

Enabling Spatiotemporal Analysis and Visualization of Air Pollution in China and India

Merrick Lex Berman, Devika Kakkar, Wendy Guan & Fei Carnes
Center for Geographic Analysis
Harvard University

The 25th International Conference on Geoinformatics
Buffalo, NY
August 2017

All News Images Videos Maps More

Settings Tools

View saved SafeSearch

- china city
- china pollution
- china fact
- polluted
- smog
- china coal
- china statistics
- china chart
- skyline
- bad air
- beijing today
- fog
- beijing
- xi an
- shanghai
- harbin
- wuhan
- guangzhou
- huai river
- air particle



diwali
smog
industrial air
polluted
outdoor
environmental
contaminated
increase
transportation
new york
village
delhi case study
worldwide
household farming chemical
clean
mumbai
new delhi
hyderabad



TOXIC AIR		WORST PERFORMERS		
EPA's ranking for air quality effect on health				
Rank	Country	Score	Country	
120	China	1.7	123	Latvia
121	France	1.7	124	Bosnia and Herzegovina
122	Spain	1.7	125	India
123	Germany	1.7	126	Kuwait
124	Nepal	1.7	127	Yemen
125	India	1.7	128	South Africa
			129	Kazakhstan

Objectives

- Develop a free archive of ground-based air pollution data collected from real-time monitoring stations in China and India.
- Develop a methodology for processing air pollution measurements from streaming raw data into pollution exposure maps.
- Provide the spatiotemporal data and maps for public health and other related research.

China Air Quality Data

- Starting in mid-2013, some websites began posting China air quality monitoring data
 - <http://aqicn.org>
 - <http://106.37.208.233:20035/> (National Urban Air Monitoring Platform)
 - <http://www.pm25s.com/> (later)
- <http://aqicn.org> initially included more than 900 ground-based monitoring stations across China (official and volunteers)
- Each monitoring station reports its location, 6 air quality measurements and 5 weather variables.
- Data covers the past 48 hours including the current hour.
- However, it did not provide access to past data beyond 48 hours.

Beijing Air Pollution: Real-time Air Quality Index (AQI)

- BEIJING
北京
- SHANGHAI
上海
- GUANGZHOU
广州
- CHENGDU
成都
- HONGKONG
香港
- NANJING
南京
- LOCATE THE NEAREST CITY
- SEARCH FOR YOUR CITY

Yongledianzhen, Tongzhou AQI: Yongledianzhen, Tongzhou Re

153

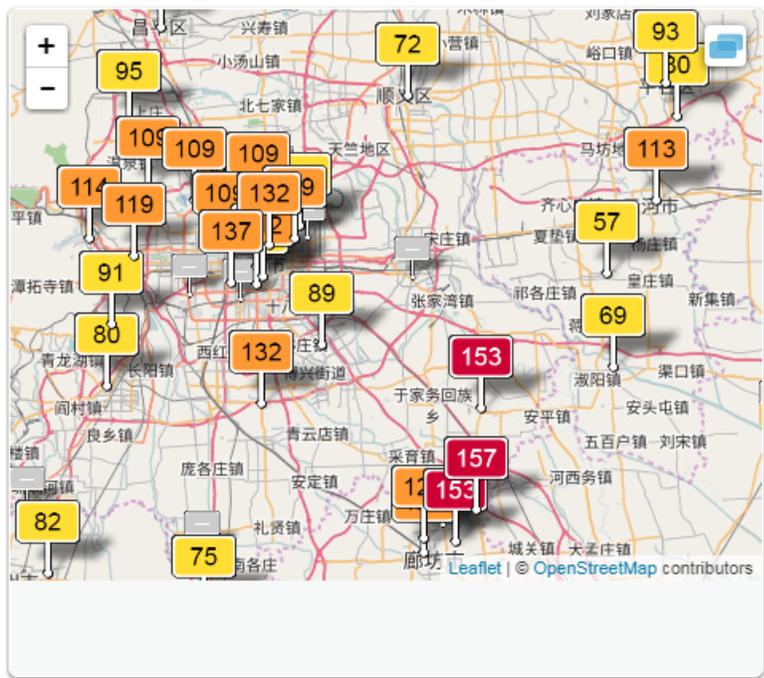
Unhealthy

Updated on Sunday 1:00

Temp.: 24°C

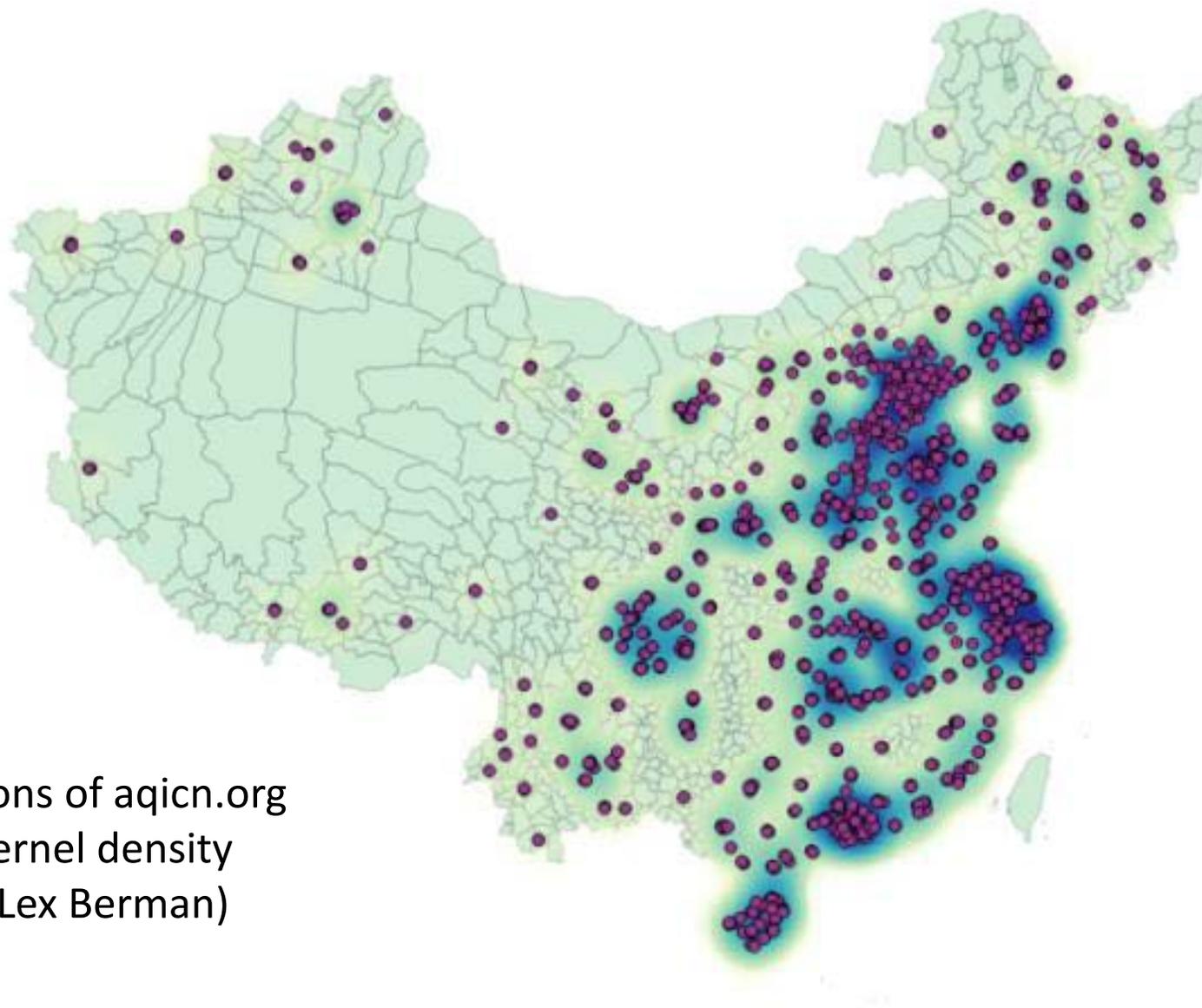
Current Past 48 hours data

Parameter	Current	Min	Max
PM2.5 AQI	153	21	153
PM10 AQI	37	37	77
O3 AQI	1	1	46
NO2 AQI	30	8	34
SO2 AQI	1	1	6
CO AQI	10	1	12
Temp.	24	22	27
Dew	22	19	25
Pressure	1008	1002	1011
Humidity	89	70	100
Wind	2	1	6



AIR QUALITY FORECAST





Monitoring Stations of aqicn.org
[2014-2015] at kernel density
200km (map by: Lex Berman)

Data Dictionary for China Air Quality Data

Variable Name	Description and Unit
LocatID	Location ID generated by unique latitude-longitude pairs for each monitor location
StationName	Station Name as defined by the monitor owner (might change, might not be unique)
ChName	Chinese language station name (UTF8 encoding)
Latitude	Latitude North in decimal degrees WGS84
Longitude	Longitude East in decimal degrees WGS84
PM2.5	Particulate Matter 2.5 micron diameter $\mu\text{g}/\text{m}^3$ (micrograms per cubic metre)
PM10	Particulate Matter 10 micron diameter $\mu\text{g}/\text{m}^3$ (micrograms per cubic metre)
O3	Ozone pphm (parts per hundred million)
NO2	NOX NitrogenDioxide pphm (parts per hundred million)
SO2	SOX SulfurDioxide pphm (parts per hundred million)
CO	CarbonMonoxide ppm (parts per million)
Temperature	degrees Celsius
DewPoint	degrees Celsius
Pressure	millibars
Humidity	absolute humidity in grams/meter ³
Wind	km / hour
UMT_time	data collection time Greenwich Mean Time

Building the China AQI Archive

- Developed a script to “scrape” these websites automatically
- Launched in January 2014, downloading data once every 8 hours
- In two years we archived about 2.3 million observations from over 2,000 monitoring stations across China
- Since September 2015 our data scraping changed to hourly from <http://www.pm25s.com/>, with over 1,500 monitoring stations
- Archived data is publically downloadable from Harvard’s Dataverse https://dataverse.harvard.edu/dataverse/cga_aqi

Dataset is collected three times / day

Posted for public use:
<http://aqi.cga.harvard.edu/china/>

Index of /china

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 Parent Directory		-	
 20140501010000.tar	01-May-2014 01:12	68K	
 20140501090000.tar	01-May-2014 09:14	68K	
 20140501170000.tar	01-May-2014 17:15	68K	
 20140502010000.tar	02-May-2014 01:15	68K	
 20140502090000.tar	02-May-2014 09:09	68K	
 20140502170000.tar	02-May-2014 17:21	68K	
 20140503010000.tar	03-May-2014 01:12	68K	
 20140503090000.tar	03-May-2014 09:14	68K	
 20140503170000.tar	03-May-2014 17:16	68K	
 20140504010000.tar	04-May-2014 01:11	68K	
 20140504090000.tar	04-May-2014 09:20	68K	
 20140504170000.tar	04-May-2014 17:19	68K	
 20140505010000.tar	05-May-2014 01:15	68K	
 20140505090000.tar	05-May-2014 09:17	68K	
 20140505170000.tar	05-May-2014 17:17	68K	
 20140506010000.tar	06-May-2014 01:08	68K	

Other China AQI Archives

- Berkeley Earth
 - <http://berkeleyearth.org/air-pollution-overview/>
- Beijing City Lab
 - <https://www.beijingcitylab.com/projects-1/13-pm2-5/>
- 中国环境监测总站全国城市空气质量实时发布平台
 - <http://beijingair.sinaapp.com/>

India Air Quality Data

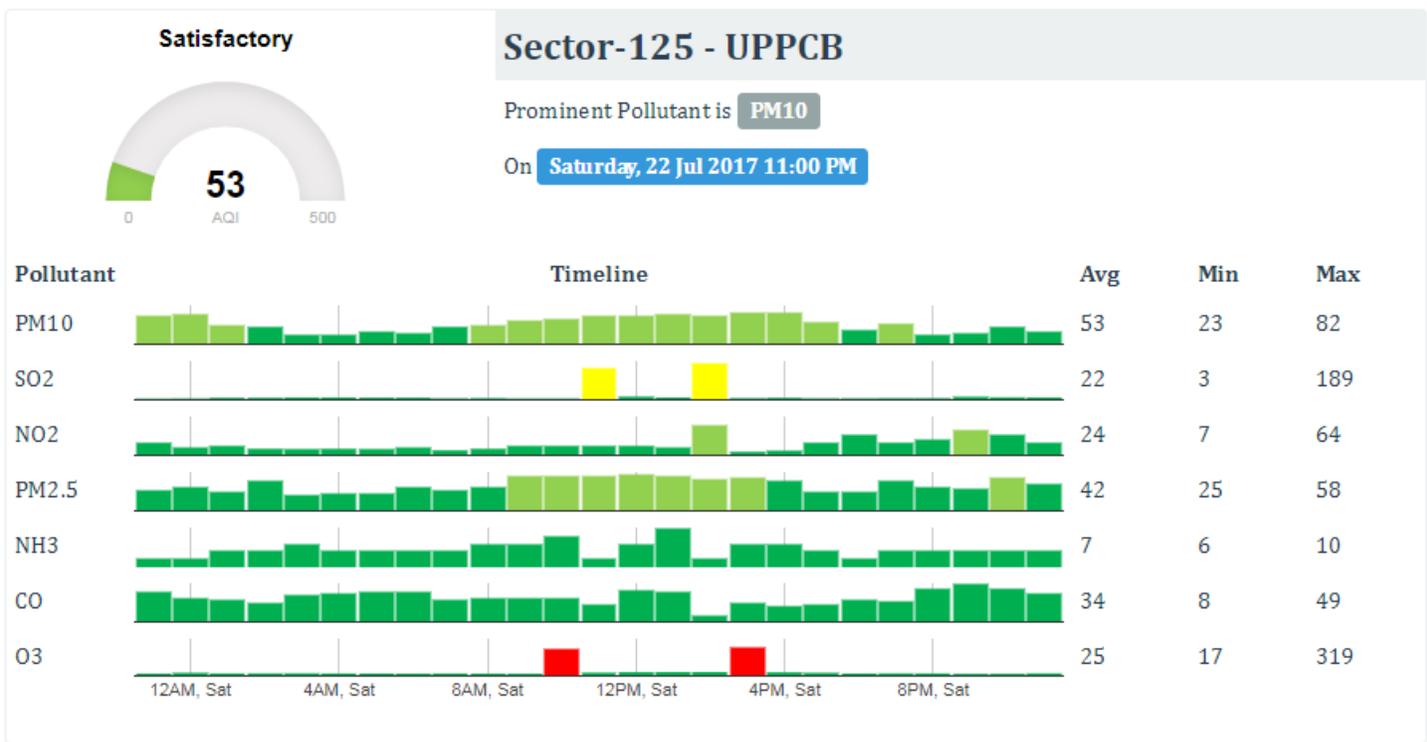
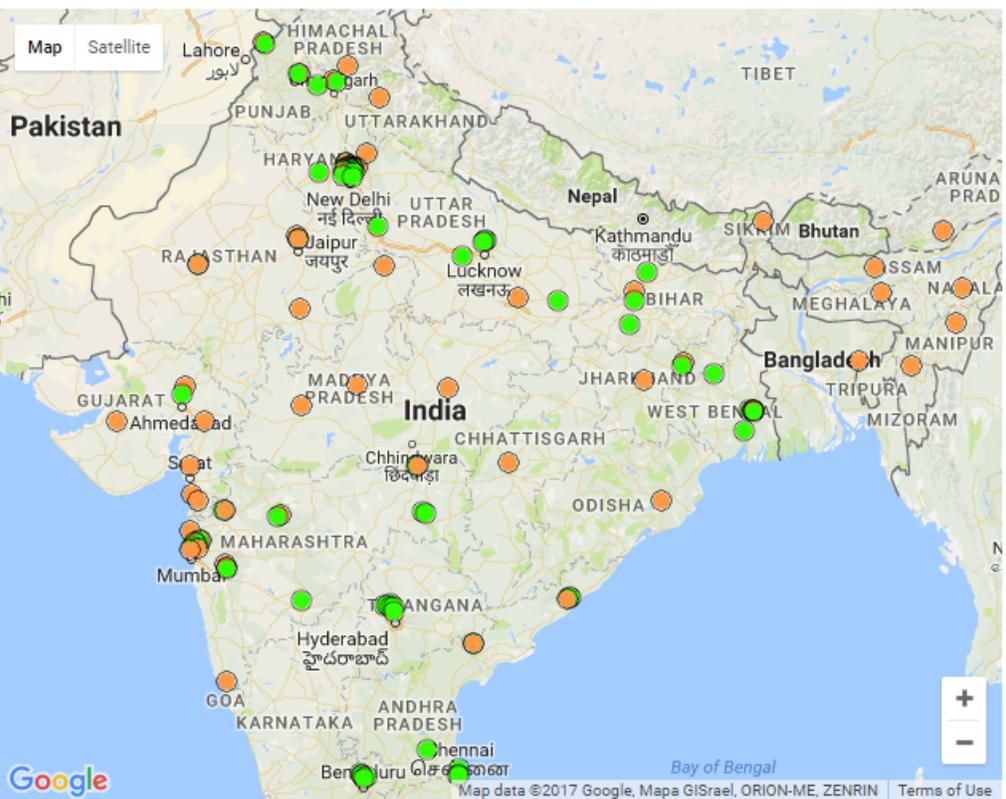
- National Air Quality Index <http://164.100.160.234:9000/> launched in April 2015
- Managed by the Central Pollution Control Board, Ministry of Environment, Forests and Climate Change
- Includes about 50 monitoring stations across India
- Reports 7 air quality measurements
- Shows the past 24 hours till the current hour

National Air Quality Index



Central Pollution Control Board,
Ministry of Environment, Forests and Climate Change

State: **Uttar Pradesh** |
 City: **NOIDA** |
 Station: **Sector-125 - UPPCB** |
 Date: **22/07/2017**



AQI	Remark	Color Code	Possible Health Impacts
0-50	Good		Minimal impact
51-100	Satisfactory		Minor breathing discomfort to sensitive people
101-200	Moderate		Breathing discomfort to the people with lungs, asthma and heart diseases

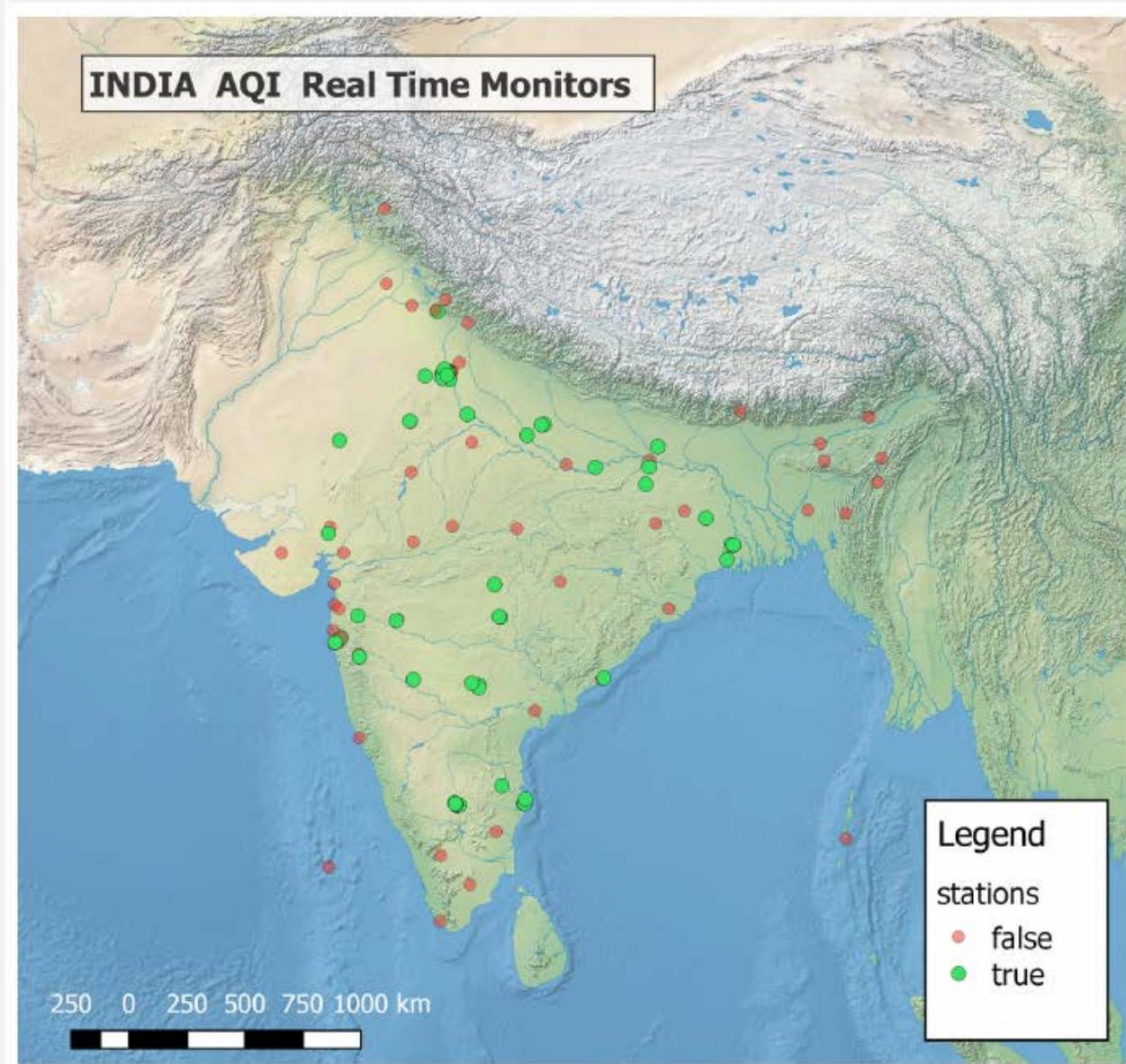
India Air Quality Index Variables

Variable Name	Description	AQI	Remark	Possible Health Impacts
Count	Serial number	0-50	Good	Minimal impact
Date	date of recording	51-100	Satisfactory	Minor breathing discomfort to sensitive people
Hour	hour of the day			
ID	Station ID	101-200	Moderate	Breathing discomfort to the people with lungs, asthma and heart diseases
Name	Station name			
PM10	Particulate Matter 10 micron diameter			
SO2	SO2	201-300	Poor	Breathing discomfort to most people on prolonged exposure
NO2	NO2			
PM25	Particulate Matter 2.5 micron diameter	301-400	Very Poor	Respiratory illness on prolonged exposure
CO	CO	401-500	Severe	Affects healthy people and seriously impacts those with existing diseases
O3	O3			
Status	Reporting or Not Reporting			

India Air Quality Index Data Archives

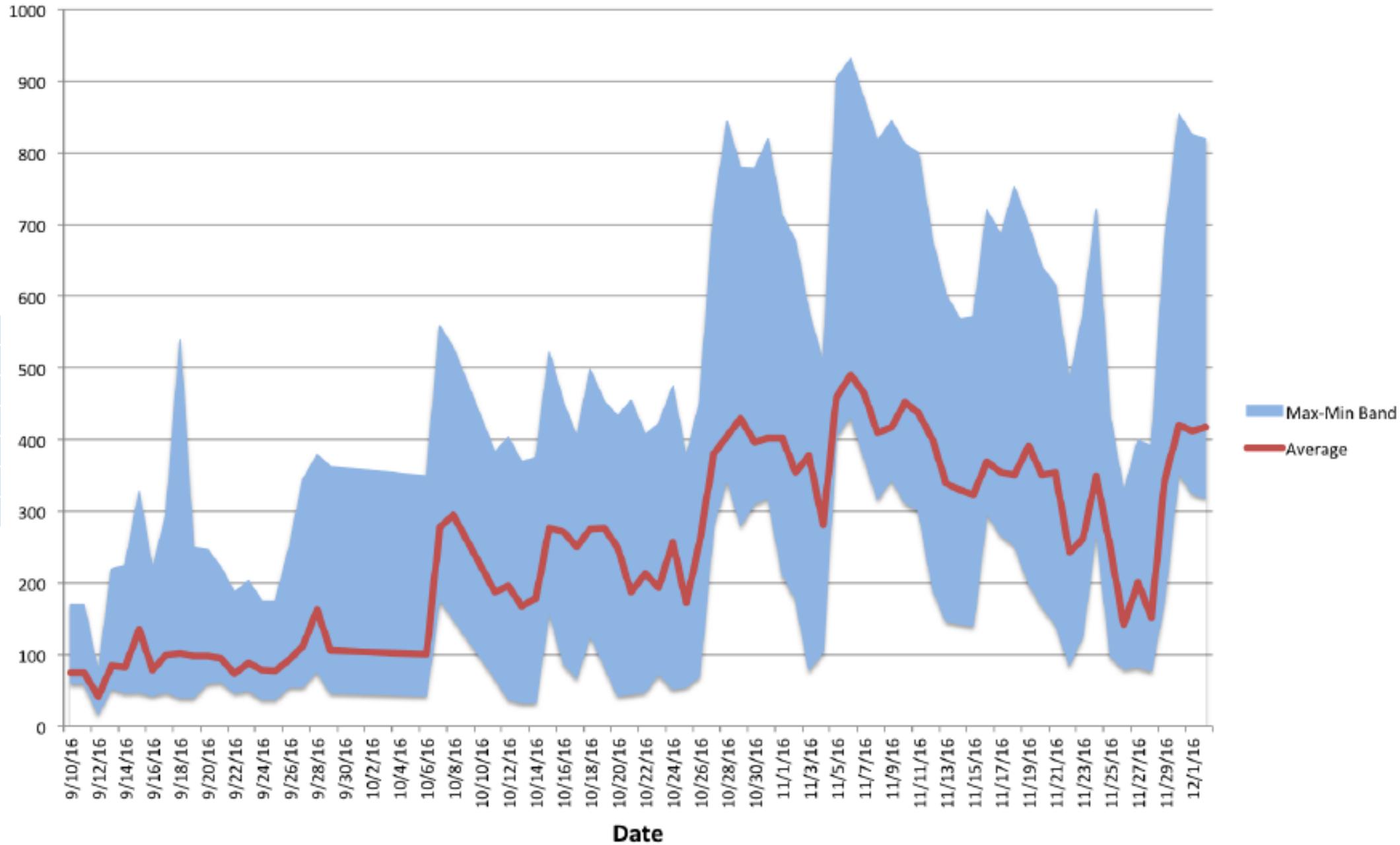
- Developed a script to “scrape” the National Air Quality Index website
- Launched in 2016, downloading data hourly
- Also gets synchronous meteorological information from OpenWeatherMap <http://openweathermap.org/>
- This archive is not available to the public at this time, but will be in the future.

India AQI monitoring locations



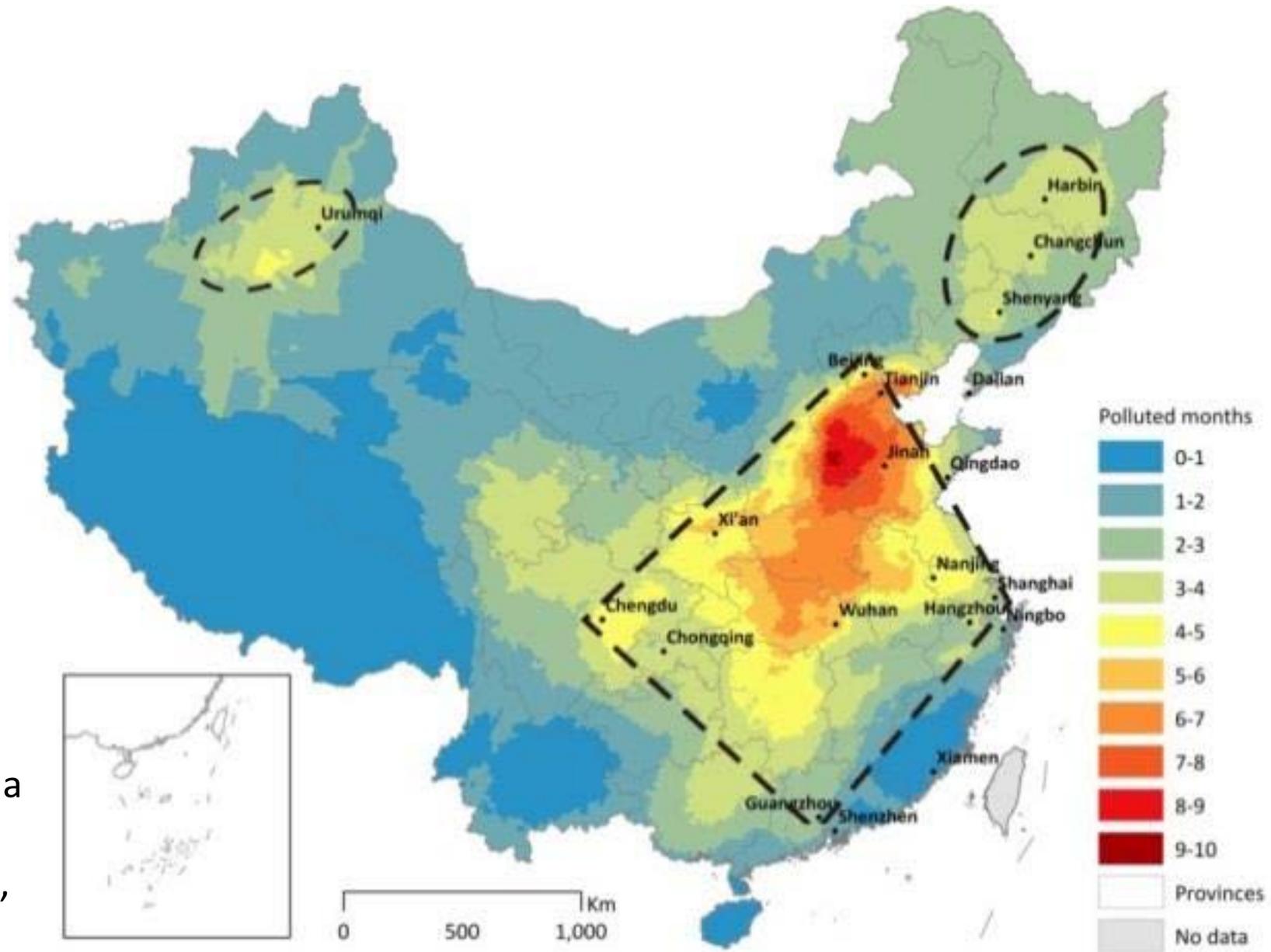
PM2.5- Delhi College of Engineering, New Delhi

AQI	Remark
0-50	Good
51-100	Satisfactory
101-200	Moderate
201-300	Poor
301-400	Very Poor
401-500	Severe



Initial Data Visualization

- We applied a universal kriging at country level in China
- It created an interpolated surface of PM2.5 concentration at a series of time steps.
- The result was an animated map, effective in visual communication <https://youtu.be/CRrrPNEdINE>.
- However, it exposed problems in the overly generalized interpolation model due to the extremely large range of distances between monitoring stations.
- Further spatial-statistical analysis is needed for the evaluation of spatial autocorrelation and regionalization patterns.

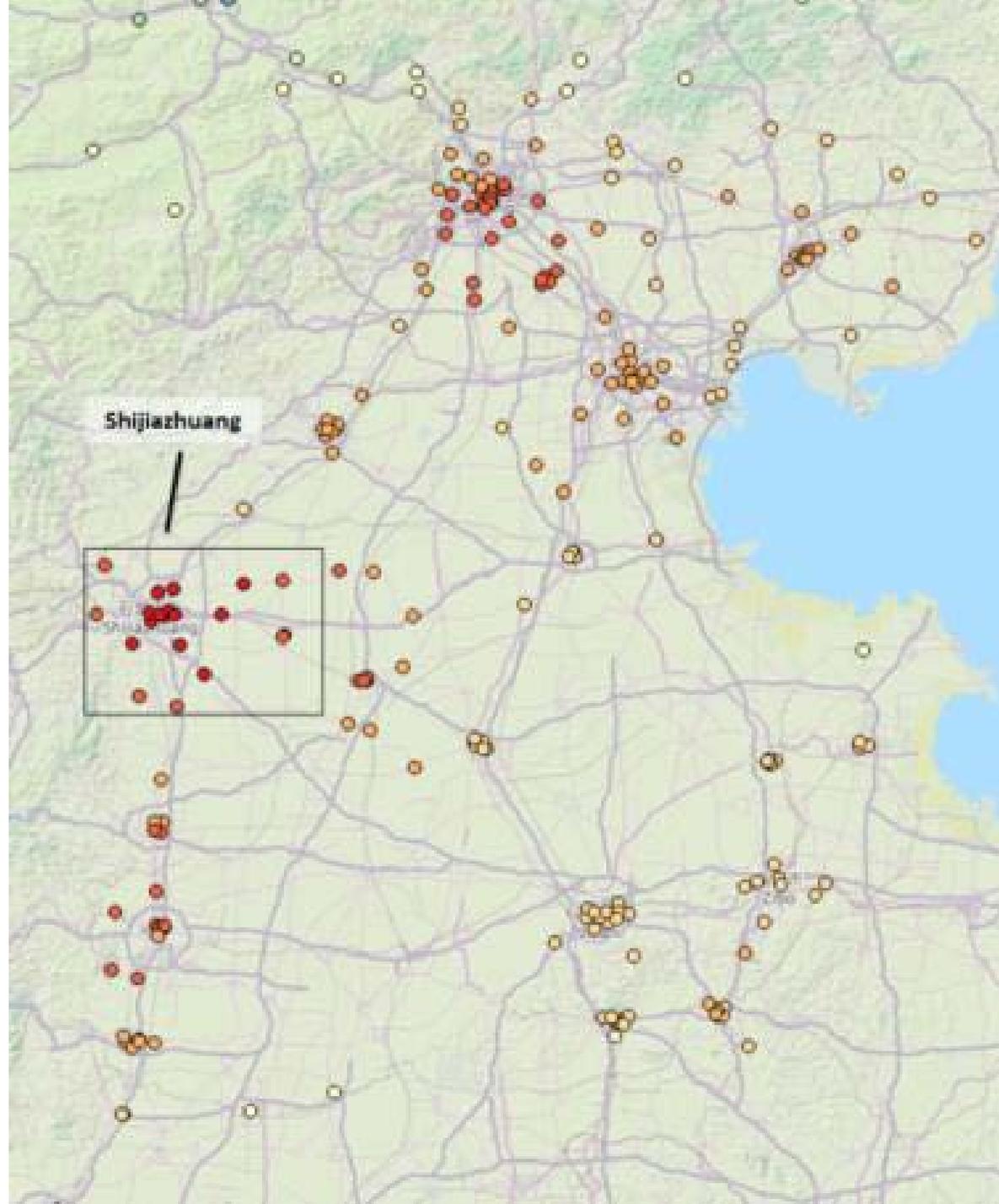


Polluted months in China
 [Apr 2013 – Apr 2014]
 (map by: Kang Wu, et al,
 Beijing City Lab)

Evaluation of Spatial Autocorrelation and Regionalization Patterns

- The approach is to assess hot spots of pollutants using a multi-scaled statistical process
- A geographically weighted regression [GWR] model is applied to estimate local effects of several independent variables, and to interpolate surface grids for the coefficients of those variables.
- These surface grid maps will reveal spatial variation in relation to the pollution variables.
- This approach accounts for volatility in the variations of AQI values over time, and find “break points” as distances within which regionalization patterns and clustering can be detected.

Regionalization based on
standard normal
deviation scales (map by:
Rene Westerholt)

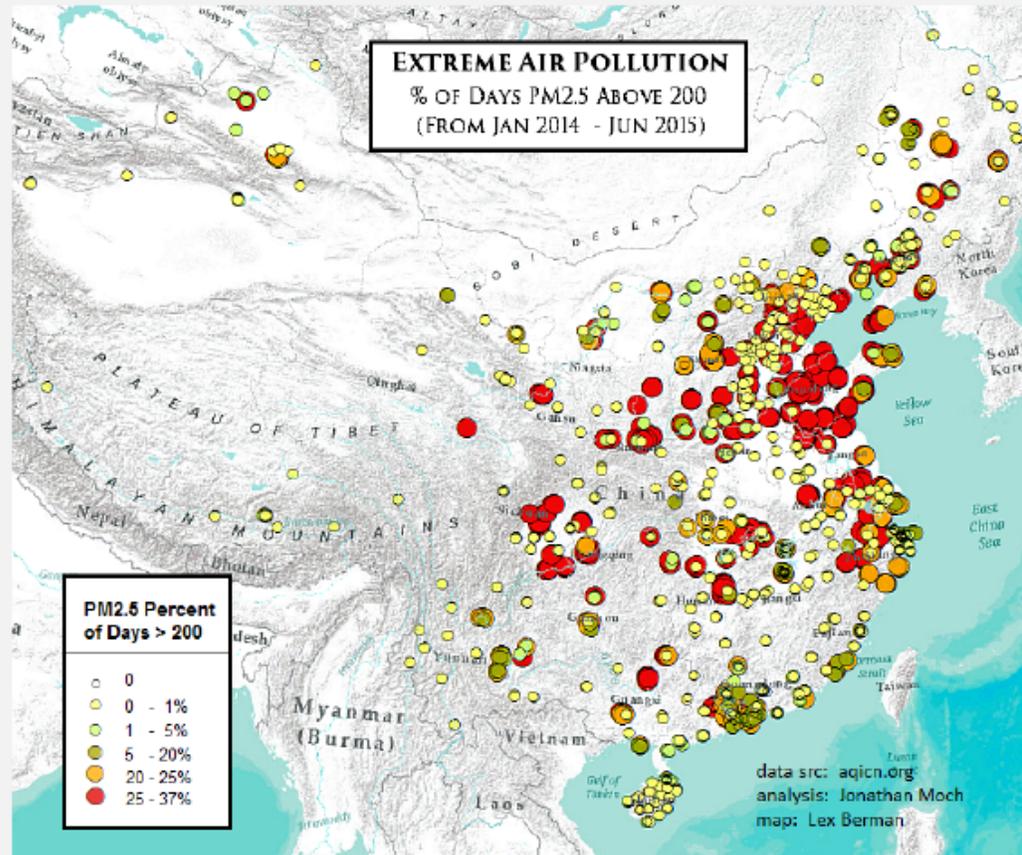


Results

- The AQI data harvesting tools capable of
 - integrating multiple web sources,
 - selecting or aggregating multiple times as harvesting target, and
 - processing raw data into standard archiving files.
- The published AQI archive for China (and will be for India too) on the Dataverse Network
 - support other researcher's work
- The analytical model for
 - detecting regionalization from the AQI data, and
 - interpolating from the monitoring point measurements to a reliable air quality field at various scales.
 - this model is still under development.

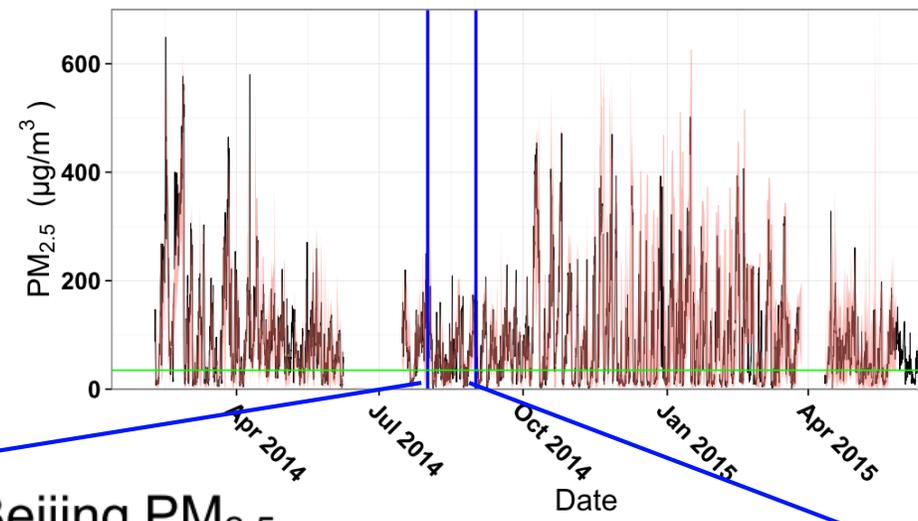
Statistical analysis of archived readings

Jonathan Moch (EPS PhD Student)

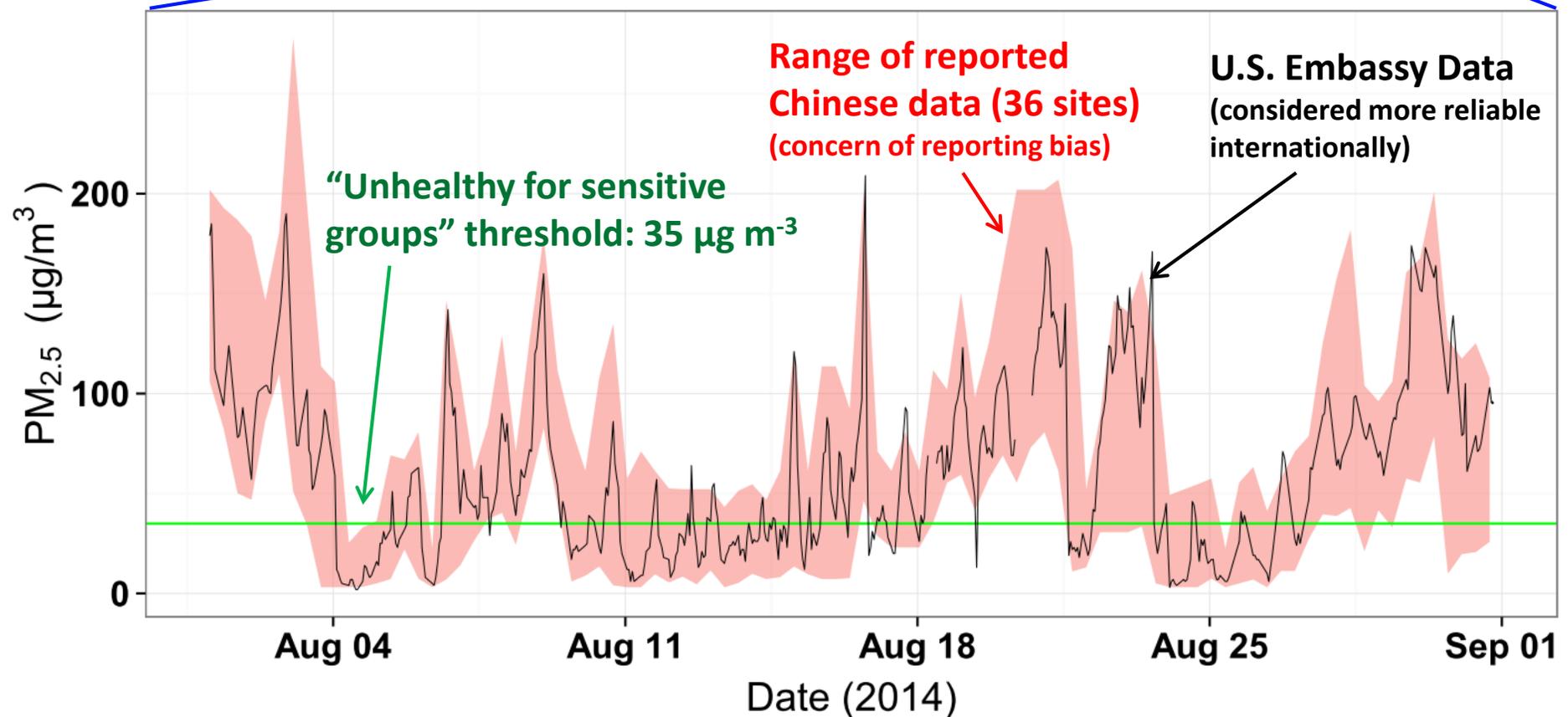


Jan 2014 to May 2015 1.5 million observations

We find that Chinese government and US Embassy measurements in Beijing agree

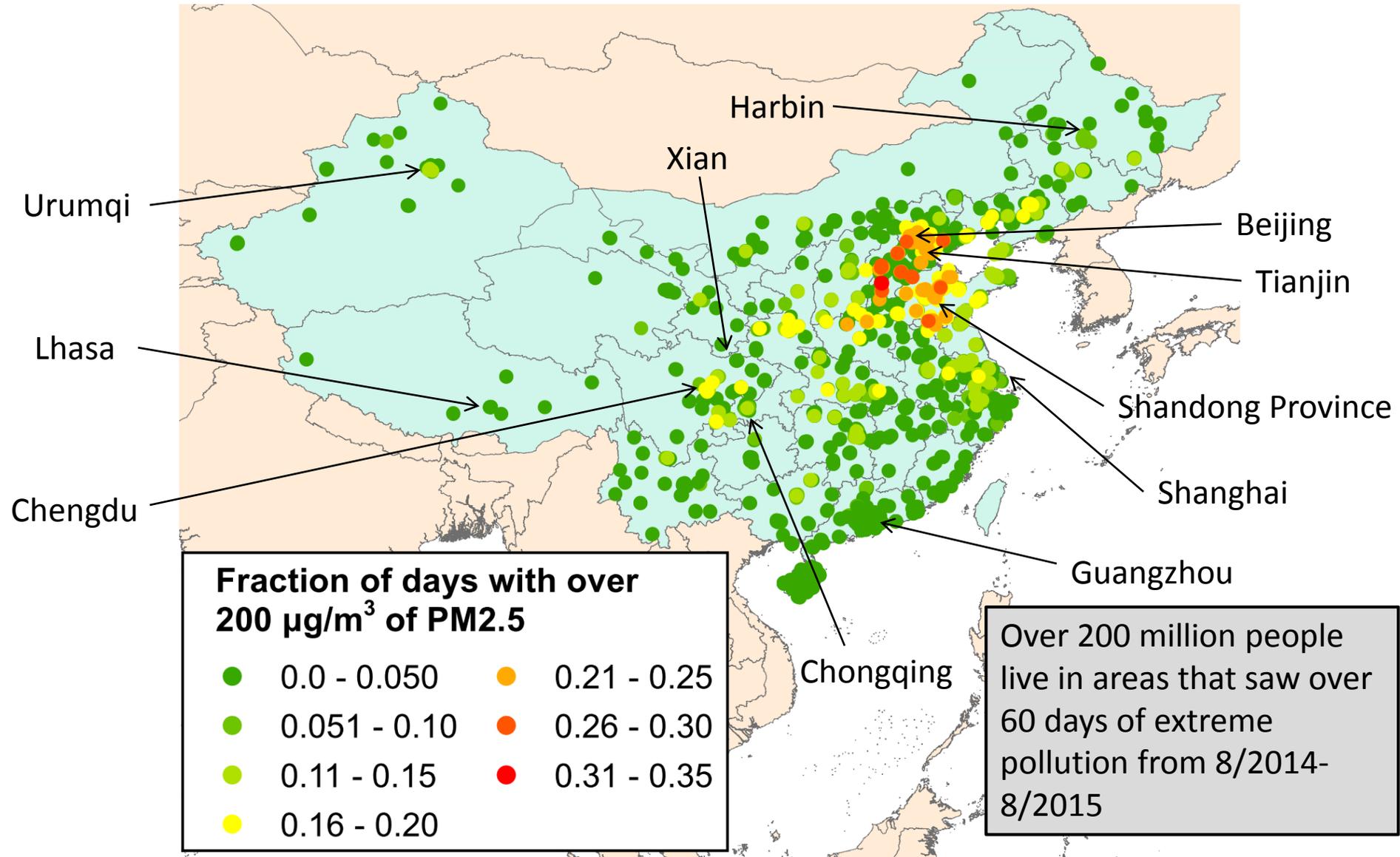


Beijing PM_{2.5}



We find that extreme pollution events are concentrated around Beijing and Tianjin

- Extreme events defined here as $PM_{2.5} > 200 \mu g m^{-3}$



Future Work

- Create channels for open data exchange
- Identify partners for collaborative analysis using the data
- Continue to develop the analytical model
- Promote discussions on air pollution towards reducing its harmful impact on societies in China, India, and around the world.

Acknowledgement

- **Zhenyang Hua** contributed to the initial data-archiving program during his internship at Harvard CGA in 2014.
- **David Strohschein** contributed to the hourly data scraping program during his fellowship at Harvard CGA in 2015.
- **René Westerholt** of the Heidelberg University contributed in the development of a geographically weighted regression [GWR] model for data analysis following his visiting fellowship at Harvard CGA in 2013.
- Slides #24 & 25 are by **Jonathan Moch**, Ph.D. student of the Department of Earth and Planetary Science, Harvard University.

References

- E. Wong, “Smog So Thick, Beijing Comes to a Standstill,” New York Times, December 9, 2015, p. A1.
- S. Wee, “China to invest \$277 billion to curb air pollution: state media,” Reuters, United States Edition, Environment, R. Popeski Ed., Wed Jul 24, 2013, 10:15pm EDT.
- Daily News & Analysis, “India's first Air Quality Index launched; will monitor pollution levels across country,” India, New Delhi, Apr 6, 2015.
- Center for Geographic Analysis, “Air Quality Index Archive,” <http://aqi.cga.harvard.edu/china/about/>, Harvard University, last accessed on March 15, 2017.
- R. A. Rohde and R. A. Muller, “Air Pollution in China: Mapping of Concentrations and Sources,” Berkeley Earth, 2015, <http://berkeleyearth.org/wp-content/uploads/2015/08/China-Air-Quality-Paper-July-2015.pdf> , last accessed on March 15, 2017.
- Beijing City Lab, “China PM2.5,” <https://www.beijingscitylab.com/projects-1/13-pm2-5/>, Projects, last accessed on March 15, 2017.
- Central Pollution Control Board, “National Air Quality Index,” <http://164.100.160.234:9000/>, Ministry of Environment, Forests and Climate Change, Government of India, last accessed on: March 15, 2017.
- Central Pollution Control Board, “National Air Quality Monitoring Programme Introduction,” <http://www.cpcb.nic.in/air.php>, Ministry of Environment, Forests and Climate Change, Government of India, last accessed on: March 13, 2017.
- OpenWeatherMap, Inc., “Open Weather Map,” <http://openweathermap.org/>, last accessed on: March 15, 2017.
- R. Westerholt, B. Resch and A. Zipf, “A local scale-sensitive indicator of spatial autocorrelation for assessing high- and low-value clusters in multiscale datasets,” International Journal of Geographical Information Science, vol. 29, pp.868-887, 2015 Issue 5.
- M. Bravo, M. Fuentes, Y. Zhang, M. J. Burr and M. Bell, “Comparison of exposure estimation methods for air pollutants: Ambient monitoring data and regional air quality simulation,” Environ Res. Vol. 116, pp.1–10, 2012 July.
- B. Lex, “China AQI PM25s Archive Dataverse,” https://dataverse.harvard.edu/dataverse.xhtml?alias=china_pm25s, Harvard Dataverse, V1, 2017.

Questions?

Enabling Spatiotemporal Analysis and Visualization of Air Pollution in China and India

Merrick Lex Berman, Devika Kakkar, Wendy Guan & Fei Carnes
Center for Geographic Analysis
Harvard University

The 25th International Conference on Geoinformatics
Buffalo, NY
August 2017