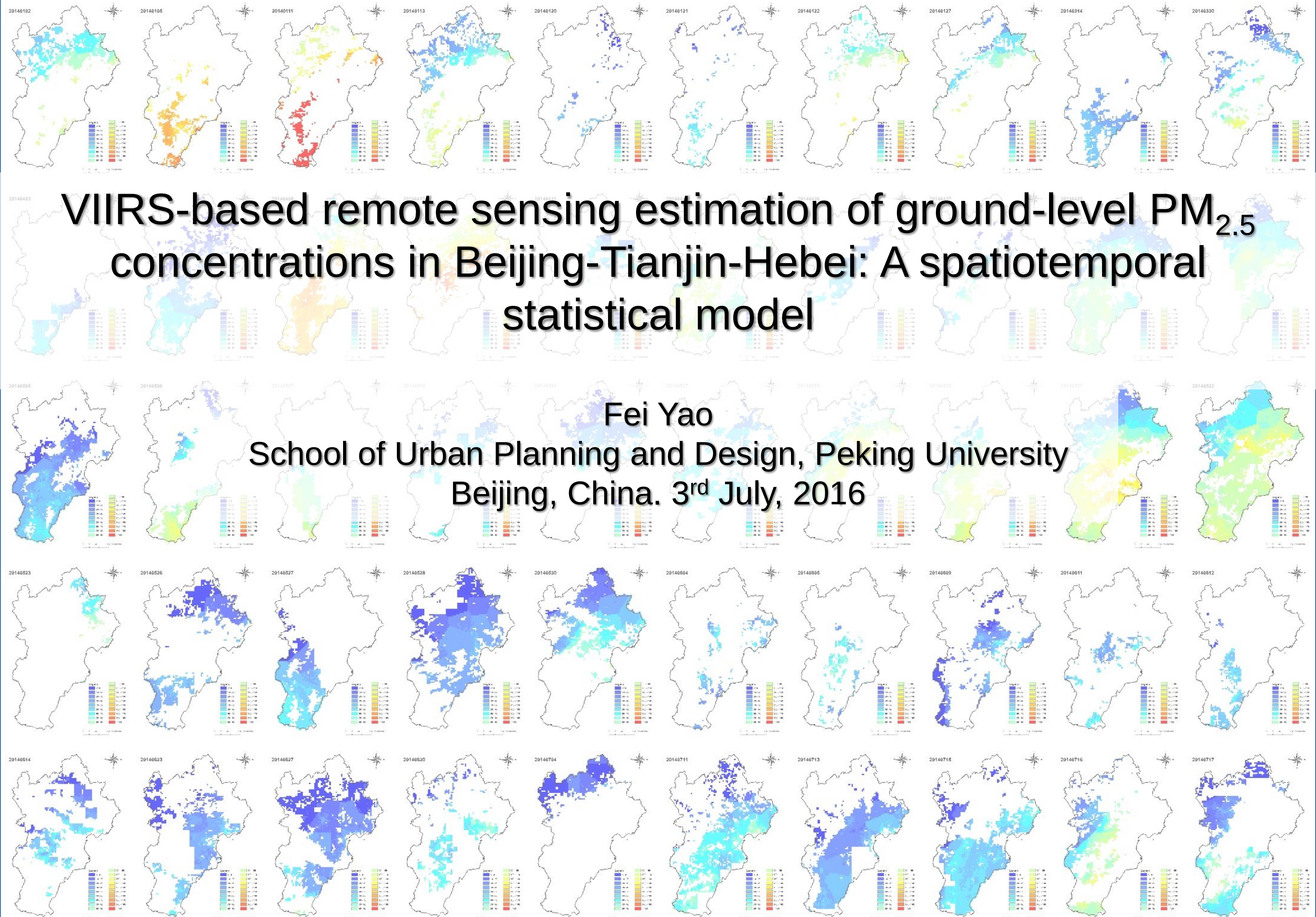


VIIRS-based remote sensing estimation of ground-level $PM_{2.5}$ concentrations in Beijing-Tianjin-Hebei: A spatiotemporal statistical model

Fei Yao
School of Urban Planning and Design, Peking University
Beijing, China. 3rd July, 2016

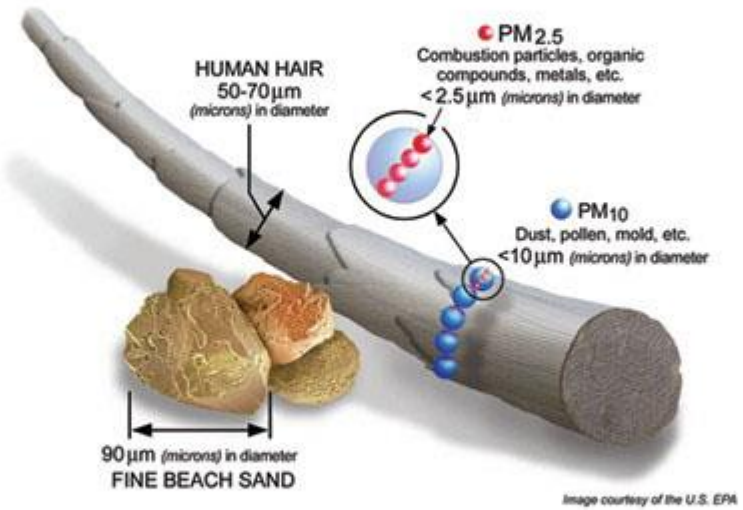
Contents

Contents

Part One
Introduction

What is $PM_{2.5}$?

- Particles with aerodynamic diameters of less than 2.5 μm



Source: US MA

Part One
Introduction

Adverse outcomes associated with $PM_{2.5}$

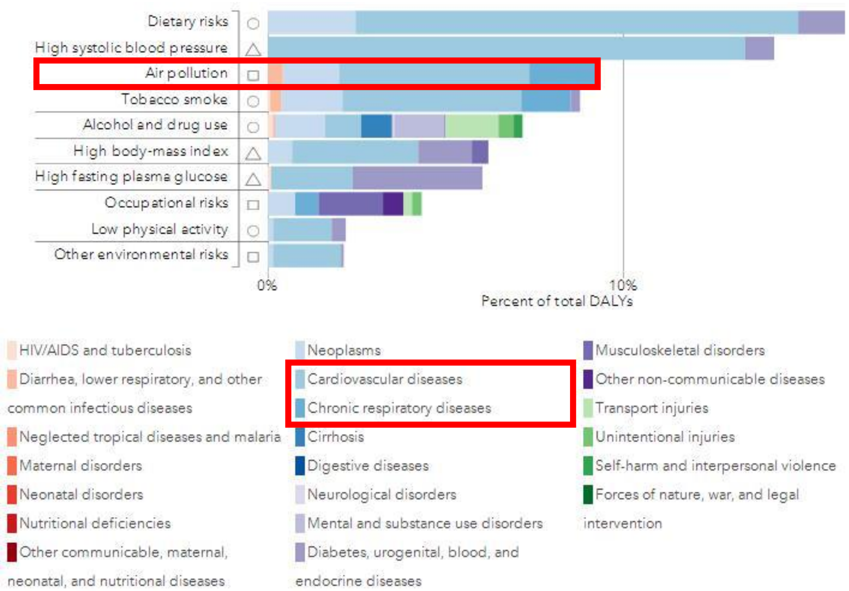
- Decreasing the visibility of the atmosphere (Tao et al. 2007; Liu et al. 2013)

Source: Sina Weibo

Part One
Introduction

Adverse outcomes associated with $PM_{2.5}$

- Increasing cardiovascular- and respiratory-related morbidity and mortality (Pope et al. 2002; Dominici et al. 2006; Pope and Dockery 2006)



Source: http://www.healthdata.org/china

Part One

Introduction

The significance of $PM_{2.5}$ data collection

- Conducting environmental epidemiologic studies
- Drafting appropriate air pollution control polices

Part One
Introduction

Two ways for $PM_{2.5}$ data collection

- Ground monitoring

Source: http://113.108.142.147:20035/emcpublish/

Expensive operating costs

⇩

Uneven spatial distribution of monitoring sites

⇩

No temporally- and spatially- full covered $PM_{2.5}$ data

Part One
Introduction

Two ways for $PM_{2.5}$ data collection

- Using satellite-derived aerosol optical depth (AOD) to estimate
- Temporally- and spatially- full covered $PM_{2.5}$ data collection is possible

Sensor	Satellite	Retrieval algorithm	Spatial resolution	Lastest version	Remarks
MODIS	Terra/Aqua	DT	10km 3km(C6)	C6	C5 has been applied mostly
		DB	10km	C6	The accuracy of C6 is much higher than C5
		MAIAC	1km	trial version	Not yet global coverage
MISR	Terra	EOF	17.6km	V22	High prediction accuracy, however, long revisit period.
SeaWiFS	SeaStar	DB	13.5km	V004	Ended in Octobor, 2010 because of a mechanical trouble
VIIRS	Suomi-NPP	DT	6km/750m	beta version	An explanation and improvement of AVHRR and MODIS

Part One
Introduction

Two limitations of previous satellite related studies

- Previous studies used MODIS C5 and MISR AOD data mostly, however, their spatial resolutions are relatively coarse

Sensor	Satellite	Retrieval algorithm	Spatial resolution	Lastest version	Remarks
MODIS	Terra/Aqua	DT	10km 3km(C6)	C6	C5 has been applied mostly
		DB	10km	C6	The accuracy of C6 is much higher than C5
		MAIAC	1km	trial version	Not yet global coverage
MISR	Terra	EOF	17.6km	V22	High prediction accuracy, however, long revisit period.
SeaWiFS	SeaStar	DB	13.5km	V004	Ended in Octobor, 2010 because of a mechanical trouble
VIIRS	Suomi-NPP	DT	6km/750m	beta version	An explanation and improvement of AVHRR and MODIS

- We will explore the performance of VIIRS AOD

Part One
Introduction

Two limitations of previous satellite related studies

- Quantitative relationships between $PM_{2.5}$ and AOD were built using statistical models mostly, however, these models rarely simultaneously considered the temporal and spatial variations of $PM_{2.5}$-AOD relationships

Statistical Models	Representatives	Temporal variations considered	Spatial variations considered
Simple linear model	Engel-Cox et al. 2004	No	No
Multiple linear regression model	Jia et al. 2014	No	No
Generalized linear regression model	Liu et al. 2005 Liu et al. 2007	No	No
Geographically weighted regression model	Hu et al. 2013 Song et al. 2014 Ma et al. 2014	No	Yes
Linear mixed effects model	Li et al. 2015	Yes	No
Generalized additive model	Liu et al. 2009	Yes	Yes
Two-stage model	Hu et al. 2014 Ma et al. 2016	Yes	Yes

- We will develop a spatiotemporal statistical model

Contents

Part Two
Data & Methods

Data

- All the data were collected from the Internet

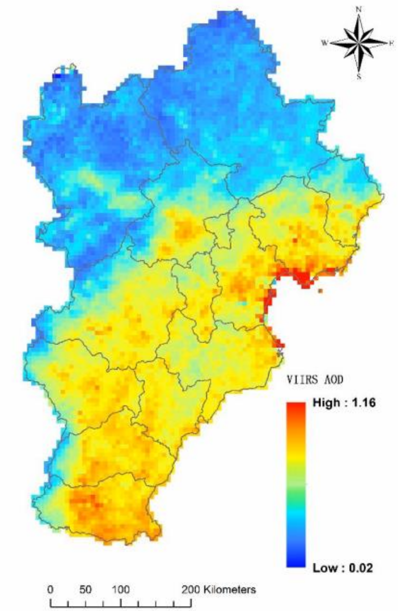
Data		Type	Spatial resolutions	Source
$PM_{2.5}$		Point	\	http://113.108.142.147:20035/emcpublish/ http://zx.bjmemc.com.cn/
VIIRS AOD		Raster	6 km	http://www.class.ngdc.noaa.gov/saa/products/welcome
Surface meteorolgical data		Point	\	http://www.escience.gov.cn/metdata/page/index.html
Aerological data	RH	Raster	1.25° × 1.25°	http://disc.sci.gsfc.nasa.gov/daac-bin/FTPSubset.pl?LOOKUPID_List=MAI3CPASM
	PBLH		0.5° × 0.5°	
Satellite-derived NDVI		Raster	250 m	https://ladsweb.nascom.nasa.gov/data/search.html
Satellite derived NO_2		Raster	0.25° × 0.25°	http://www.temis.nl/airpollution/no2col/no2regioomi_v2.php

Part Two

Data & Methods

Data integration

- Nearest neighbor approach



Y	X													
$PM_{2.5}$	AOD	TP	SRH	RF	PBLH	RH_PBLH	NDVI	NO_2	TOE	TOS	TOW	TON	seq	site

Time

Space

- We finally obtained a spatial panel dataset

Part Two

Data & Methods

Model development

- Stage I: Time fixed effects regression model
  - $PM_{2.5,st} = Intercept_t + \beta_{\mathrm{AOD}} * \mathrm{AOD_{st}} + \beta_{\mathrm{TP}} * \mathrm{TP_{st}} + \beta_{SRH} * SRH_{st} + \beta_{RF} * RF_{st} + \beta_{PBLH} * PBLH_{st} + \beta_{RH_{PBLH}} * RH_{PBLH}{}_{st} + \beta_{NDVI} * NDVI_{st} + \beta_{NO2_{Lag}} * NO2_{Lag}{}_{st} + \beta_{TOE} * TOE_{st} + \beta_{TOS} * TOS_{st} + \beta_{TOW} * TOW_{st} + \beta_{TON} * TON_{st} + \varepsilon_{st}$
- Stage II: Geographically weighted regression model
  - $Residual_{ss'} = \beta_{0,s} + \beta_{\mathrm{AOD},s} * AOD_{ss'} + \varepsilon_{ss'}$

Spatial variations

Temporal variations

Part Two

Data & Methods

Model development

- Stage I: Time fixed effects regression model
  - $PM_{2.5,st} = Intercept_t + \beta_{\mathrm{AOD}} * \mathrm{AOD}_{\mathrm{st}} + \beta_{\mathrm{TP}} * \mathrm{TP}_{\mathrm{st}} + \beta_{SRH} * SRH_{st} + \beta_{RF} * RF_{st} + \beta_{PBLH} * PBLH_{st} + \beta_{RH_{PBLH}} * RH_{PBLH\,st} + \beta_{NDVI} * NDVI_{st} + \beta_{NO2_{Lag}} * NO2_{Lag\,st} + \beta_{TOE} * TOE_{st} + \beta_{TOS} * TOS_{st} + \beta_{TOW} * TOW_{st} + \beta_{TON} * TON_{st} + \varepsilon_{st}$

- Stage II: Geographically weighted regression model
  - $Residual_{ss'} = \beta_{0,s} + \beta_{\mathrm{AOD},s} * AOD_{ss'} + \varepsilon_{ss'}$

Spatial variations

Temporal variations

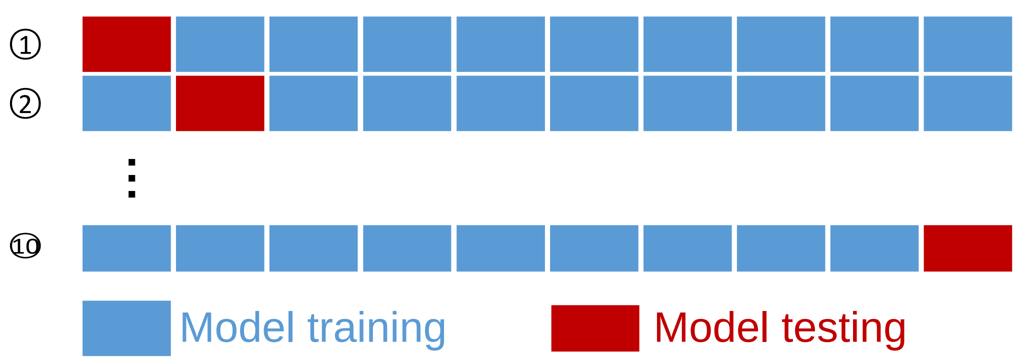
- Final $PM_{2.5}$=$PM_{2.5}$ from Stage I + Residual from Stage II

Part Two

Data & Methods

Model validation

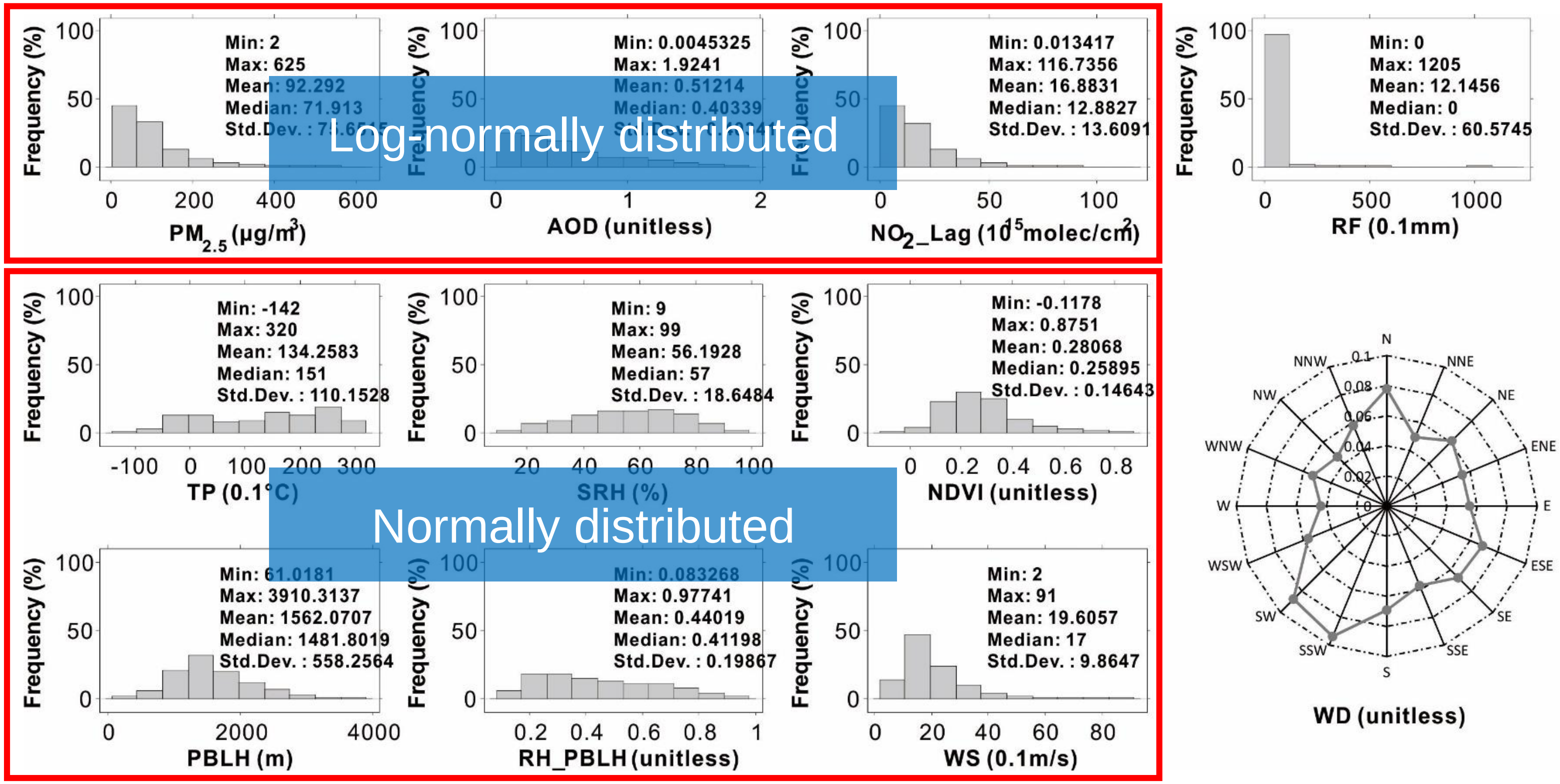
- Statistical indicators
  - Coefficient of determination (R^2)
  - Mean predication error (MPE)
  - Root-mean-square error (RMSE)
  - Residual spatial autocorrelation (Moran’s I)
- Ten-folder cross validation



Contents

Part Three
Results

Descriptive statistics



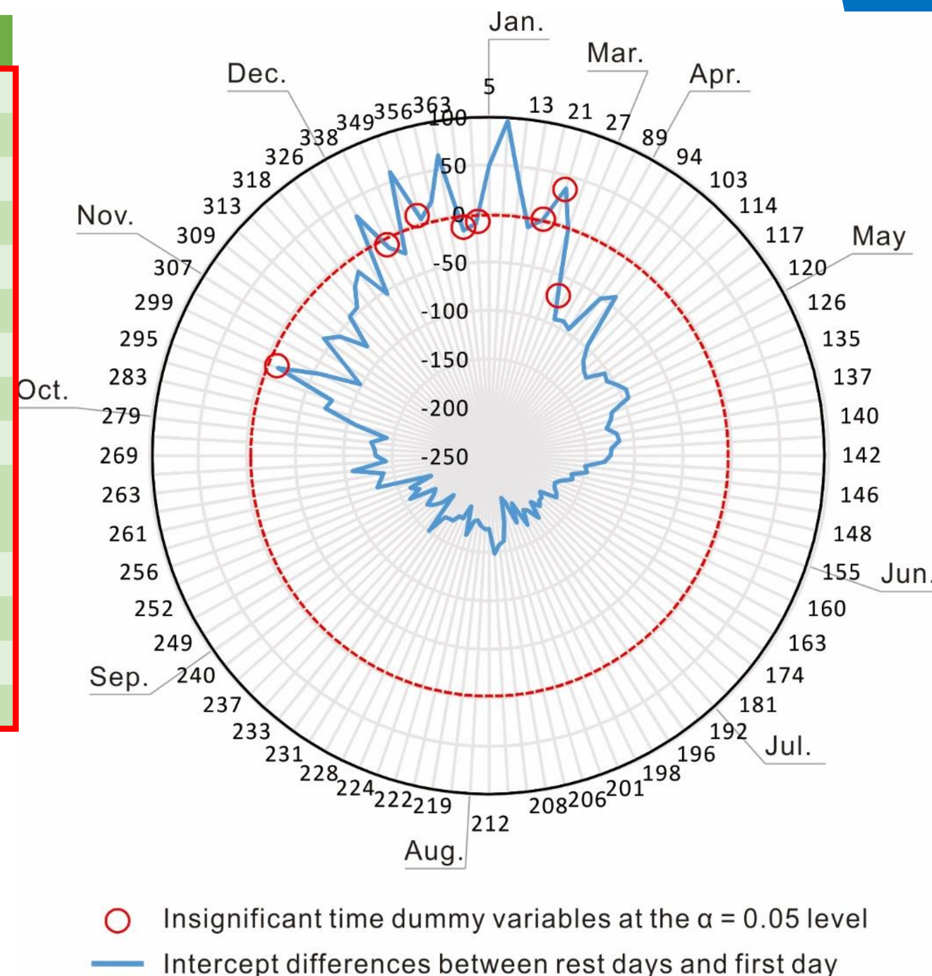
Part Three
Results

Model fitting – Time fixed effects regression model

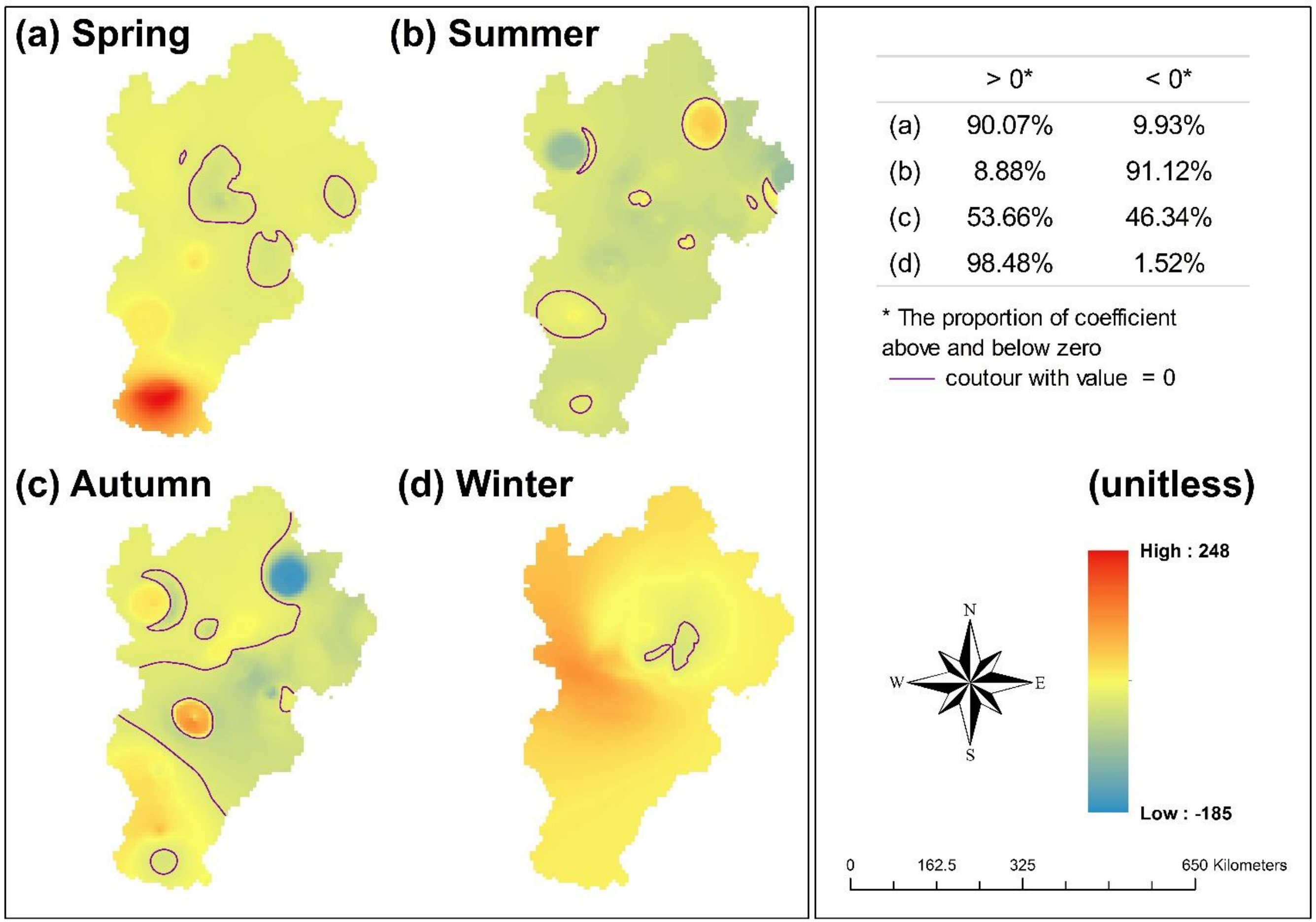
Time fixed effects regression model			
	b	P-value	Magnitude
Intercept*	40.813	0.000	
AOD(unitless)	26.499	0.000	50.778
TP(0.1℃)	0.514	0.000	213.919
SRH(%)	1.059	0.000	82.632
RF(0.1mm)	-0.048	0.003	-33.498
PBLH(m)	-0.004	0.004	-13.951
RH_PBLH(%)	-28.165	0.000	-21.421
NDVI(unitless)	-5.910	0.051	-4.843
NO_2_Lag (10^{15}molec/cm^2)	0.123	0.098	10.195
TOE(0.1m/s)**	-0.078	0.267	-3.651
TOS(0.1m/s)	-0.414	0.000	-23.582
TOW(0.1m/s)	-0.228	0.007	-9.903
TON(0.1m/s)	-0.215	0.005	-8.546

* Intercept of the first day

** Not significant

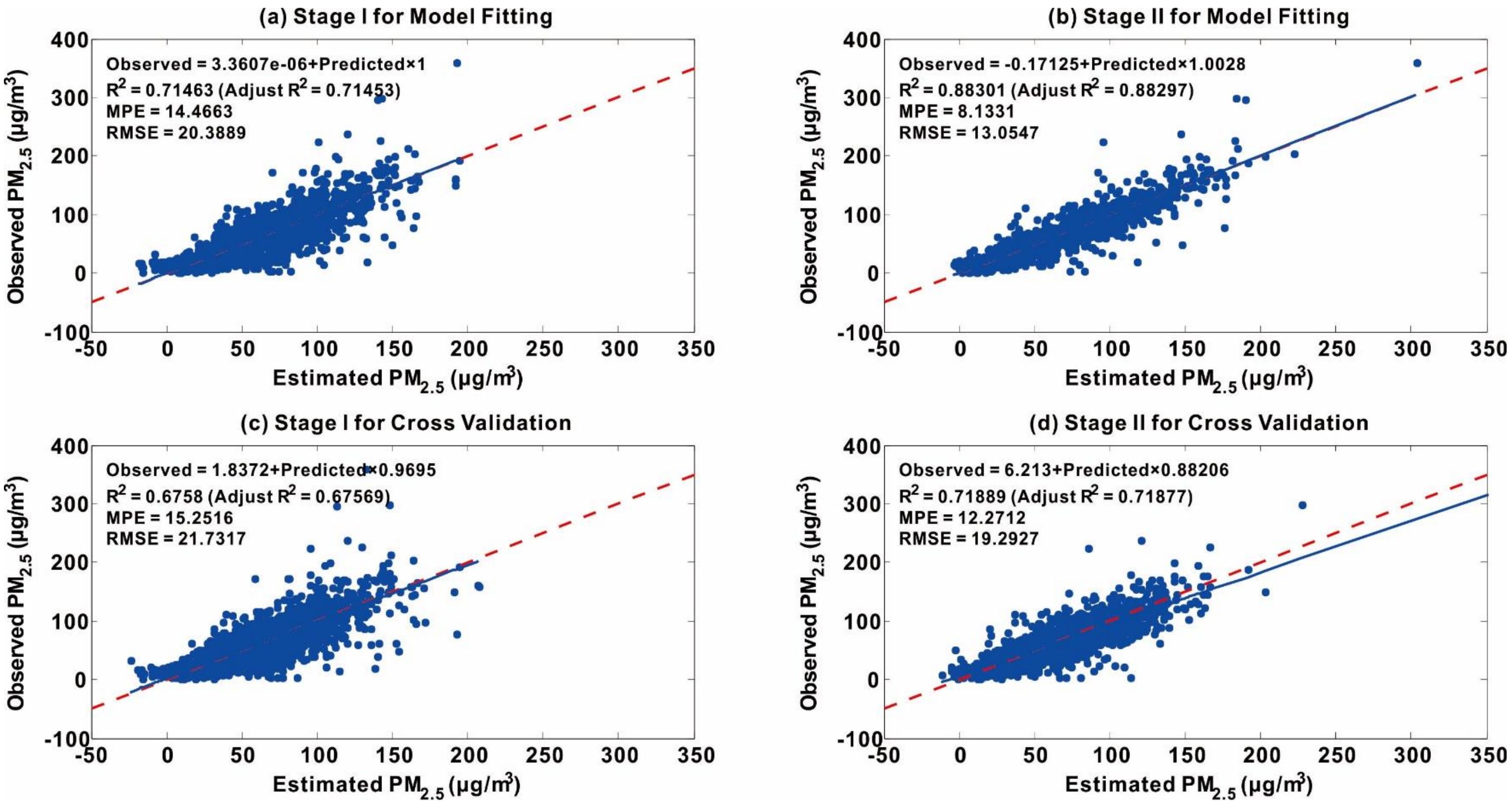


Model fitting – Geographically weighted regression model



Part Three
Results

Model validation – Overfitting degree

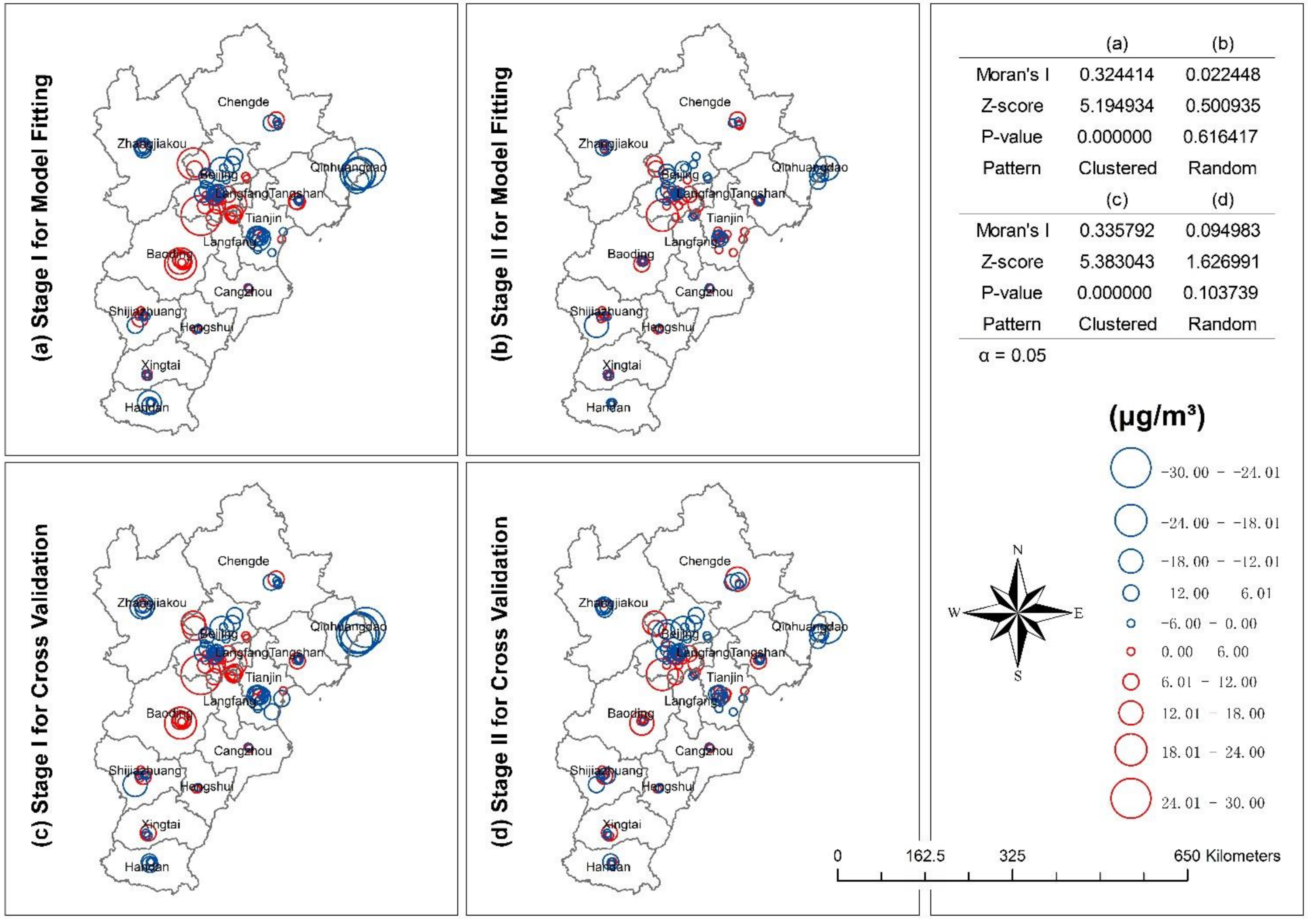


R^2 decreased by 0.03883

R^2 decreased by 0.16412

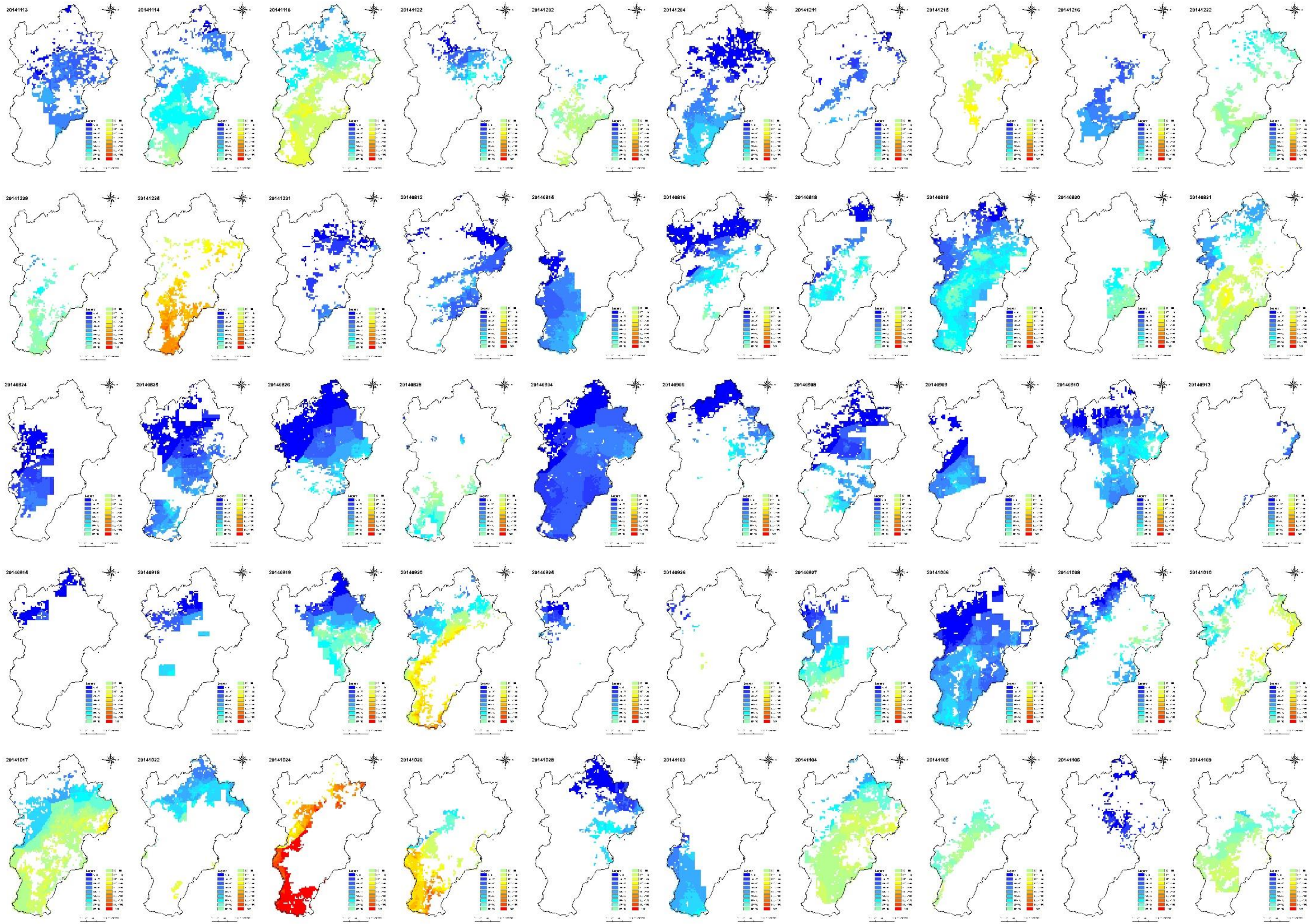
Part Three
Results

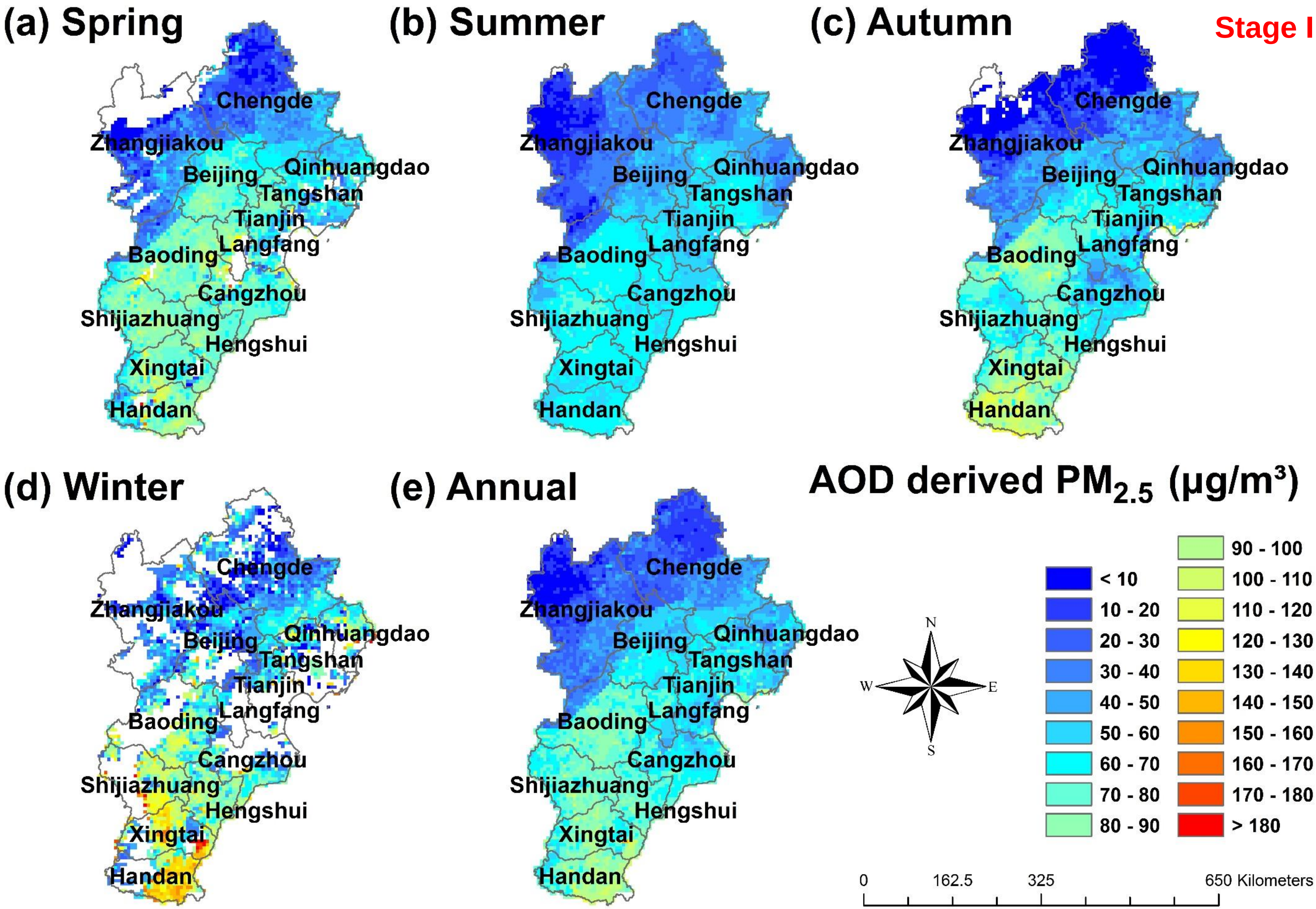
Model validation – Residual spatial autocorrelation

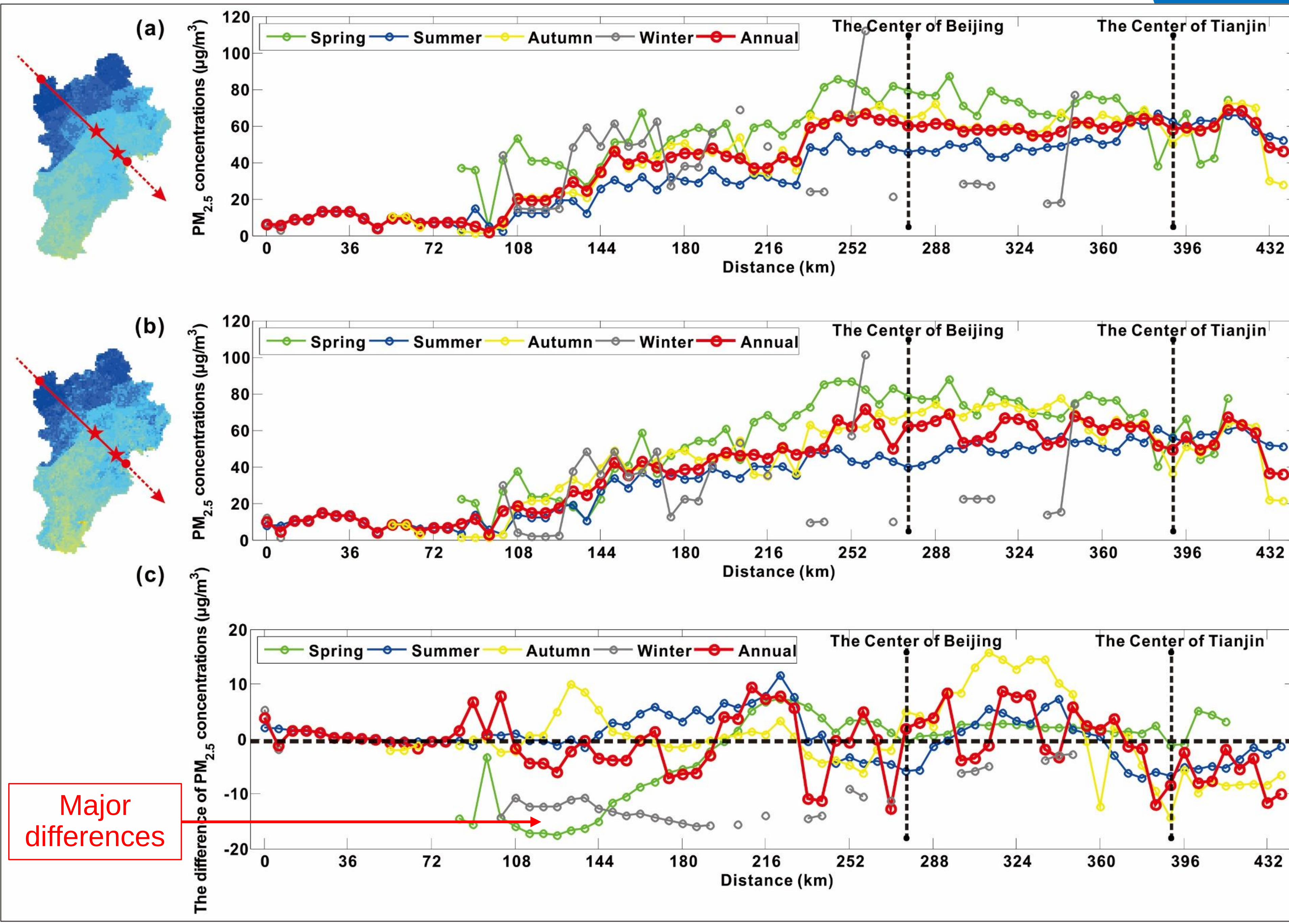


	(a)	(b)
Moran's I	0.324414	0.022448
Z-score	5.194934	0.500935
P-value	0.000000	0.616417
Pattern	Clustered	Random
	(c)	(d)
Moran's I	0.335792	0.094983
Z-score	5.383043	1.626991
P-value	0.000000	0.103739
Pattern	Clustered	Random

α = 0.05


20141113
20141114
20141118
20141122
20141202
20141204
20141211
20141215
20141216
20141222
20141229
20141225
20141231
20140812
20140815
20140816
20140818
20140819
20140820
20140821
20140824
20140825
20140826
20140828
20140904
20140906
20140908
20140909
20140910
20140913
20140915
20140918
20140919
20140920
20140925
20140926
20140927
20141006
20141008
20141010
20141017
20141022
20141024
20141026
20141028
20141103
20141104
20141105
20141108
20141109


Stage II
(a) Spring
(b) Summer
(c) Autumn
(d) Winter
(e) Annual
Chengde
Zhangjiakou
Beijing
Qinhuangdao
Tangshan
Tianjin
Langfang
Baoding
Cangzhou
Shijiazhuang
Hengshui
Xingtai
Handan
AOD derived PM2.5 (μg/m³)
< 10
10 - 20
20 - 30
30 - 40
40 - 50
50 - 60
60 - 70
70 - 80
80 - 90
90 - 100
100 - 110
110 - 120
120 - 130
130 - 140
140 - 150
150 - 160
160 - 170
170 - 180
> 180
N
E
S
W
0
162.5
325
650 Kilometers

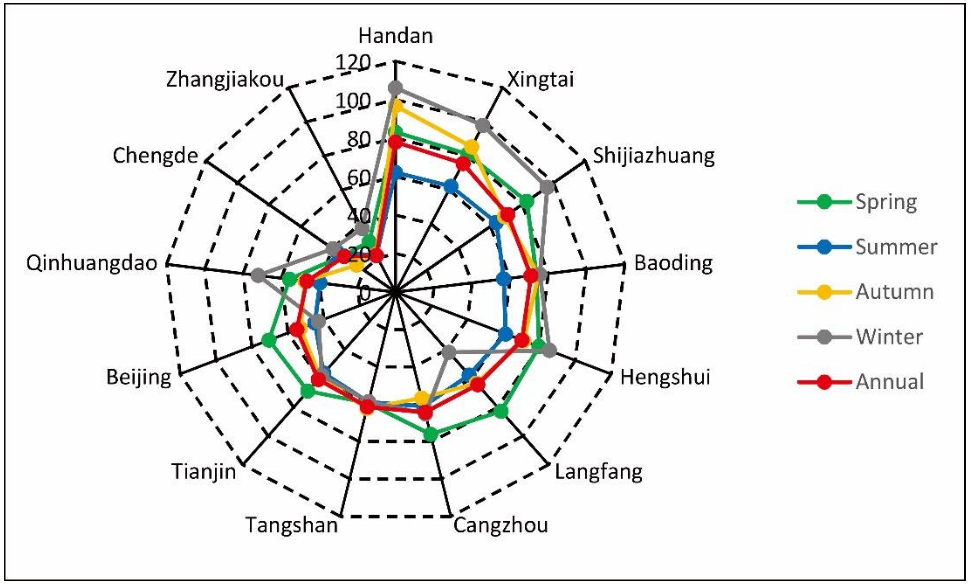


(a)
(b)
(c)
Spring
Summer
Autumn
Winter
Annual
The Center of Beijing
The Center of Tianjin
PM2.5 concentrations (μg/m3)
The difference of PM2.5 concentrations (μg/m3)
Distance (km)
0
36
72
108
144
180
216
252
288
324
360
396
432
120
100
80
60
40
20
10
-10
-20
Major differences

Part Three
Results

Predication maps of $PM_{2.5}$ concentrations

- $PM_{2.5}$ concentrations among all prefecture-level cities



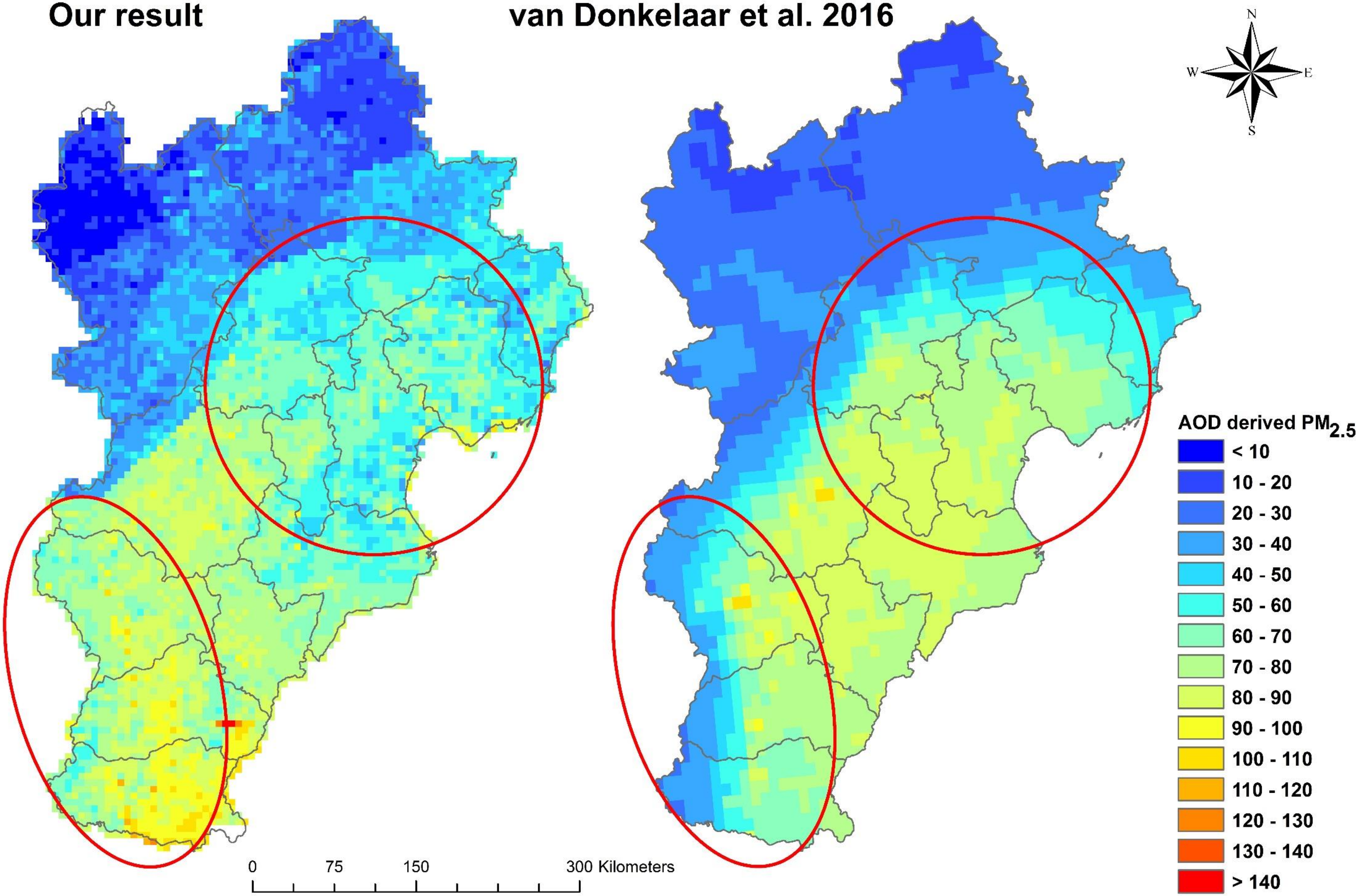
- Beijing and Tianjin were in medium level

Contents

Part Four
Discussion

The novelty of methodology

- Previous two-stage models often employed linear mixed effects model in their first stage while we employed time fixed effects regression model, which is computationally lighter and operationally easier for model calibration and prediction. And the model's performance was comparable or even better.


Our result
van Donkelaar et al. 2016
N
W
E
S
AOD derived $PM_{2.5}$
< 10
10 - 20
20 - 30
30 - 40
40 - 50
50 - 60
60 - 70
70 - 80
80 - 90
90 - 100
100 - 110
110 - 120
120 - 130
130 - 140
> 140
0
75
150
300 Kilometers

Part Four
Discussion

The application of our work

- Our work is a demonstration of the method and can be extended to other regions but cautions should be paid on whether the region has the characteristic of urban-industrial conditions.
- We could also estimate $PM_{2.5}$ concentrations of the past and near future if we assume that the spatiotemporal variations of $PM_{2.5}$-AOD relationship was constant in each year.

Part Four
Discussion

The limitations of our work

- The deficiency of matched data records per day
- The data integration method is relatively simple

Possible solutions

- Seeking a trade-off between the minimum number of matched data records per day and the model’s overfitting degree
- Adopting the mean value of some variables over a certain range from the monitoring site. Adopting spline interpolation for the meteorological data.

Contents

Part Five

Conclusions

Listed as below

- Time fixed effects regression model captured the temporal variations of $PM_{2.5}$-AOD relationships
- Geographically weighted regression model captured the spatial variations of $PM_{2.5}$-AOD relationships
- The ground-level $PM_{2.5}$ concentrations were significantly affected by meteorological factors, land use characteristics, and other air pollutants
- The prediction maps revealed that fine particulate pollution in Beijing–Tianjin–Hebei is severe and the pollution pattern presents relatively strong seasonal heterogeneity and southeast–northwest spatial heterogeneity

Thanks

FAQ