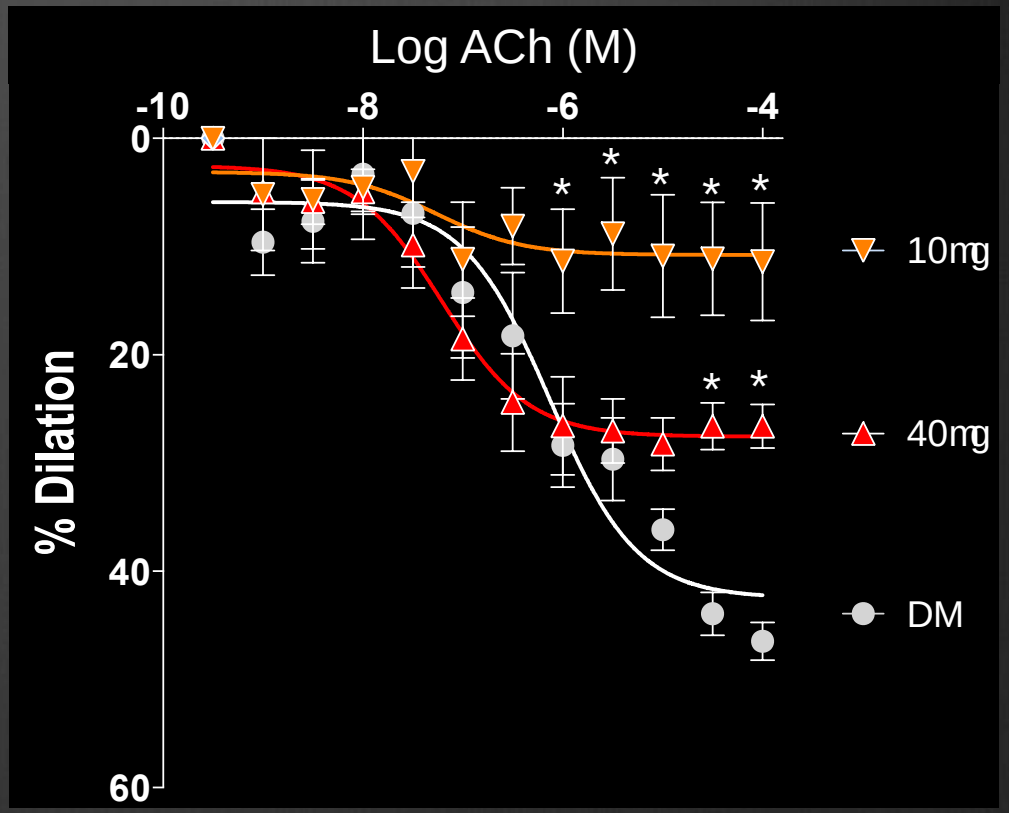
Gene Ontology	#Genes	P value	Bayes Factor
Cell Proliferation	18	<0.0001	33
M phase	11	<0.0001	30
Cell cycle	15	<0.0001	28
Mitosis	9	<0.0001	26
M Phase of Mitotic Cell Cycle	9	<0.0001	26
Nuclear Division	9	<0.0001	23
Mitotic Cell Cycle	9	<0.0001	23
Cytokinesis	6	<0.0001	20
Microtubule-based process	5	<0.0001	8



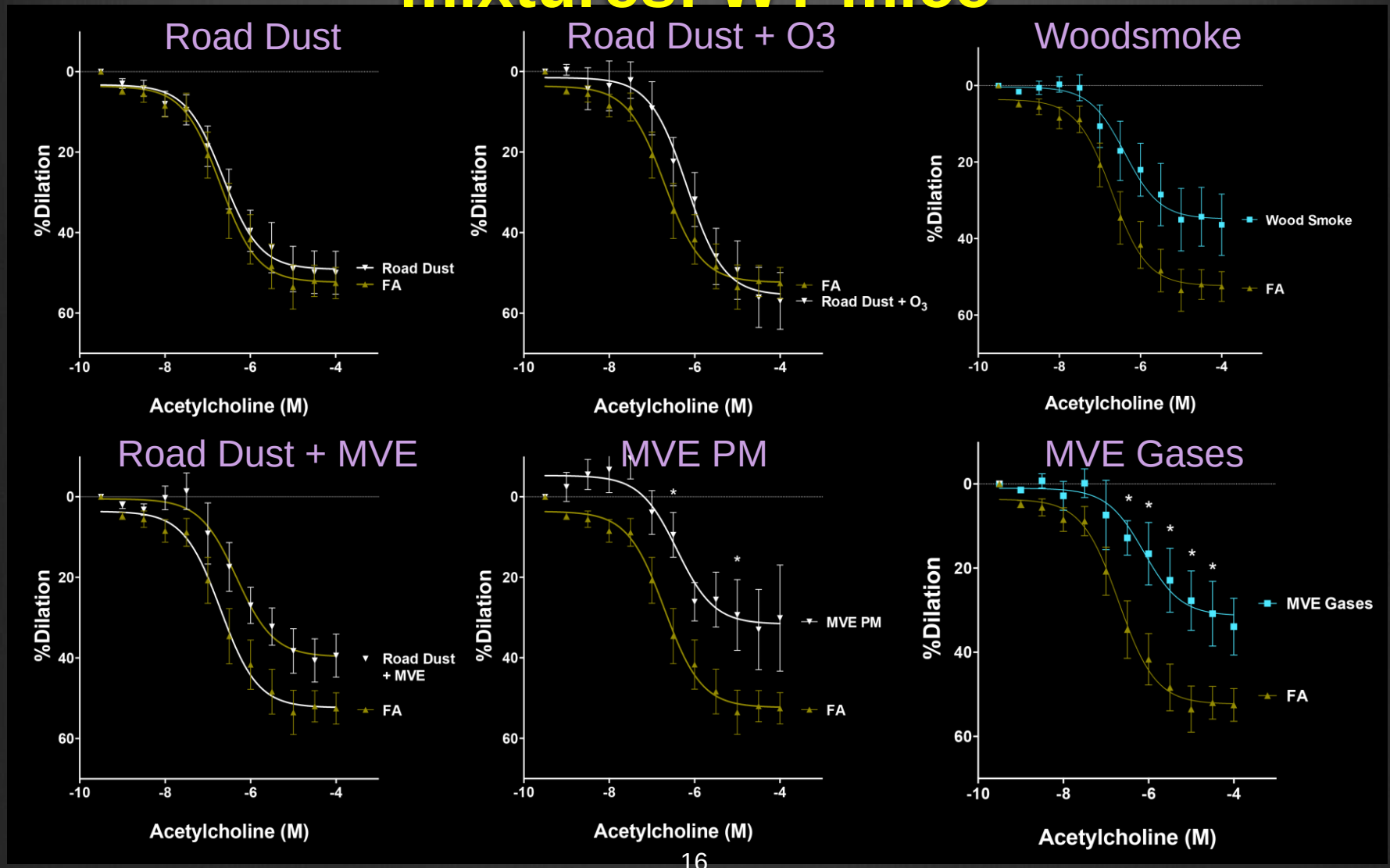
Different MWCNT Doses Induce Different Compositional Changes, Bioactivity



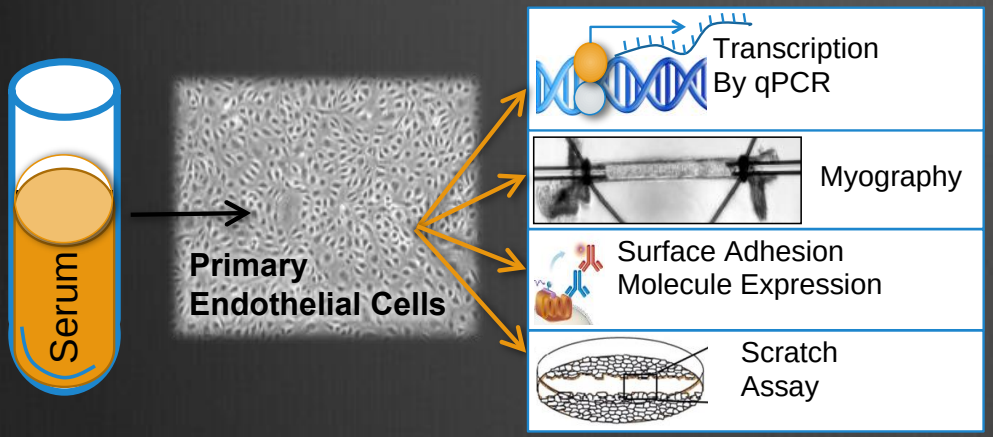
- Biphasic serum compositional changes
- Myography with serum (1%) on naïve aortic rings reveals a non-linear dose response

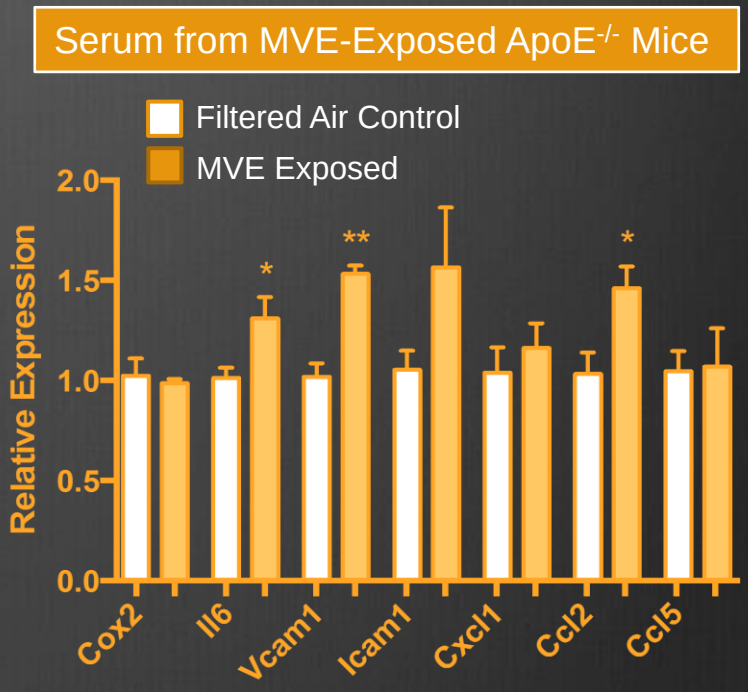


Serum bioactivity after single 6h exposure to various pollutant mixtures: WT mice



Comparative Impacts of Various Pollutants on Serum Bioactivity





- Serum from WT and ApoE$^{-/-}$ mice
  - Single 6 h exposure to one of 6 atmospheres
- Treated on murine cerebrovascular endothelial cells for 4h
- Isolated mRNA, probing 7 target genes

Metabolomic Changes in Serum after MVE: Study Overview

Study Objective

The goal of this study was to gain insight into metabolic changes that take place in mouse plasma when exposed to automotive exhaust.

Study Design

- Global biochemical profiles were determined in serum samples from mice exposed to filtered air, 100 µg PM/m^3 or 300 µg PM/m^3 of exhaust.
- Exhaust exposure time was for 6 hours
- Serum was collected immediately following 6 hours exposure (0 hr) and at 18 hours following exposure.
- Comparisons examine both dose and time effects.

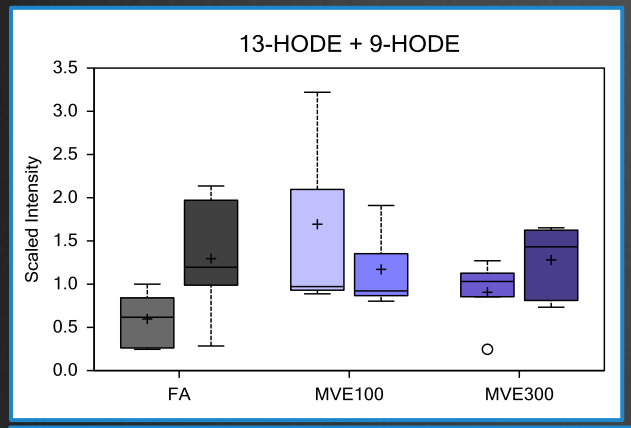
	Filtered Air	100 µg PM/m^3	300 µg PM/m^3
0 hours post	N=6	N=6	N=6
18 hours post	N=6	N=6	N=6

Metabolomics Statistical Summary

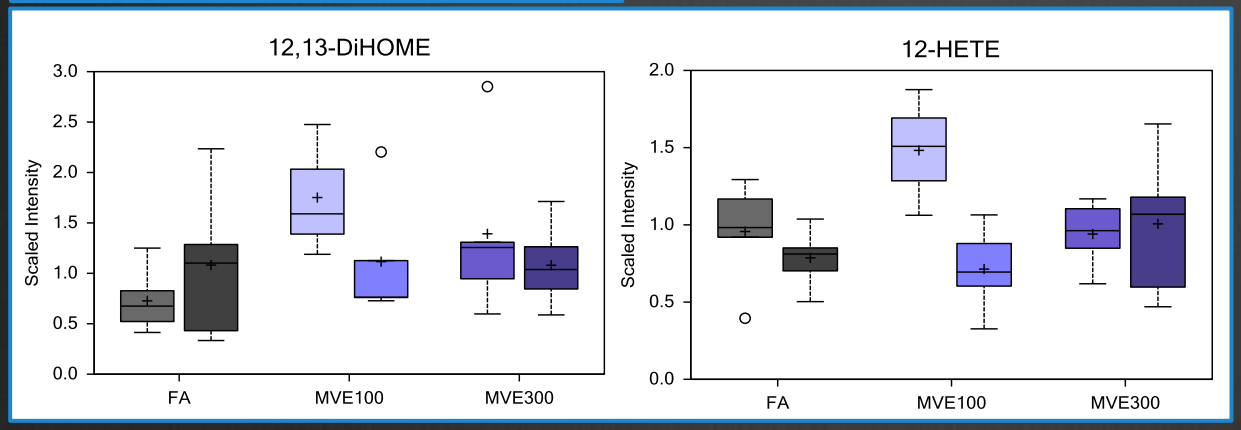
Statistical Comparisons					
Two-Way ANOVA Contrasts (no outliers)	**FA-18 / FA-0**	**MVE100-18 / MVE100-0**	**MVE300-18 / MVE300-0**	**MVE100-0 / FA-0**	**MVE300-0 / FA-0**
Total biochemicals $p \leq 0.05$	123	126	80	40	26
Biochemicals (↑|↓)	107 | 16	78 | 48	47 | 33	25 | 15	26 | 0
Total biochemicals $0.05 < p < 0.10$	26	31	32	30	25
Biochemicals (↑|↓)	19 | 7	13 | 18	18 | 14	23 | 7	23 | 2
Two-Way ANOVA Contrasts (no outliers)	**MVE300-0 / MVE100-0**	**MVE100-18 / FA-18**	**MVE300-18 / FA-18**	**MVE300-18 / MVE100-18**	**MVE300-18 / MVE100-0**
Total biochemicals $p \leq 0.05$	24	27	6	23	101
Biochemicals (↑|↓)	22 | 2	8 | 19	0 | 6	13 | 10	71 | 30
Total biochemicals $0.05 < p < 0.10$	24	20	12	26	27
Biochemicals (↑|↓)	16 | 8	7 | 13	3 | 9	13 | 13	12 | 15
ANOVA Main Effects (no outliers)	**Dose Main Effects**		**Time Main Effects**		**Dose:Time Interaction**
Total biochemicals $p \leq 0.05$	17		170		31
Total biochemicals $0.05 < p < 0.10$	17		29		36

*From a total of **382** named biochemicals*

Lipid metabolites associated with inflammation and peroxidation



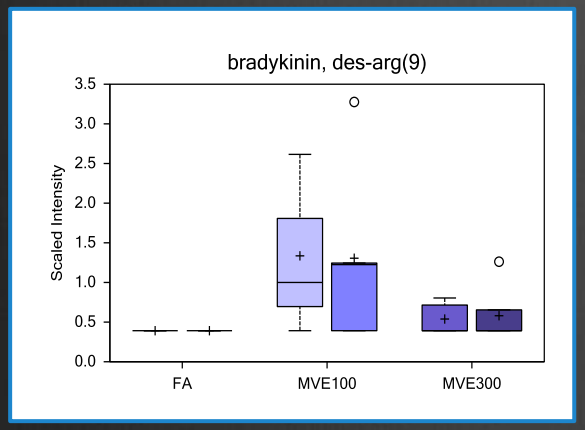
Sub Pathway	Biochemical Name		Fold Change Two-Way ANOVA Contrasts (no outliers)			
			MVE100-0	MVE300-0	MVE100-18	MVE300-18
			FA-0	FA-0	FA-18	FA-18
Fatty Acid, Monohydroxy	13-HODE + 9-HODE		2.84	1.52	0.9	0.99
Fatty Acid, Dihydroxy	12,13-DiHOME		2.41	1.91	1.03	1
Eicosanoid	12-HETE		1.55	0.98	0.91	1.28
Endocannabinoid	palmitoyl ethanolamide		1.24	1.27	1.01	1.12
Ascorbate and Aldarate Metabolism	oxalate (ethanedioate)		2.19	1.55	0.98	0.92

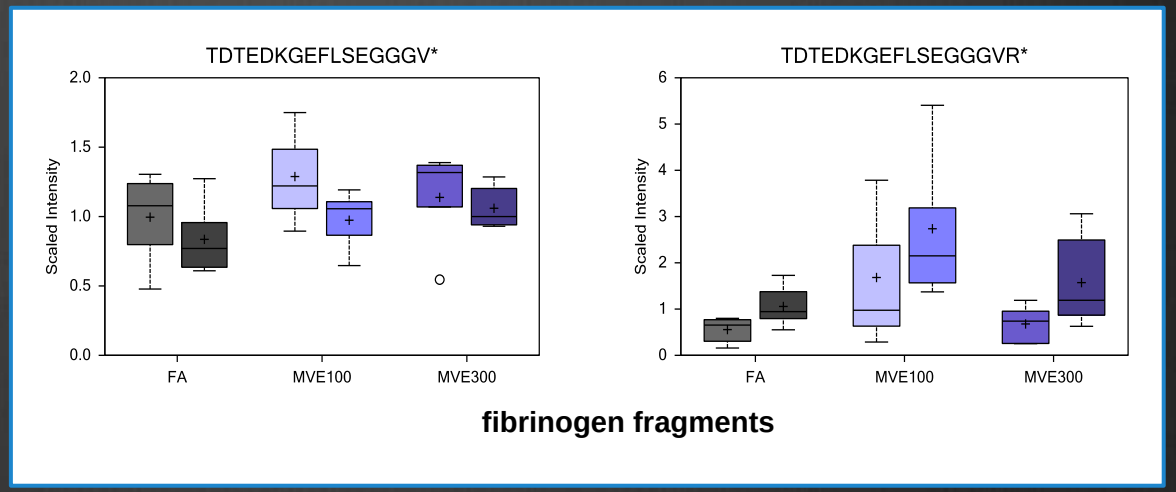


- 13-HODE/9-HODE are measures of lipid peroxidation and were increased in MVE serums.
- 12, 13-DiHOME has been reported to be a neutrophil attractant and may indicate increased inflammation.
- 12-HETE is derived from LO metabolism of arachidonic acid and is reported to act as a vasoconstrictor.

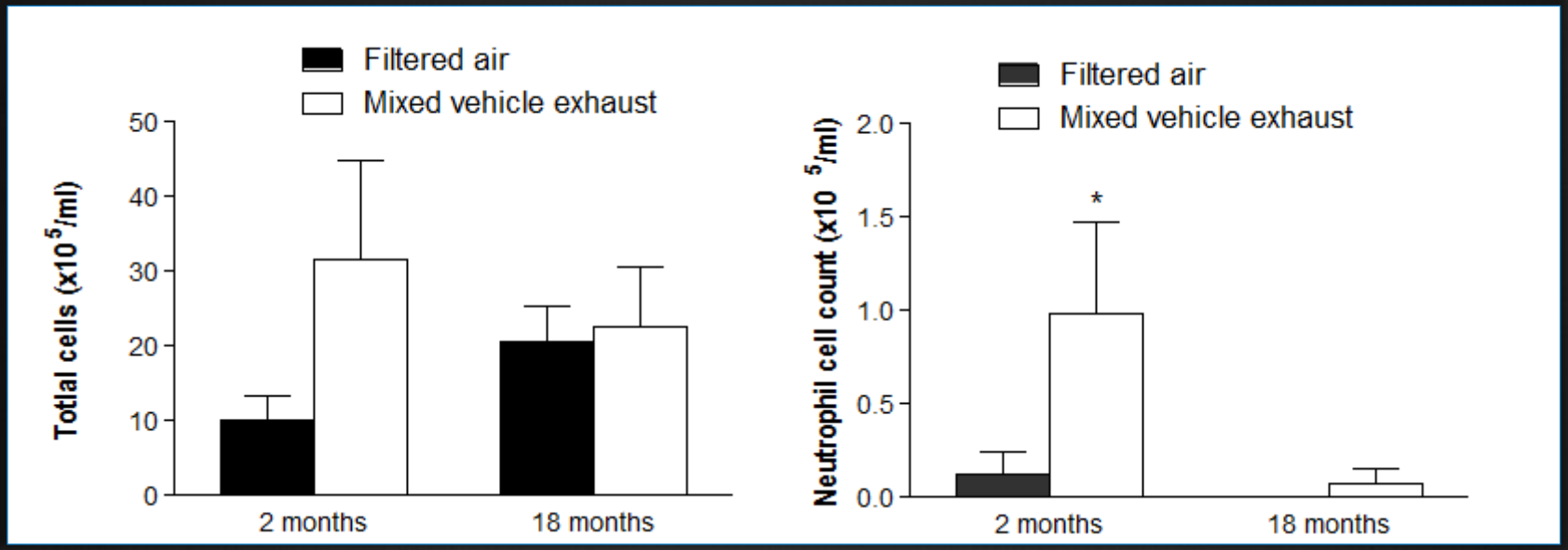
Metabolites with potential cardiovascular significance

Sub Pathway	Biochemical Name		Fold Change Two-Way ANOVA Contrasts (no outliers)			
			MVE100-0 FA-0	MVE300-0 FA-0	MVE100-18 FA-18	MVE300-18 FA-18
Polypeptide	bradykinin, des-arg(9)		3.41	1.38	3.33	1.48
Fibrinogen Cleavage Peptide	TDTEDKGEFLSEGGGV*		1.29	1.14	1.16	1.27
	TDTEDKGEFLSEGGGVR*		3.02	1.22	2.59	1.49
Purine Metabolism, Adenine containing	adenosine		0.14	0.66	0.86	0.68
Urea cycle; Arginine and Proline Metabolism	arginine		0.75	0.95	0.89	1.01
Histidine Metabolism	histidine		0.82	1	0.99	0.87
	1-methylimidazoleacetate		1.74	1.17	1.08	1.11



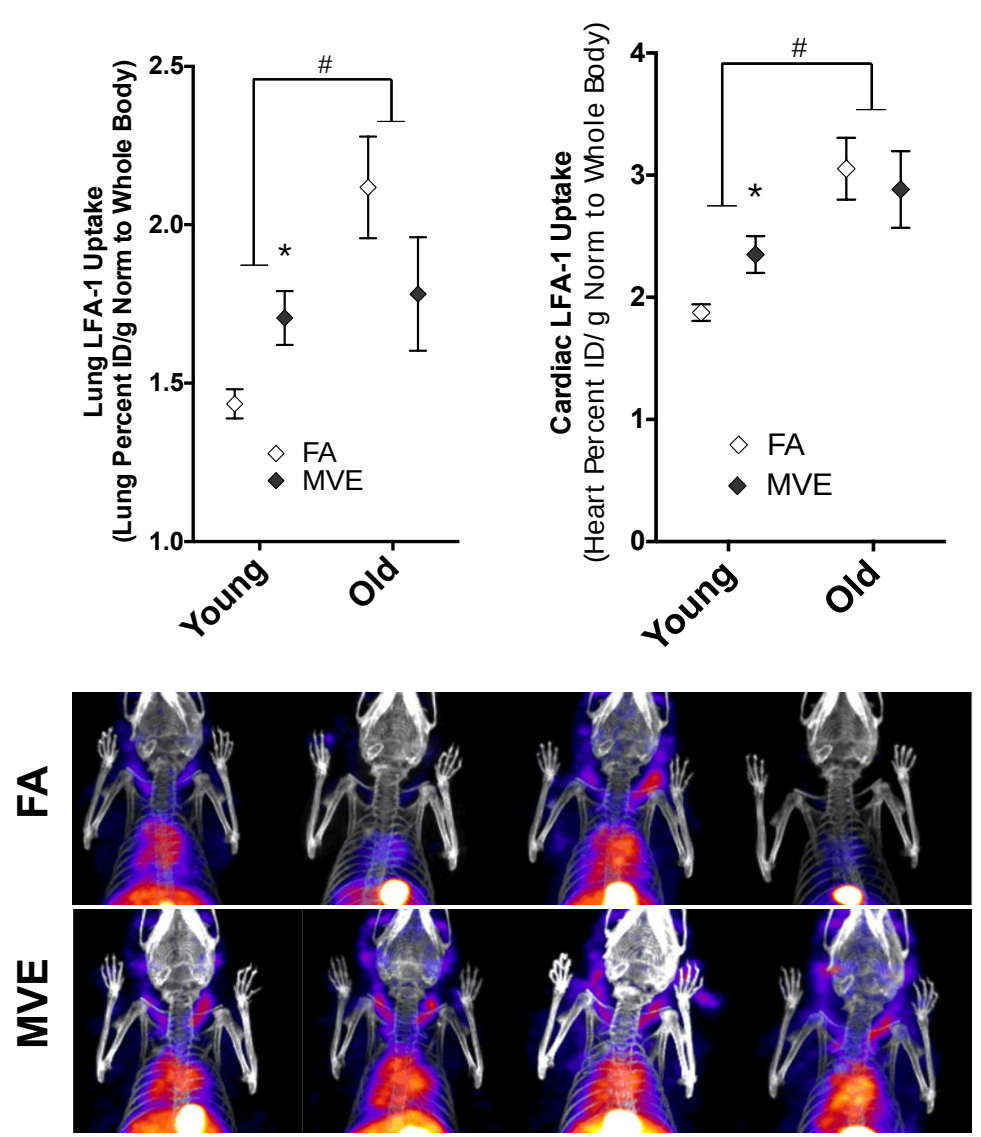


Old vs Young: MVE-Induced Pulmonary Inflammation

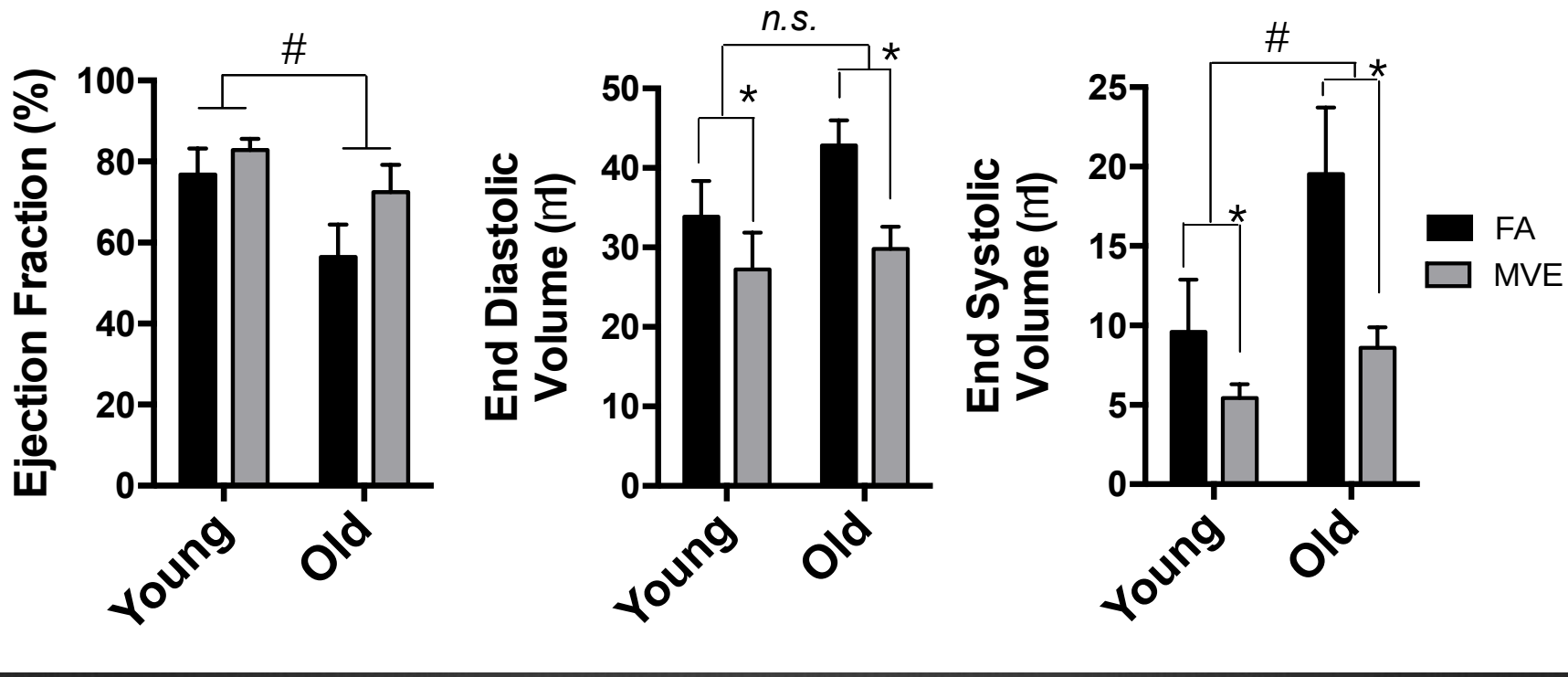


- C57BL/6 mice at 2 mo or 18 mo of age
- Exposed 50 days to MVE, 300 µg PM/m^3

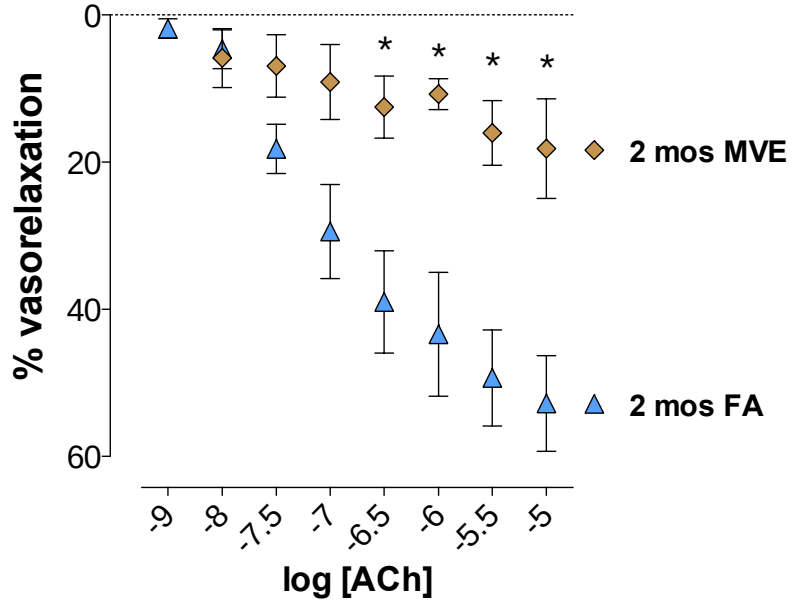
MVE-Induced Thoracic Inflammation

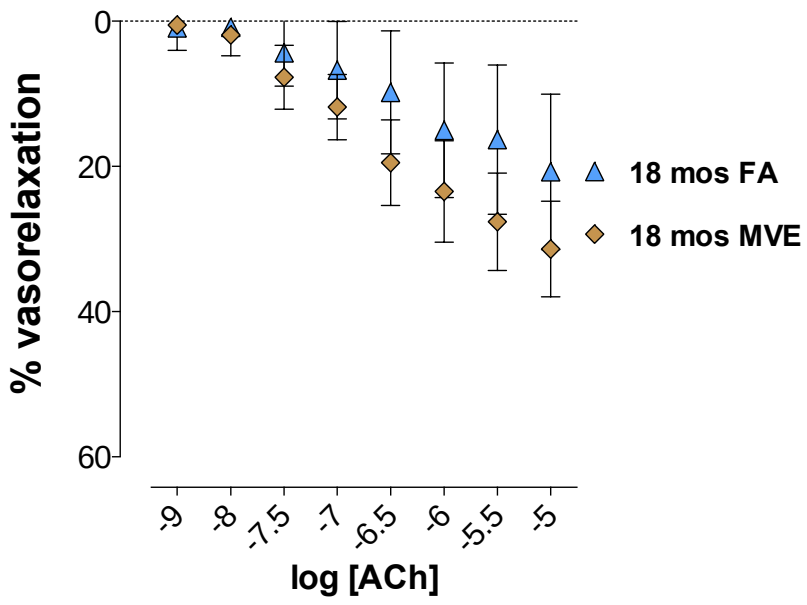


End Diastolic, Systolic Decreased by MVE



Serum-Induced Impairment of Vasorelaxation





Summary: Endothelial Cell Biosensors as Tools for Assessing Serum "Exposome"

- Findings to date are consistent with the premise that circulating inflammatory factors are generated by pulmonary exposures to numerous pollutants
  - Suggests commonality of mode of action
  - Specific cytokines are unlikely to drive the entirety of this effect
- Use of canonical inflammatory outcomes, such as VCAM and IL-8, has been fruitful, but genomic approaches are ongoing to identify stronger responders and potentially unique patterns of response
- Strong parallels between human and animal work using this 'biosensor' assay paradigm
- Proteomic/lipidomic/metabolomic assessment of the serum, conducted in parallel with functional assays, provide many leads, but also challenges related to mechanism
- Ongoing work comparing various complex emissions suggests fresh emissions and gas-PM mixtures remain the most potent

Upcoming…

- Wrapping up!
  - Integrated studies with Project 2: Exploring the metabolomic findings – timing and dose – related to biopotential of serum across atmospheres
  - Banked serum from Human studies:
    - Rob Brooks' CAPs studies
    - Bob Devlin and Mike Madden's Ozone and Diesel studies
  - Coronary vascular effects of O_3 and serum contribution – waiting on NO scavenging data and considering some proteome assays
  - Old vs young mice
- Future work:
  - Genomics, proteomics on complex emissions
  - Bioactivity-guided fractionation (currently looking at bioactivity of small fragments, <10kD)

Acknowledgements

UNM

- Mario Aragon
- Heidi Cung
- Michael Paffett, PhD
- Sarah Robertson, PhD
- Sage Colombo, MD
- Meghan Channell, MS
- Selita Lucas
- Joe Anderson, Pharm.D.
- Jim Nawarskas, Pharm.D.
- Gavin Pickett, PhD, KUGR

US EPA

- Michael Madden, PhD
- Bob Devlin, PhD

High Point Univ

- James Smoliga, DVM, PhD

Lovelace

- Jake McDonald, PhD
- Izabela Chrobak, PhD
- Jeremy Brower, PhD
- Ben Moeller, PhD

Virginia Commonwealth University

- Andrew Ottens

NIOSH

- Aaron Erdely, PhD

Funding:

- **EPA** R83399001
- **NIH** ES014369
- **NIH** OH010495

UW CCAR Spatio-temporal Air Quality Modeling

CCAR Biostat Core

CCAR SAC Annual Meeting

Seattle

Oct 6, 2014

Outline

1. Elements of universal kriging: spatial or "land use" regression with kriging/spatial smoothing
2. UW Spatio-temporal model
3. Evaluation of model predictions, R^2 and cross-validation
4. Applications
   - MESA Air spatio-temporal application
   - Regionalized national NO2 model, with satellite data
   - Historical modeling/prediction of PM2.5
   - Spatially predictive PCA and k-means clustering (posters)

1. Universal kriging: spatial or "land use" regression with kriging/spatial smoothing

- *Land use regression:* $C(s) = X(s)'\beta + \nu(s)$

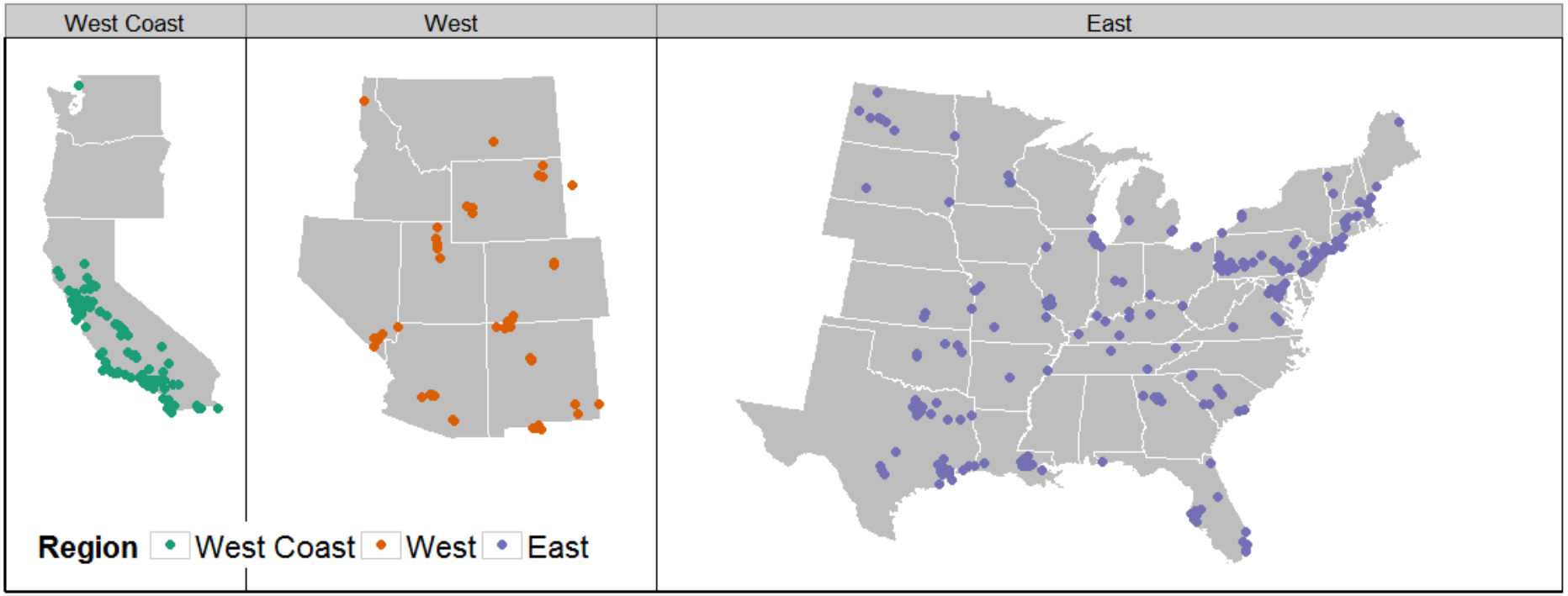
  where $X(s)$ is a vector of covariates measured at location s, as many as 300 geographic covariates, including

Distance to:	Buffer values:	Misc:
A1, A2, A3 road	A1, A2, A3 road length	Street canyon metrics (NY & IL)
Truck Route	Intersections	Absolute Elevation
Railroad	Truck Route Length	Long-term CALINE value:
Railyard	Emissions	• Traffic volumes
Airport	Residual Oil output (NY)	• Roadway locations
Port	% Land Use	• Diurnal traffic patterns
Coast	Impervious surface	• Wind speed, wind direction, mixing height
Oil Boiler (NY)	NDVI	
	Population	

LUR *must* incorporate kriging/smoothing

- Surprisingly many applications of LUR assume independent errors $\nu(s)$, or simply ignore spatial correlation structure
- Spatial prediction at location s_0 with LUR $\hat{C}(s_0) = X(s_0)'\hat{\beta} + \hat{\nu}(s_o)$ where $\hat{\nu}(s_0)$ derives from kriging equations or spline smoothing of residuals.
- Depending on spatial scale, nonstationarity in the sense of spatially varying coefficients and residual correlation structure should be considered.

Coarse approach: regionalized modeling



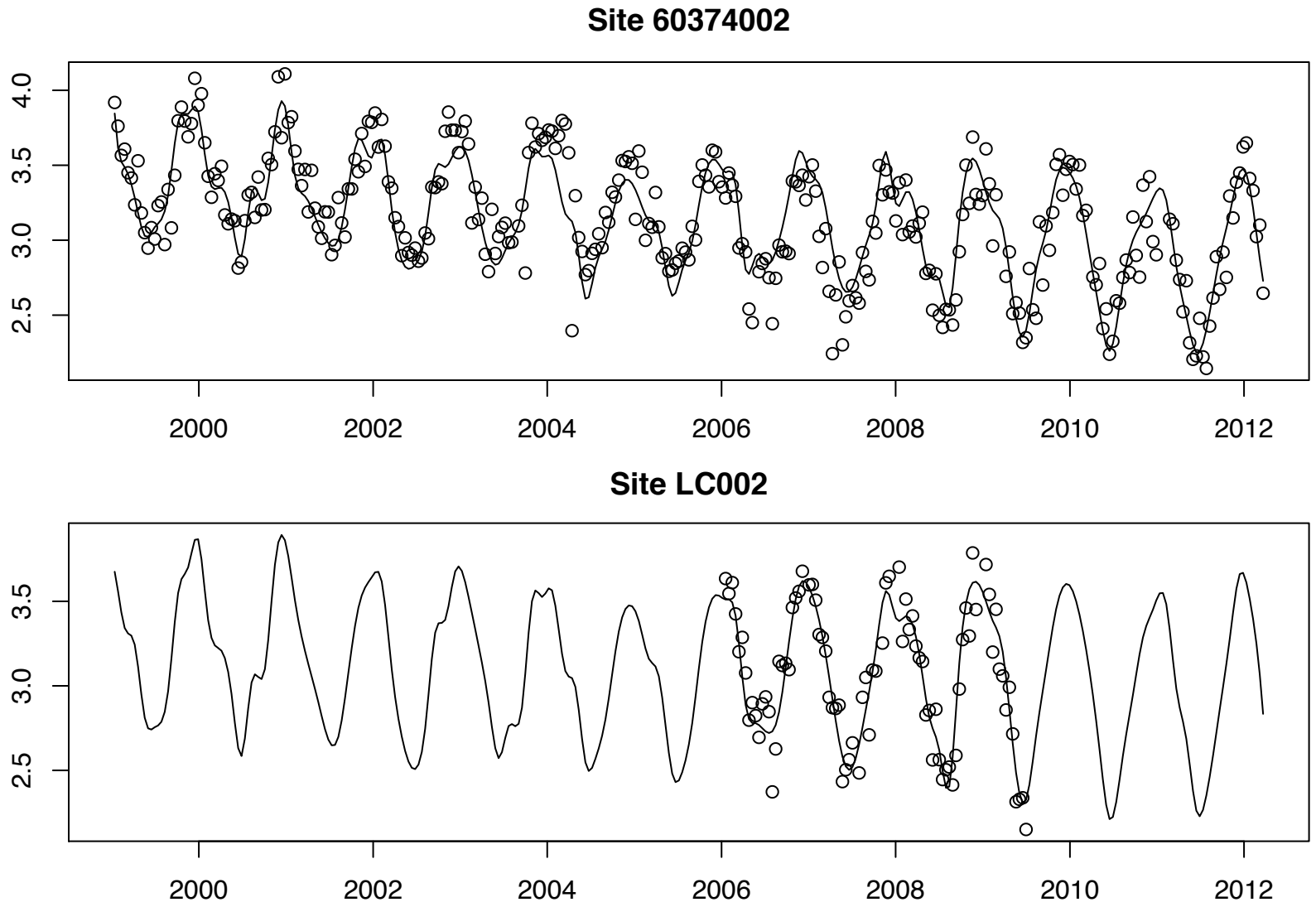
Spatially varying land-use regression and kriging parameters
(examples coming below)

2. UW Spatio-temporal model

- Strategies for spatio-temporal modeling:
  - *Spatial models* (like the above spatial regression) *varying in time*, including "dynamic" models
  - *Temporal models* (at monitoring sites and arbitrary locations) *varying in space*. The UW CCAR model is of this latter type.

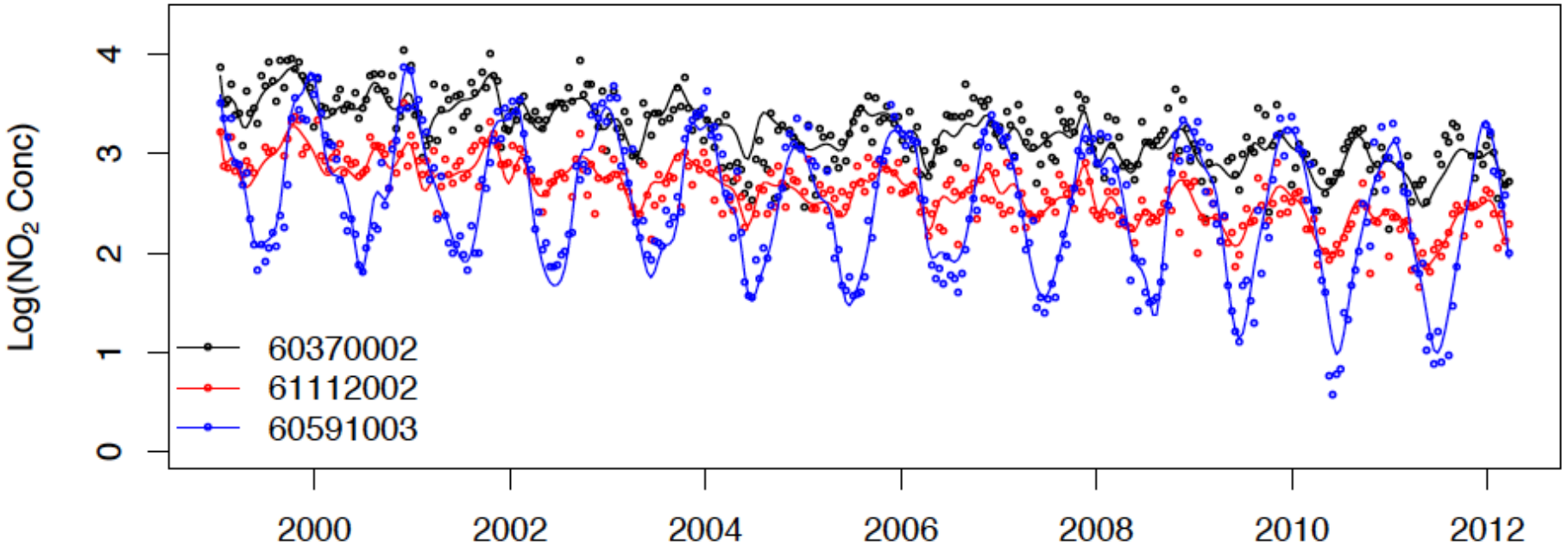
NO_2 Time Trends in CA

$$\beta_0(s) + \sum_{i=1}^{n} \beta_i(s) f_i(t) + \nu(s,t)$$



NO_2 Time Trends in CA

$$\beta_0(s) + \sum_{i=1}^{n} \beta_i(s) f_i(t) + \nu(s,t)$$

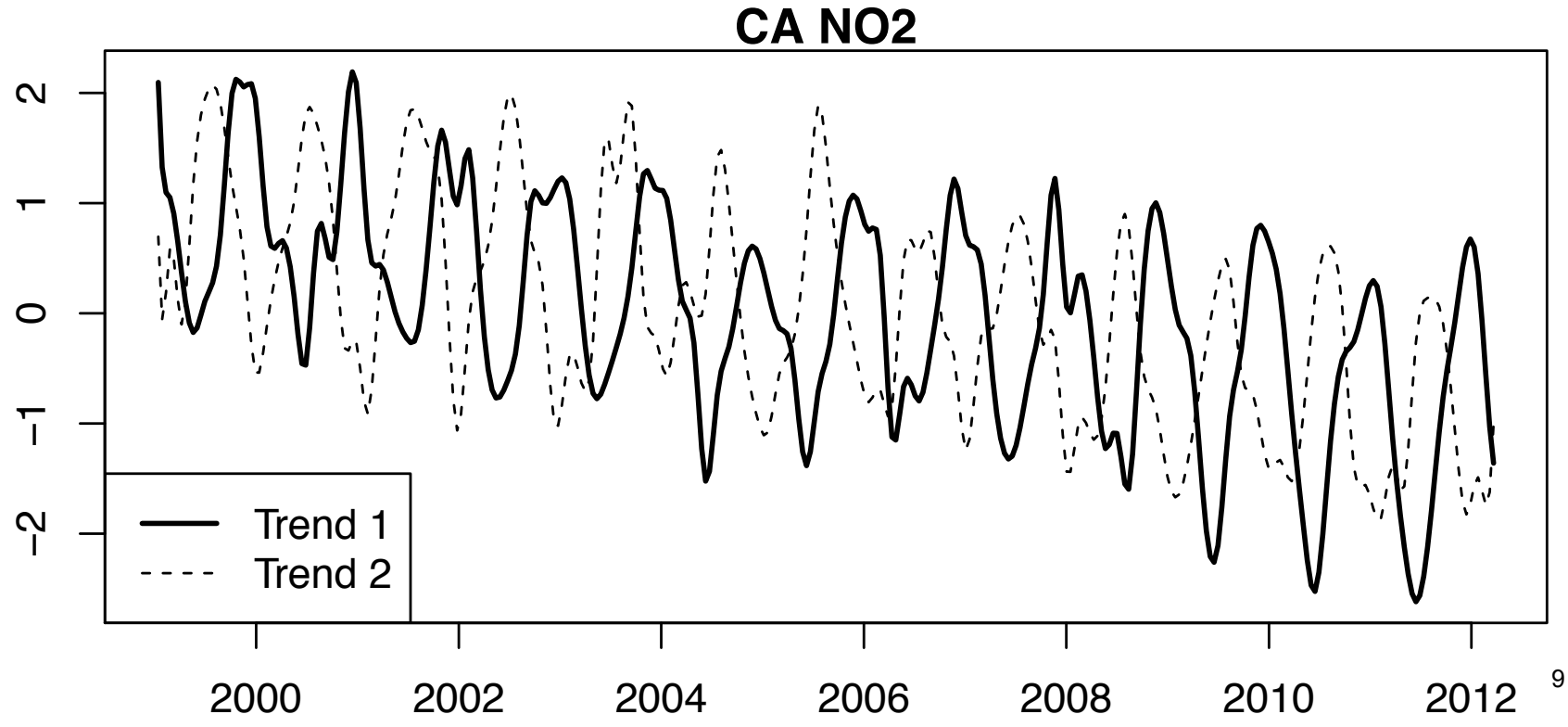


060370002: Azusa, Los Angeles county
061112002: Simi Valley, Ventura county
060591003: Costa Mesa, Orange county

NO_2 Time Trends in CA

$$\beta_0(s) + \sum_{i=1}^{n} \beta_i(s) f_i(t) + \nu(s,t)$$

- Represent time trends as combinations of smooth, temporal trend basis functions computed from SVD of AQS monitoring data



Statistical Model

$$C(s,t) = \beta_0(s) + \sum_{i=1}^{m} \beta_i(s) f_i(t) + \nu(s,t)$$

- $C(s, t)$ – (log) conc. at location s at time t
- $\beta_i(s)$ – spatial random fields
  - $\beta_i(s) \sim N\left(\mathbf{X}_i(s)\alpha_i, \Sigma(\tau_i, \sigma_i, \varphi_i)\right)$
  - $\mathbf{X}_i(s)$ – geographic ("land use") covariates
  - $\Sigma(\tau_i, \sigma_i, \varphi_i)$ – covariance function
  - $\beta_0(s)$ – long-term means at location s
- $f_i(t)$ – temporal trend basis functions
- $\nu(s,t)$ – residual space-time field

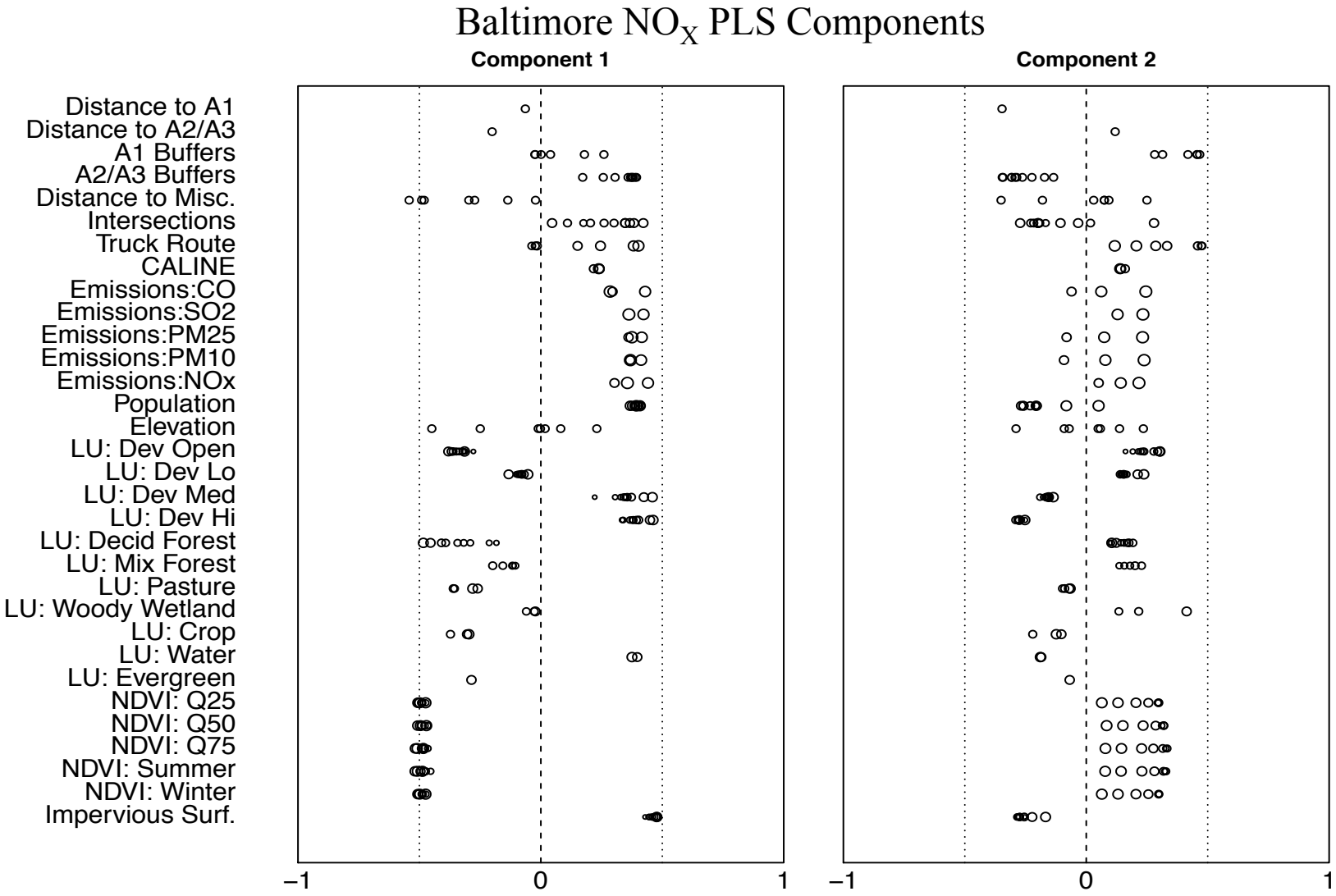
Covariate Selection

$$\beta_0(s) + \sum_{i=1}^{n} \beta_i(s) f_i(t) + \nu(s,t)$$

- Over 300 geographic covariates at each site. Variable selection or dimension reduction?
  - with a larger number of highly correlated covariates (e.g. lengths of roads in buffers of 50m, 100m, 150m, …), we usually choose dimension reduction
- Partial Least Squares (PLS)
  - Conceptually similar to PCA, but components computed as linear combinations of original covariates to *maximize covariance* between Y and X.
  - Typically retain small number of combinations, 1, 2, 3.
  - Note: PLS scores pre-computed from estimates of $\beta_i(s)$.

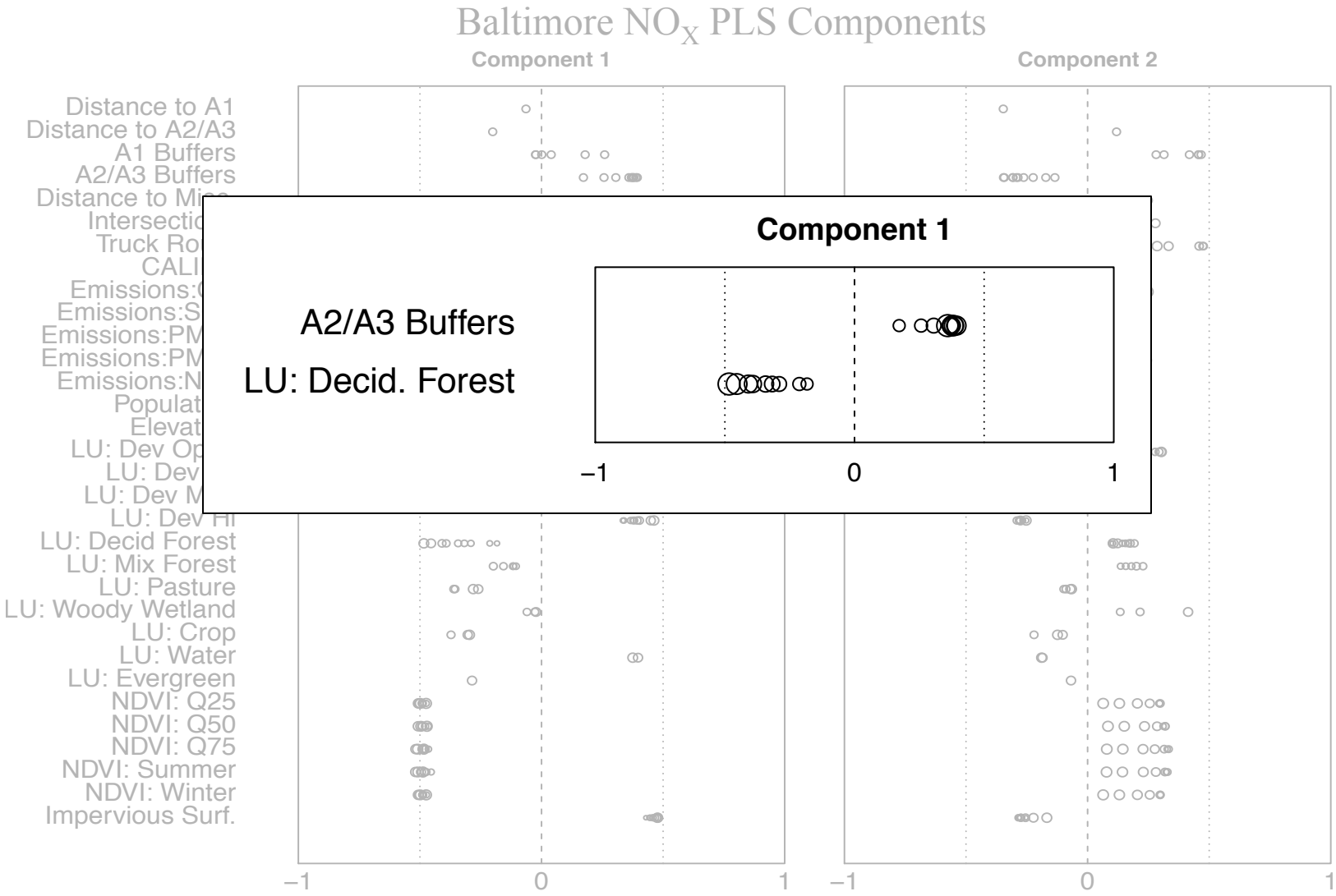
PLS Components

$$\beta_0(s) + \sum_{i=1}^{n} \beta_i(s) f_i(t) + \nu(s,t)$$



PLS Components

$$\beta_0(s) + \sum_{i=1}^{n} \beta_i(s) f_i(t) + \nu(s,t)$$



SpatioTemporal R package

- Maximum likelihood estimation with unbalanced spatio-temporal response data
- Incorporates spatio-temporal covariates $z(s,t)$ --- to date, only CALINE or satellite predictions

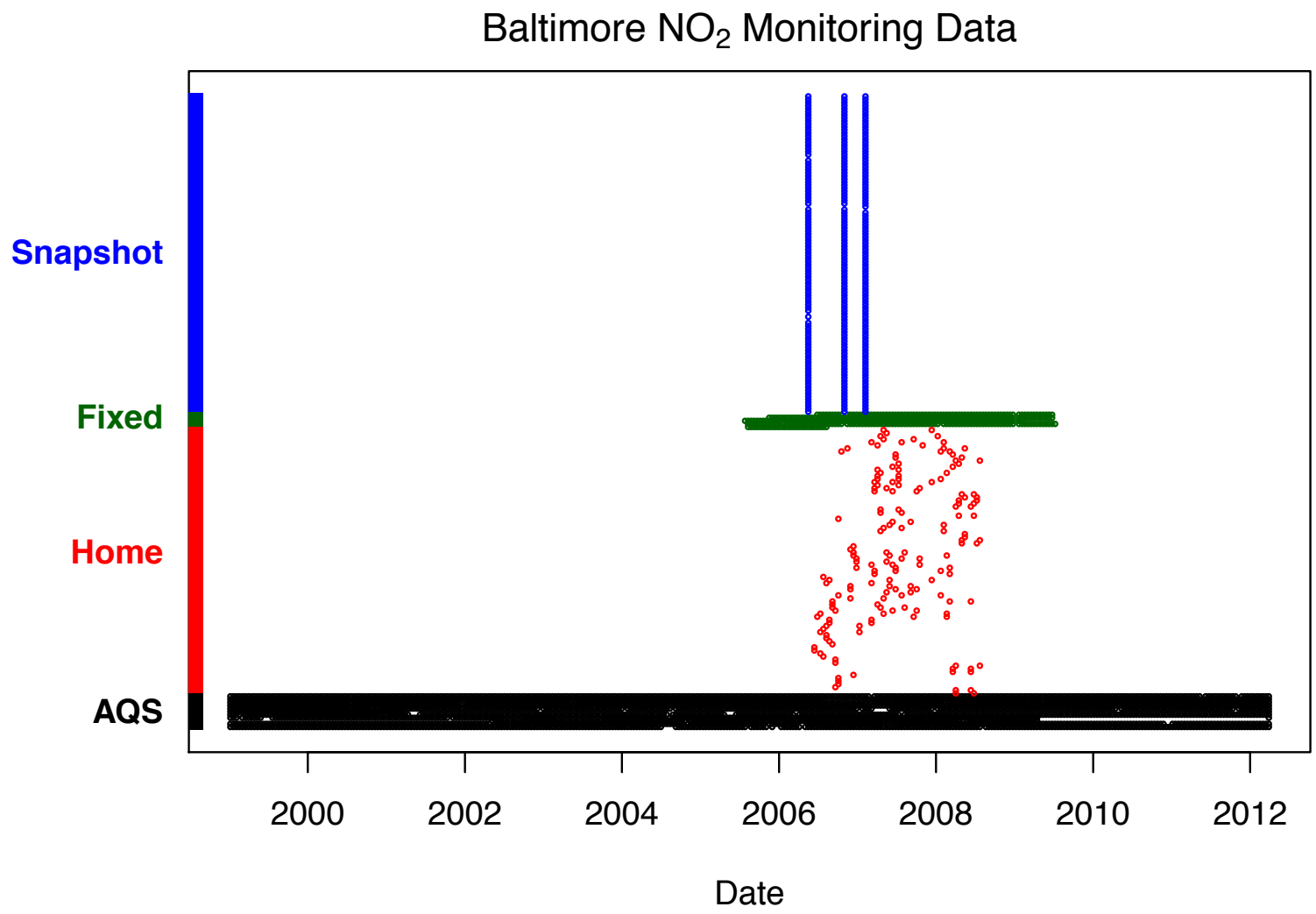
$$C(s,t) = \beta_0(s) + \sum_{i=1}^{m} \beta_i(s) f_i(t) + \gamma\, z(s,t) + \hat{v}(s,t)$$

- Predictions, "universal kriging" from ML fitting:

$$\hat{C}(s_0,t) = \hat{\beta}_0(s_0) + \sum_{i=1}^{m} \hat{\beta}_i(s_0) f_i(t) + \hat{\gamma} z(s_0,t) + \hat{v}(s_0,t)$$

where $\hat{\beta}_i(s_0)$ derive from ML-based krigings of these spatial fields and $\hat{v}(s_0,t)$ similarly.

Monitoring Data



Model Selection

- Vary model parameters
  - Variance structure
    - Spatial Smoothing (Exponential variogram) or No Smoothing (Independent)
  - Number of Time Trends: 1 or 2
  - Number of PLS components: 2 or 3
- Best model for each city and pollutant chosen by cross-validation

3. Evaluation of model predictions

- Standard cross-validation: repeatedly
  1. Split data into training and test sets
  2. Estimate parameters using training data
  3. Predict at space-time locations of test data

  Compute MSE and R^2 of predicted vs. observed data
  - $R^2_{cv} = \max\left(0, 1 - MSE_{pred} / MSE_{obs}\right)$
  - measures fit to 1-1 line, not simply correlation

❑ Temporally adjusted R^2

- Challenging to separate the spatial and temporal contributions to R^2_{cv} for cross-validation of temporally sparse datasets like MESA Air home sites.
- Lindström et al. (2013) introduced three temporally-adjusted adaptations of R^2_{cv} using data from neighboring AQS and fixed sites for the reference MSE instead of MSE_{obs} in order to focus on *spatial prediction accuracy*.

$$R^2_{cv} = \max\left(0, 1 - MSE_{pred} / MSE_{obs}\right)$$

- R^2_{avg} uses spatial average at each time point of observations at AQS/fixed sites within region.
- R^2_{close} uses closest AQS/fixed site
- R^2_{smth} uses smooth temporal trend at the closest site

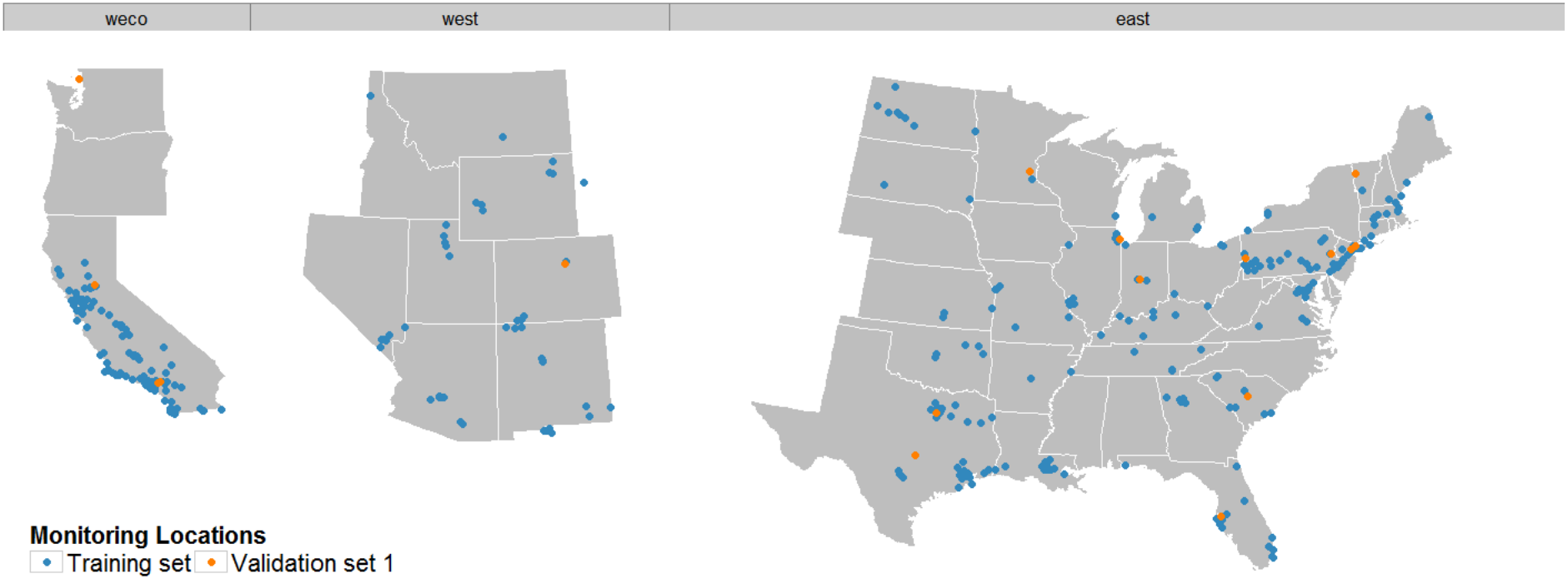
- The resulting R^2s represent the *improvement in predictions* provided by our model compared with central site or nearest neighbor exposure schemes often used in epi studies.

❑ Assessing extrapolation using clustered cross-validation

K-fold cross-validation on a spatially clustered network

- In a spatially-clustered monitoring network, left-out sites tend to be proximal in space to training sites. Despite validation sites being "out of sample"
  - LUR coefficients are estimated from nearby monitoring data
  - Spatial smoothing (kriging) averages values from nearby sites
- Conventional cross-validation methods assess out-of-sample performance *when predicting to points close to monitoring locations* due to spatial clustering in the AQS network.

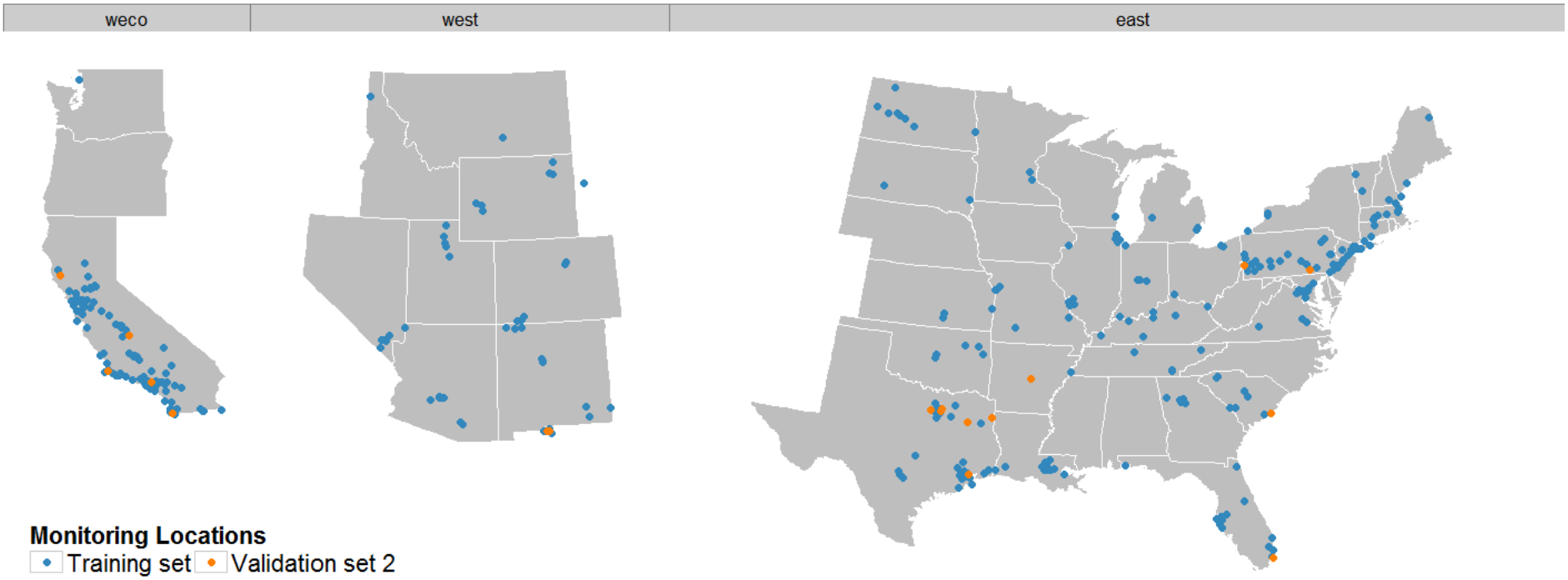
k-fold Cross-Validation



AQS NO_2 sites with complete* data in 2006

In k-fold cross-validation, a group is left out and the model is fit without that validation set. Predictions are then made to the validation set.

k-fold Cross-Validation

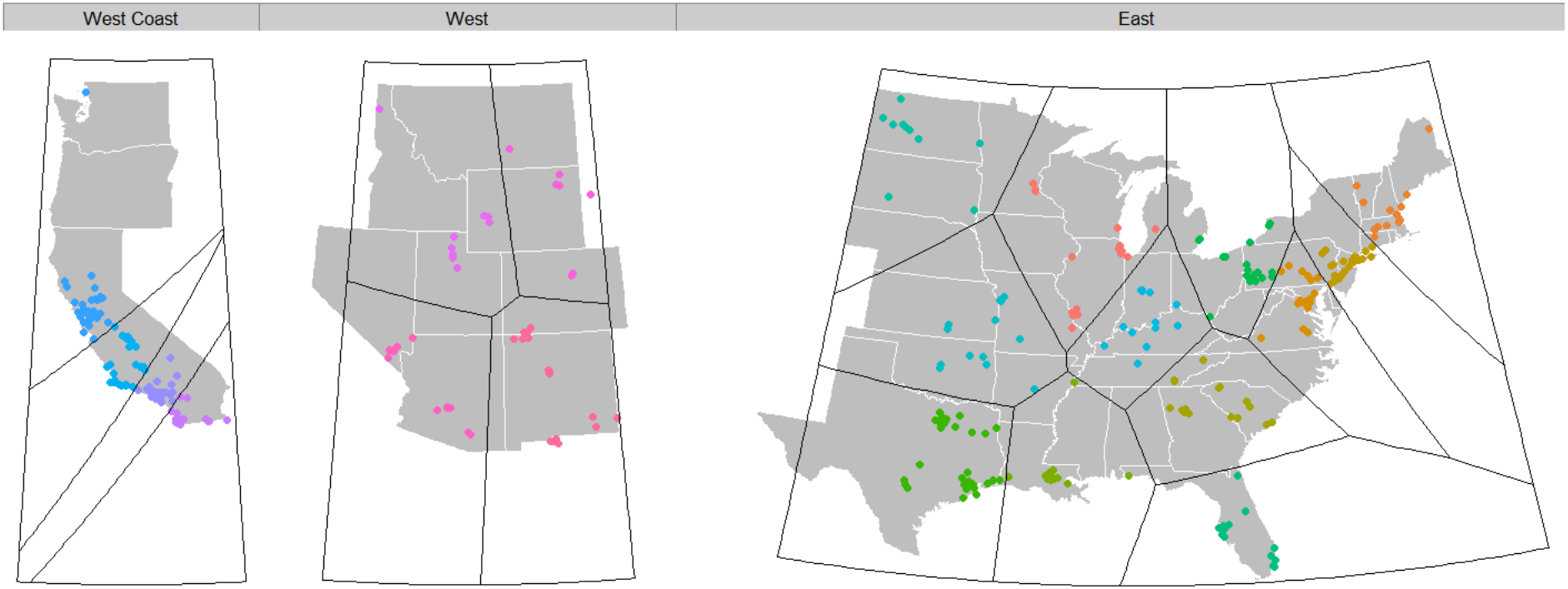


Repeat for all k groups (k=20 here), until we have out-of-sample predictions for the entire dataset.

Spatially-Clustered Cross-Validation

- *Goal:* assess model performance in areas far from monitoring locations
  - (this isn't directly possible)
- *Approach:* Simulate prediction to points far from monitoring sites by *defining cross-validation groups to be spatially clustered*
- *Result:* Estimate of model performance when extrapolating to gaps in monitoring coverage.

Spatially-Clustered Cross-Validation



Each color represents a cross-validation group

(Created using k-means on lat/long for each region)

4. Applications

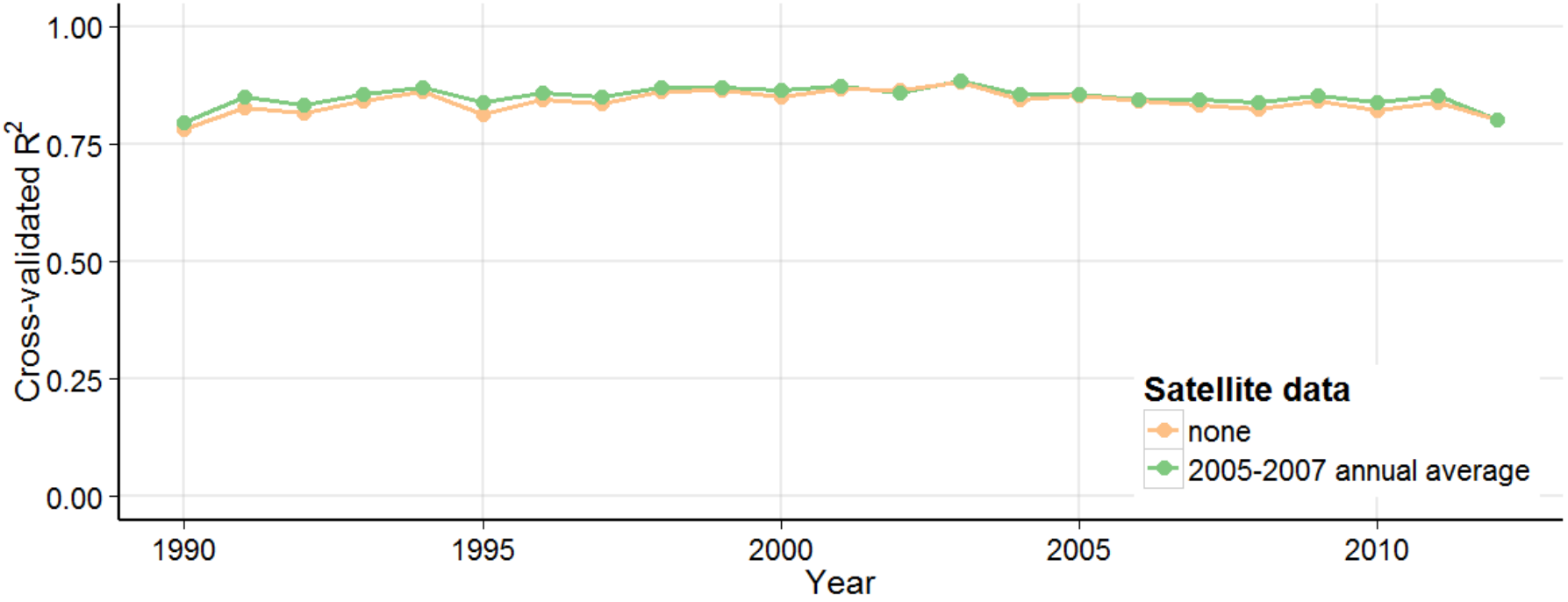
a) MESA Air spatio-temporal applications (Keller)
b) Regionalized national NO_2 model, with satellite data (Young)
c) Historical modeling/prediction of $PM_{2.5}$ (Kim)
d) Spatially predictive PCA (Jandarov) and k-means clustering (Keller)

a) A National Prediction Model Based on Universal Kriging and Land-Use Regression Using Satellite-Based NO_2 Measurements

- Satellite Data[2]
  - Years: 2005-2007 averaged
  - 13x24 km^2 spatial resolution
  - OMI tropospheric NO_2 column data converted to surface levels, 12:00-15:00 local time
  - Raster-to-point conversion using average Moore Neighborhood

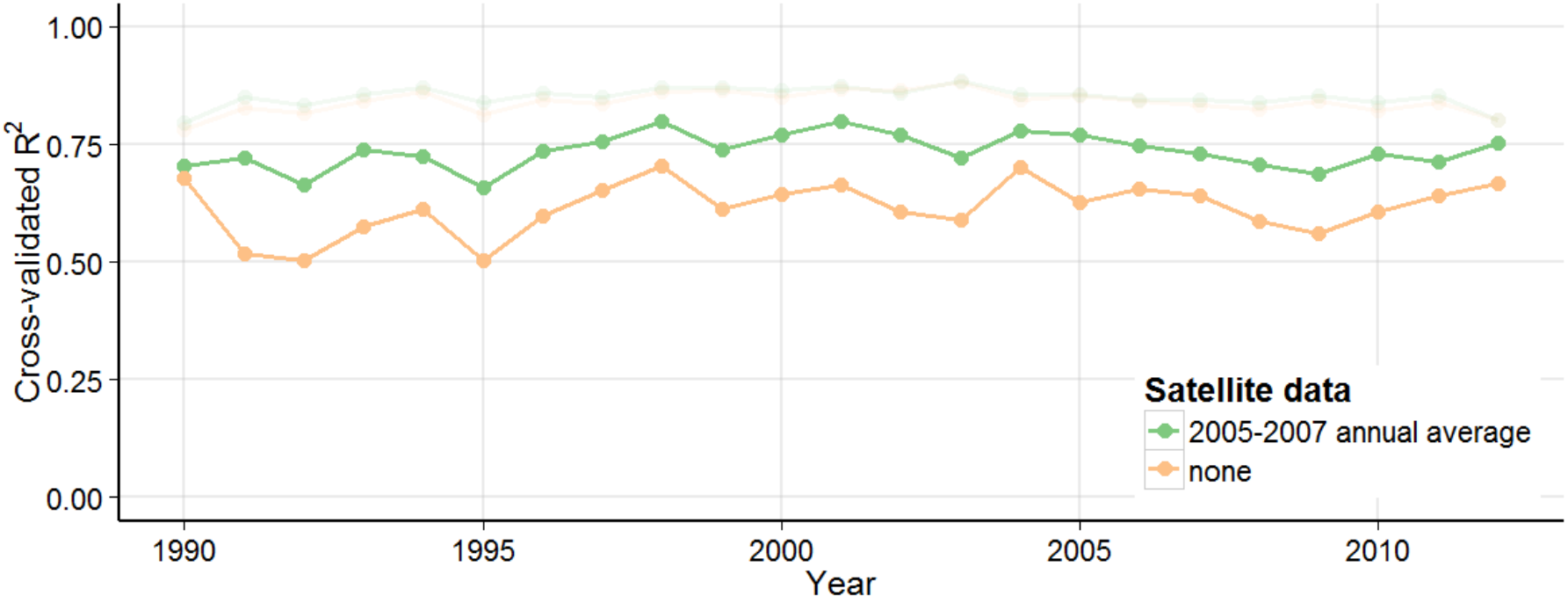
[2]National Satellite-Based Land-Use Regression: NO2 in the United States. Eric V. Novotny, Matthew J. Bechle, Dylan B. Millet, and Julian D. Marshall. *Environmental Science & Technology* 2011 *45* (10), 4407-4414

Typical 20-fold cross-validation



- Model performance was insensitive to inclusion of satellite data using 20-fold cross-validation
- This does not answer the question regarding model performance in areas of poor coverage

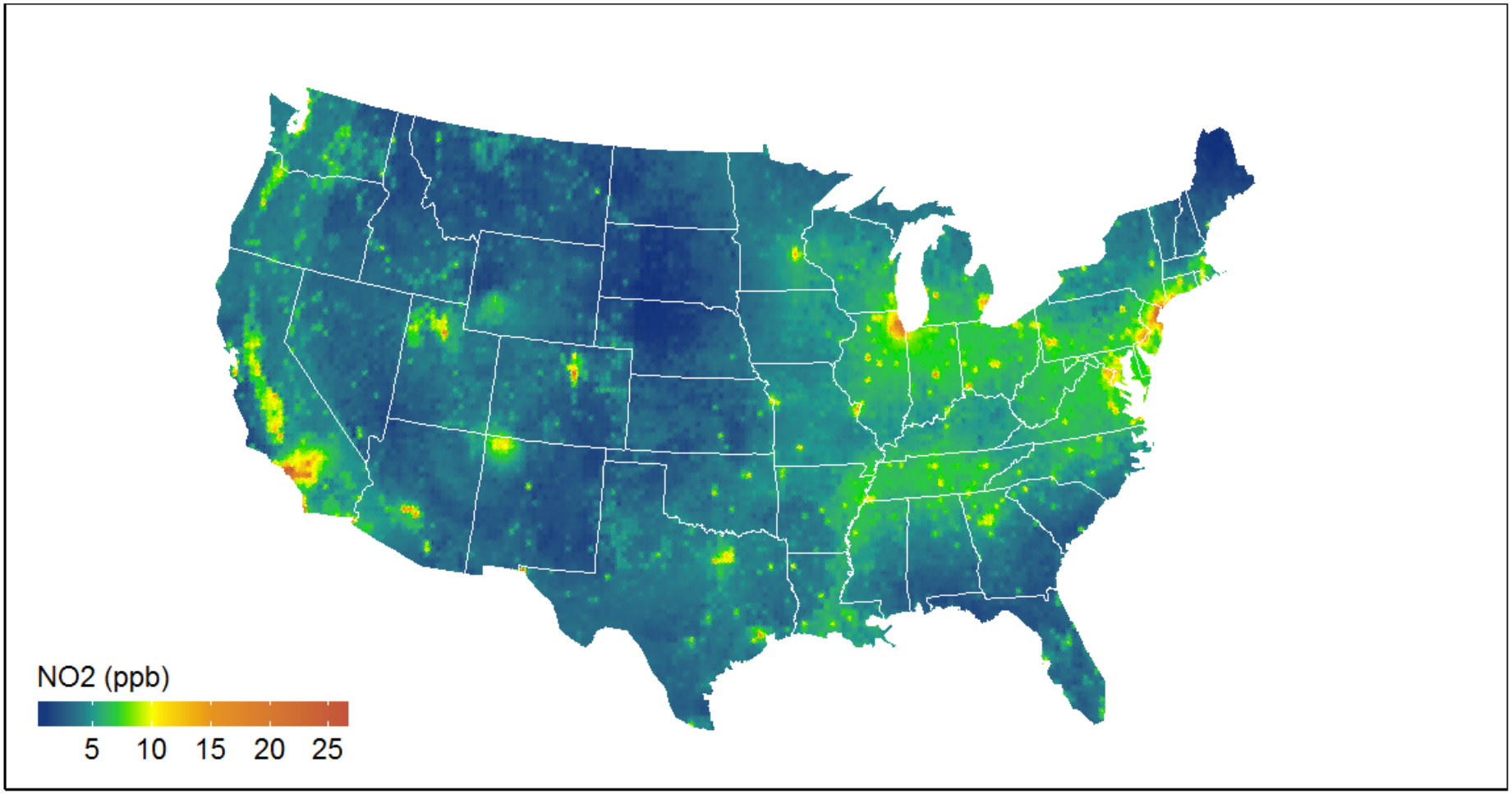
Model Performance: Spatially Clustered Cross-Validation



$R^2 = 1 - MSE / Var(Y)$

- Increases performance consistently across time, despite satellite data being from 2005-2007

NO_2 Predictions on a 25 km grid

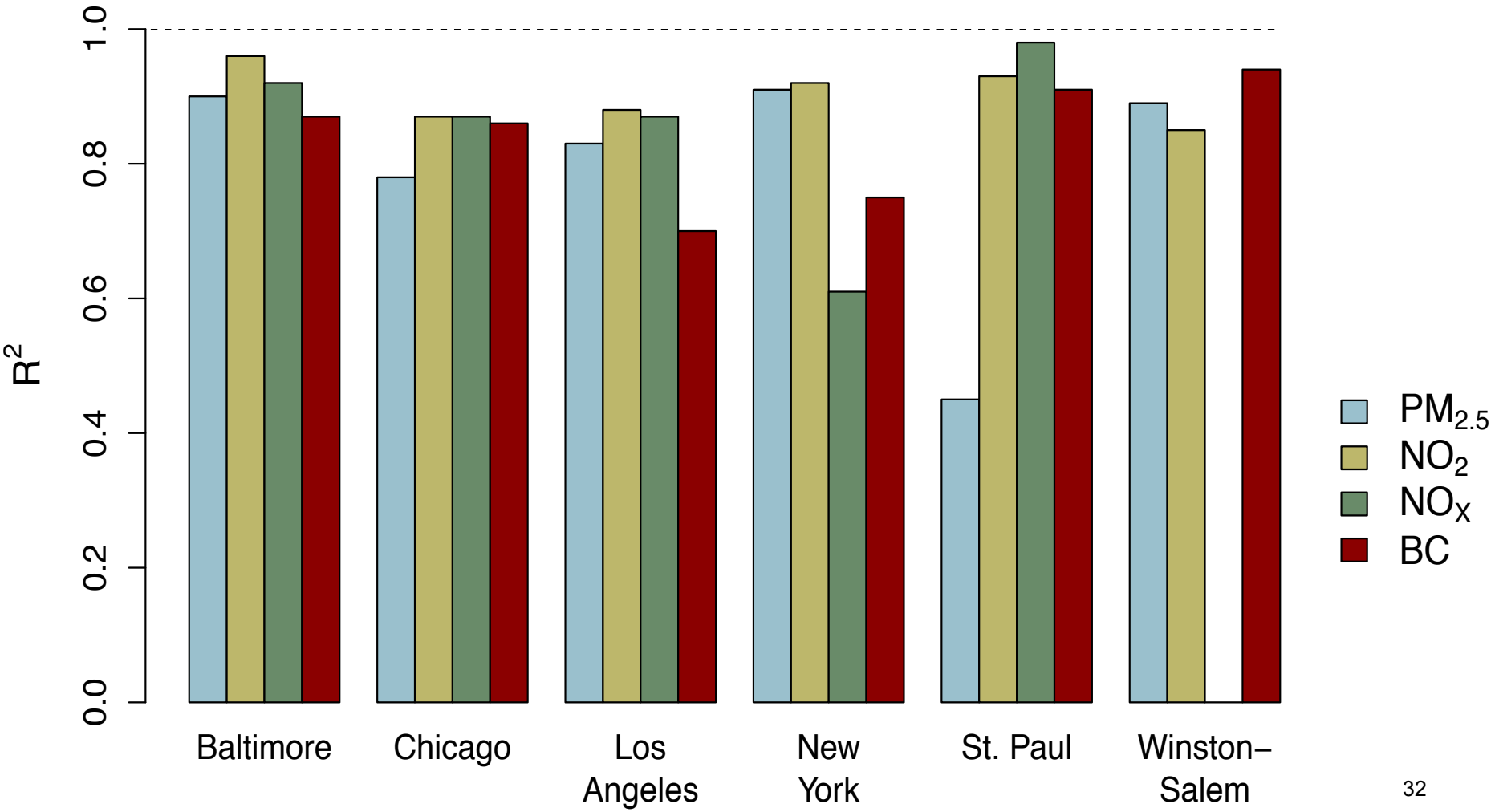


b) Spatio-temporal modeling for MESA Air

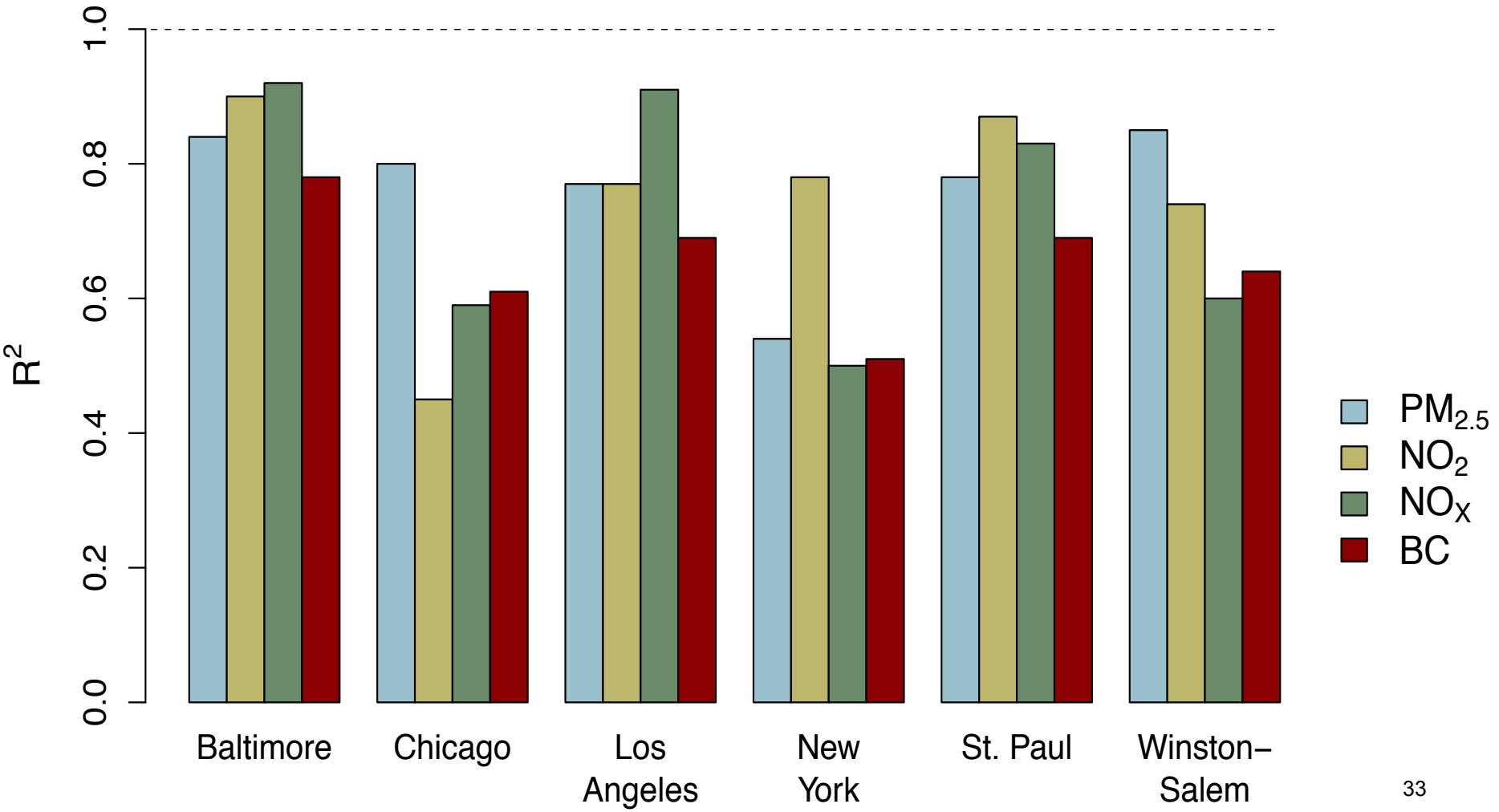
- 2005 – 2009
- $PM_{2.5}$, NO_2, NO_X, BC
- 2-week measurements
- **Fixed** Sites
  - 3 – 7 per city, 1 collocated with AQS site
- **Home** Outdoor Sites
  - 1 – 3 measurements from ~100 participant residence locations in each city
- **Snapshot** Sites (NO_X and NO_2 only)
  - Clusters around roadways

- AQS and Fixed sites
  - Leave-one-out CV
- Home sites
  - 10-fold CV
- Snapshot Sites
  - 10-fold CV by cluster

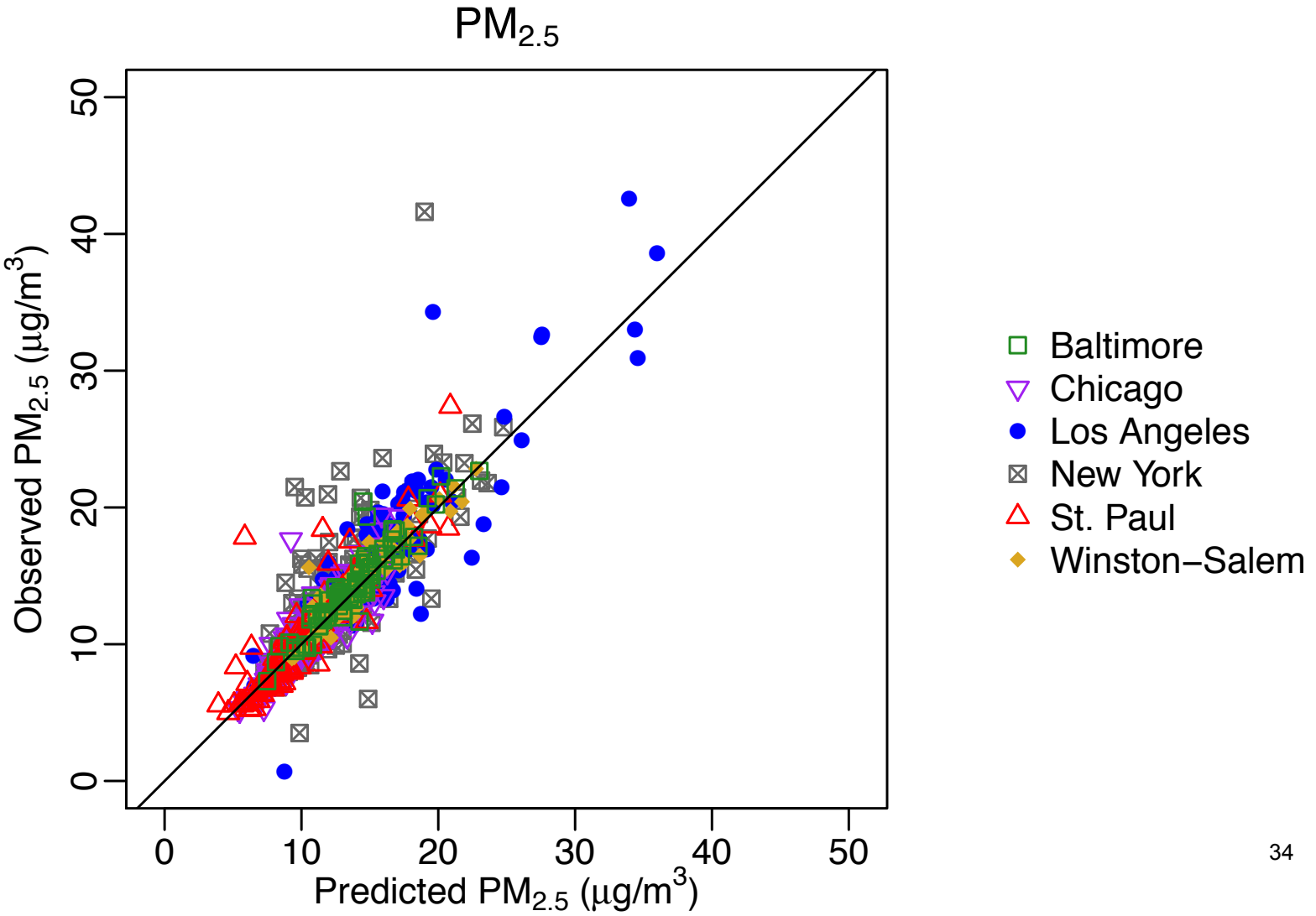
CV Results: AQS/Fixed Site R^2



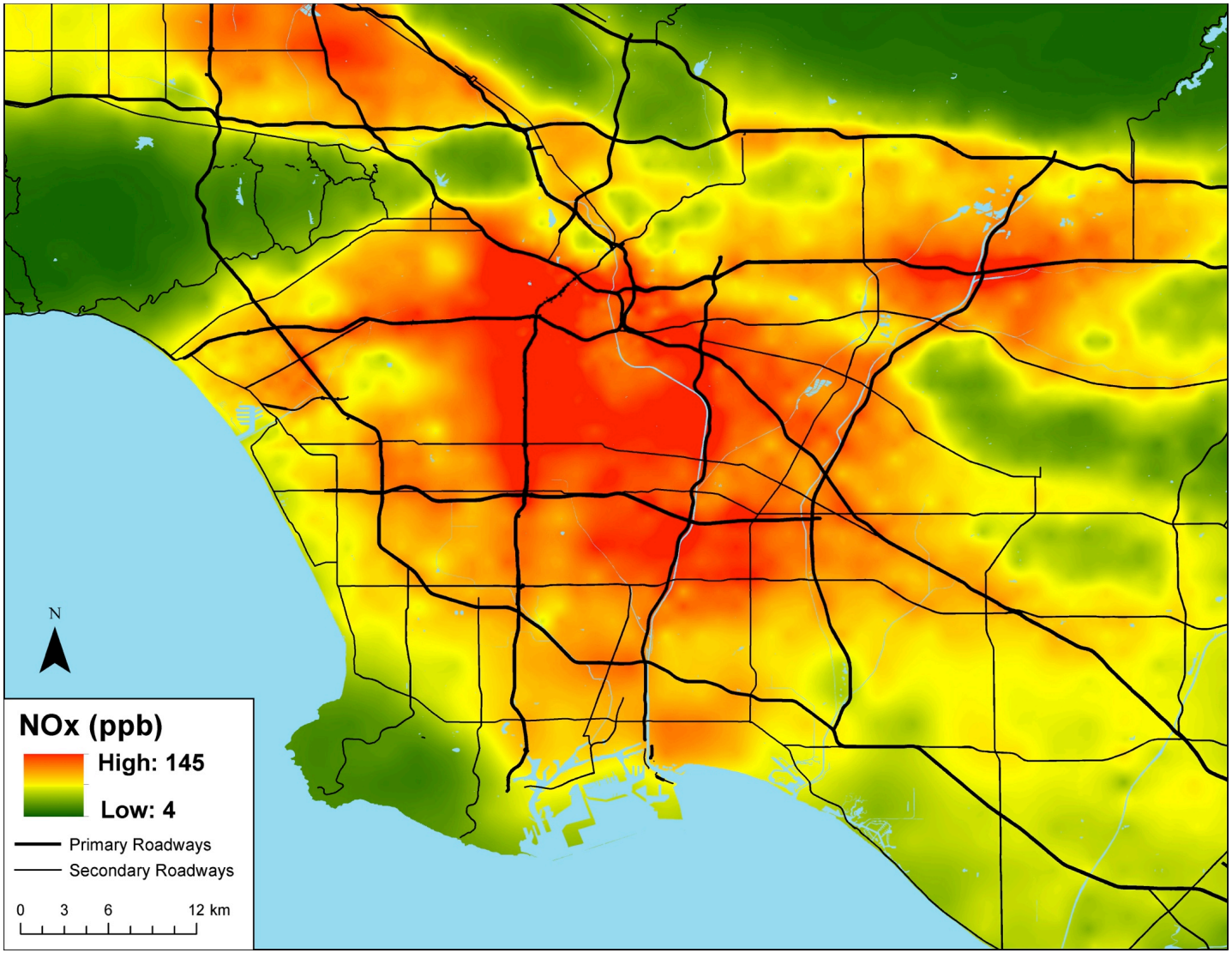
CV Results: Home Site R^2



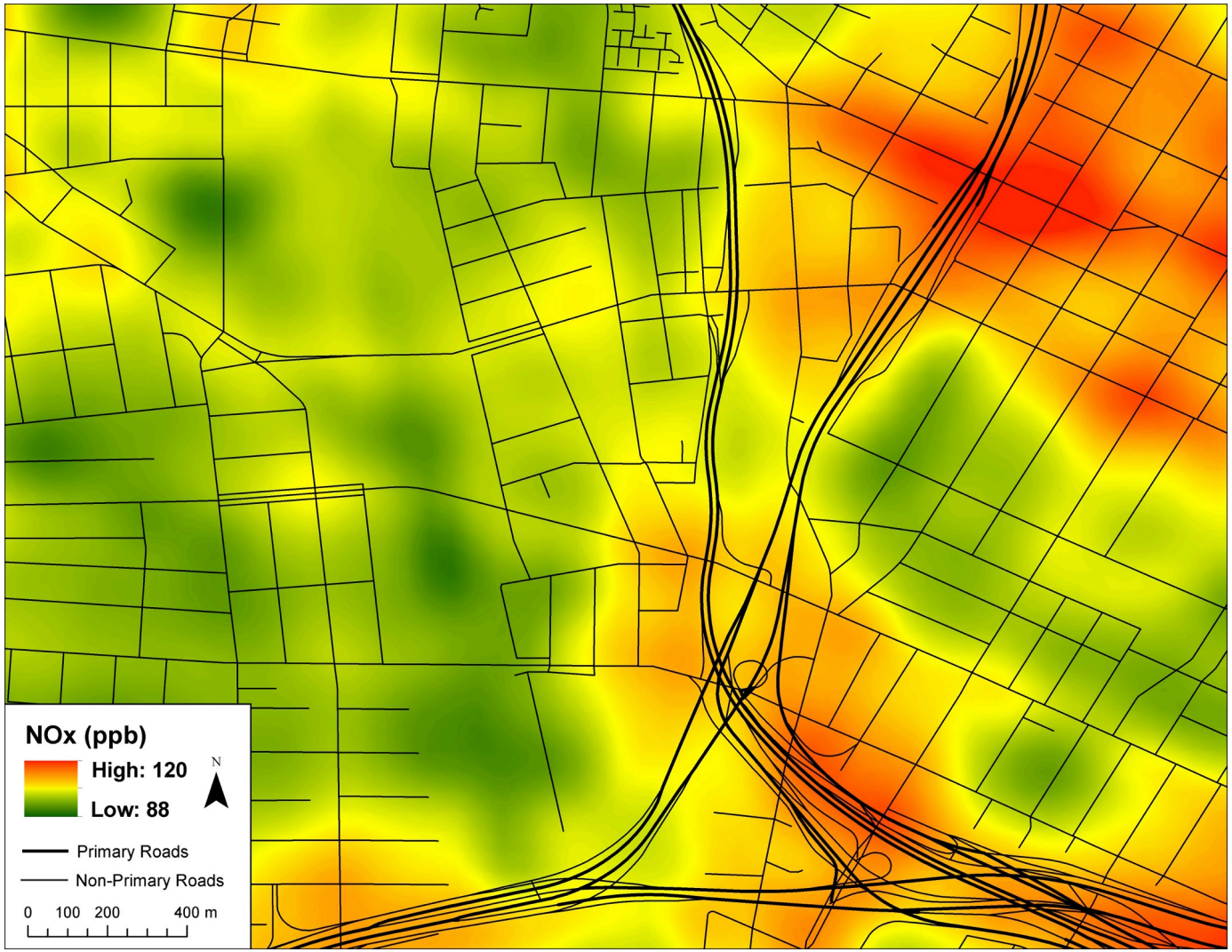
CV Results: Home Sites



NO_X Predictions: 2000 Average



NO_X Predictions: 2000 Average



Summary

- Model fit very good for most cities and pollutants
- Spatiotemporal model
  - Incorporates cohort-specific monitoring data
  - Allows for unbalanced monitoring design
  - Provides predictions at flexible time scales and with fine-scale spatial resolution

c) Historical modeling/prediction of PM2.5

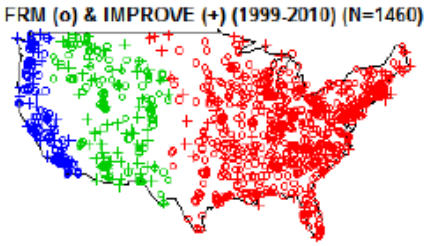
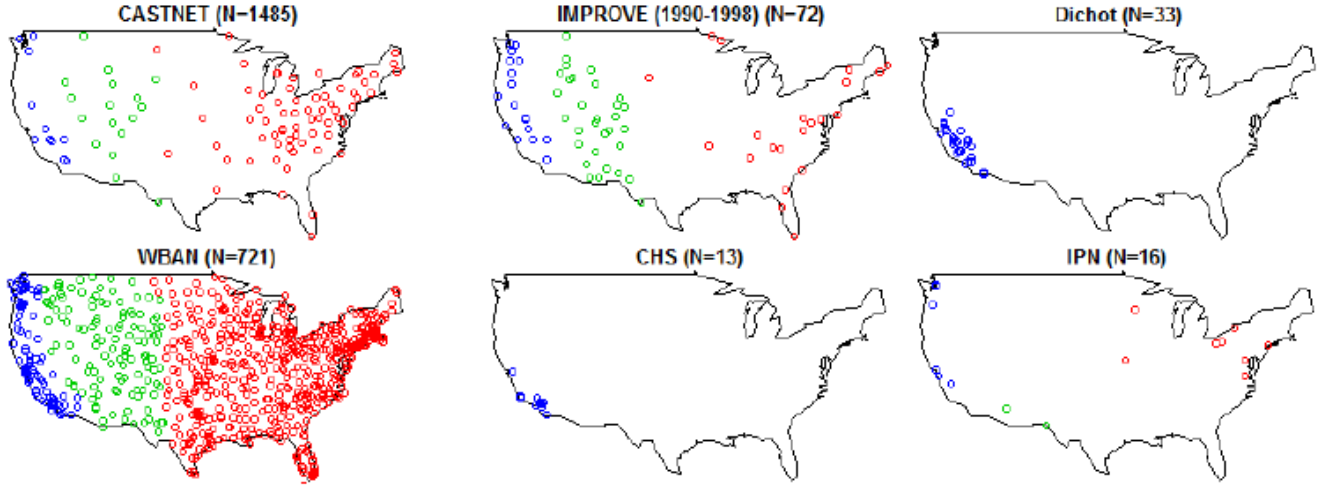


Figure 1. Sites for model development, trend estimation, and model evaluation by three regions (East, West Mountain, and West Coast)



Spatio-temporal modeling on annual means with different strategies to estimate/extrapolate trend basis function back in time.

- **Estimated temporal trends**
  - Trend shapes vary somewhat across the 3 approaches (Figure 2)

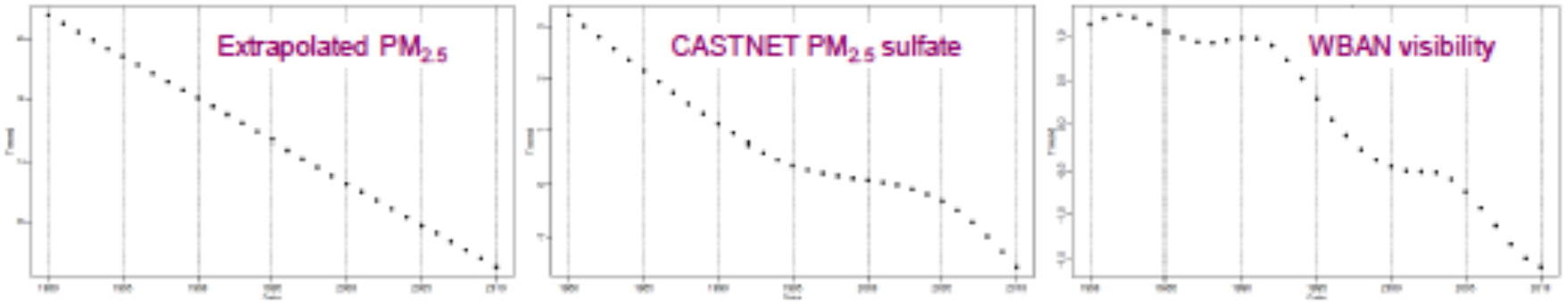


Figure 2. Estimated temporal trends based on $PM_{2.5}$ annual averages in FRM and IMPROVE, $PM_{2.5}$ sulfate annual averages in CASTNET, and visibility annual averages in WBAN

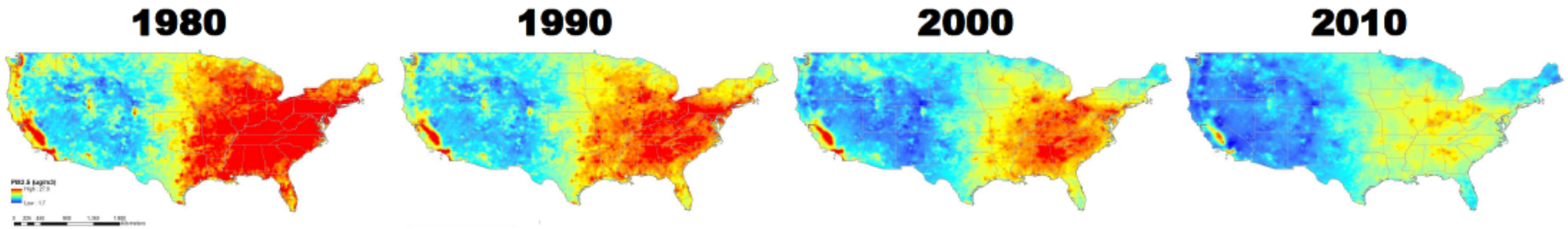
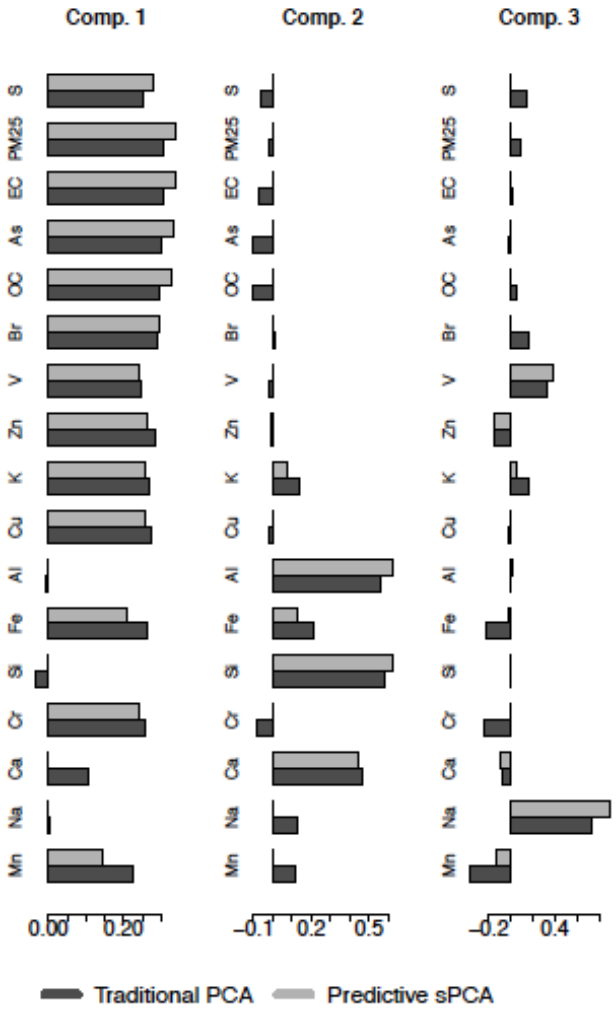


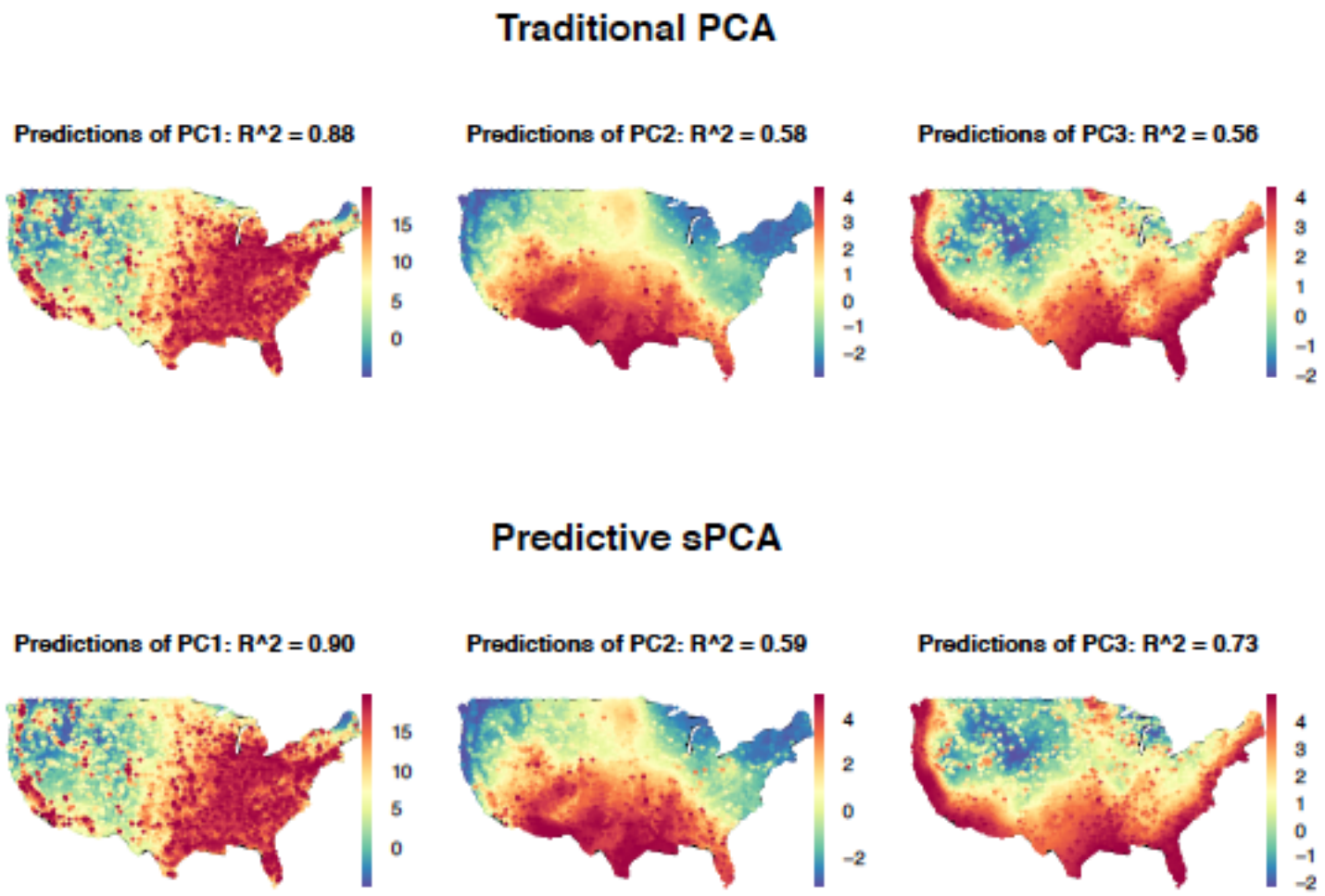
Figure 3. Predictions of $PM_{2.5}$ annual averages in 1980, 1990, 2000, and 2010

d) Spatially predictive PCA (Jandarov) and k-means clustering (Keller)

Application of predictive sPCA to PM2.5 data: Loadings



Application of predictive sPCA to Data: Heat maps



CCAR Biostat Core

- Josh Keller
- Sun-Young Kim
- Johan Lindström
- Casey Olives
- Assaf Oron
- Paul Sampson
- Lianne Sheppard
- Adam Szpiro
- Michael Young
- MESA Air Data Base team

Cross-center collaborations

Other EPA Clean Air Research Centers (CLARCs)

- Emory/Georgia Tech ("SCAPE")
- Harvard ("Harvard")
- Michigan State/Michigan ("GLACIER")

Collaboration specifics

- $50,000 per center per year
- Involves 2 or more CLARCs

UW CLARC collaborations

- Mobile sampling in Atlanta (with Emory) – Tim Larson, Mike Yost

- Toxicology (Michigan State) – Matt Campen

- Exposure measurement error correction (with Harvard and Emory) – Adam Szpiro

- Satellite (remote sensing) data for $PM_{2.5}$ (with Emory and Harvard) – Paul Sampson

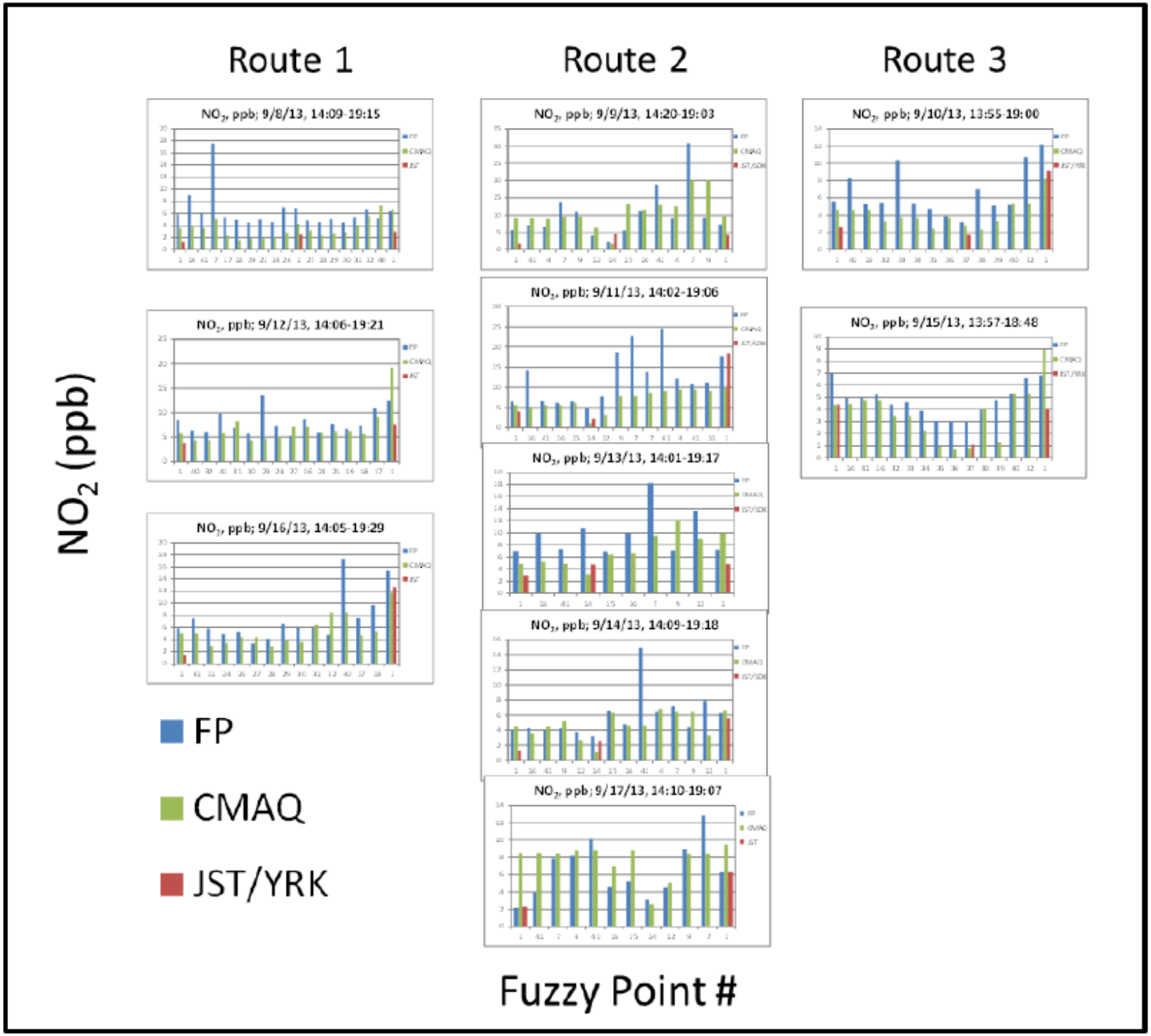
Mobile & fixed site characterization in Atlanta

- With Emory (SCAPE) – Sept 2013 x 2 weeks

- 2-week sampling at central site and 28 fuzzy points x 3 routes + passive (badge) sampling, incl 2 roadside gradient routes

- Completed comparison to CMAQ predictions (4 km grid, downscaled to 250 m using LUR model) + validation using 2 monitoring sites in Atlanta
  - NO_2 and nephelometry good agreement, BC fair

- Airplane emissions, esp. UFP spatial distribution


Route 1
Route 2
Route 3
NO2 (ppb)
NO2, ppb; 9/8/13, 14:09-19:15
NO2, ppb; 9/9/13, 14:20-19:03
NO2, ppb; 9/10/13, 13:55-19:00
NO2, ppb; 9/11/13, 14:02-19:06
NO2, ppb; 9/12/13, 14:06-19:21
NO2, ppb; 9/13/13, 14:01-19:17
NO2, ppb; 9/15/13, 13:57-18:48
NO2, ppb; 9/14/13, 14:09-19:18
NO2, ppb; 9/16/13, 14:05-19:29
NO2, ppb; 9/17/13, 14:10-19:07
FP
CMAQ
JST/YRK
Fuzzy Point #

Animal toxicology

1. Michigan State (GLACIER) coarse PM human experimental subject serum and BAL samples (R Brook, U Michigan)
   - Campen ex vivo endothelial cell & inflammatory potential assays

2. Campen collaboration with Jesus Araujo (UCLA & GLACIER)
   - HDL dysfunction & oxidized lipids on serum and BAL samples from LRRI mouse studies

Exposure measurement error correction

- With Harvard and Emory
- Georgia birth cohort (low birth weight) and EPA $PM_{2.5}$
- Common $PM_{2.5}$ exposure predictions based on UW spatio-temporal model for Atlanta and health effects analysis completed
- 3 statistical approaches for measurement error correction (with simulations for insight into differences):
  - parameter bootstrap - UW
  - simulation extrapolation - Harvard
  - Bayesian – Emory

Satellite $PM_{2.5}$ estimation (with Emory and Harvard)

- Standard set of data for North Carolina, 2006-08, MODIS AOD data downloaded for 10km grid

- common UW geographic database

- 6 candidate models for $PM_{2.5}$ prediction in progress
  - Harvard x 2 (mixed effects, multi-level)
  - Emory x 3 (spatial downscaler, mixed effects, CMAQ)
  - UW x 1 (spatio-temporal model)
    - assess added value of satellite data

- commons metrics for model evaluation