MODELING THE AIR POLLUTION IN CHINA

— BASED ON NONPARAMETRIC METHODS

Zihan Zhang

December 15, 2021

MOTIVATION



Figure: World map with AQI

DATA

- Hourly city AQI from the National Environmental Monitoring Center (CNEMC).
- City location data.

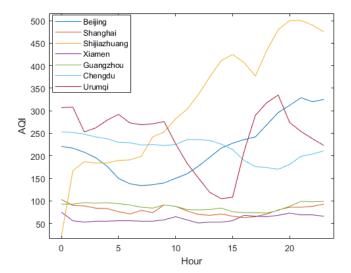


Figure: AQI of seven representative cities in China.

DATA

Table: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)
	N	mean	sd	min	max
AQI per hour in 7 representative cities					
Beijing	8,497	102.0	82.36	9	500
Shijiazhuang	8,560	132.7	102.2	15	500
Shanghai	8,559	65.88	41.98	9	452
Xiamen	8,558	46.31	23.04	7	221
Guangzhou	8,559	54.63	26.73	9	264
Chengdu	8,559	89.98	48.91	14	310
Urumqi	8,559	107.9	88.55	9	500
_					
Average AQI over year for 285 cities					
AQI	285	74.22	20.42	29.15	132.7



CITY LOCATION

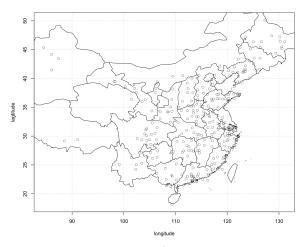


Figure: City location

METHODS

Gaussian process and Local constant kernel regression

$$\begin{bmatrix} y \\ f(x^*) \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} k(x,x) + \sigma_n^2 I & k(x,x^*) \\ k(x^*,x) & k(x^*,x^*) \end{bmatrix}\right)$$

$$f(x^*) \mid x^*, x, y \sim \mathcal{N}\left(k(x^*,x) \left[k(x,x) + \sigma_u^2 I\right]^{-1} y, \\ k(x^*,x^*) - k(x^*,x) k(x,x)^{-1} k(x,x^*)\right)$$
, where $k(x,x') = \sigma_f^2 \exp\left(-\left(x - x'\right)^2 / \left(2l^2\right)\right)$
Hyper-parameters $\hat{\theta} = \left(\sigma_f, \sigma_n, l\right)$ are chosen by,

$$\hat{\theta} = \arg \max_{\theta} \log p(y \mid t, \theta) = -\frac{1}{2} y^{T} K_{y}(\theta)^{-1} y - \frac{1}{2} \log |K_{y}(\theta)| - \frac{n}{2} \log 2\pi$$

RESULT: HOURLY TREND

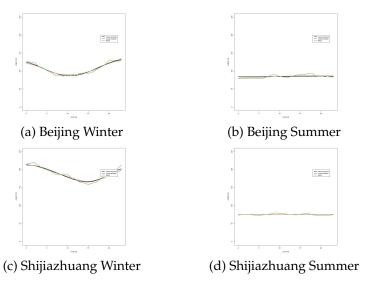


Figure: AQI hourly trend: Northern cities

RESULT: HOURLY TREND

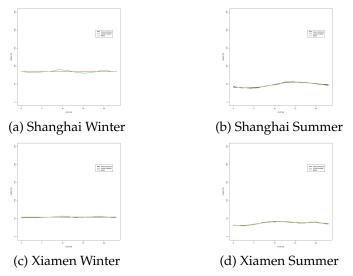
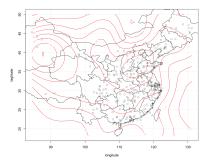
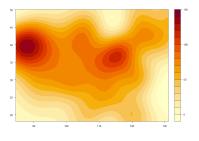


Figure: AQI hourly trend: Southern cities

RESULT: SPATIAL DISTRIBUTION





- Thanks -