

Supplemental Material to:

Spatial-temporal distribution of disability-adjusted life-years of lung cancer attributable to ambient PM_{2.5} in Guangzhou, China, 2010-2013: a population-based study

Xiao Lin^{a,1}, Hang Dong^{b,1}, Guozhen Lin^b, Yan Li^b, Qiongying Yang^b, Yu Liao^c, Ao Luo^a, Binglun Liang^a, Zhicong Yang^{b,2,*}, and Yuantao Hao^{a,d,2,*}

^aDepartment of Medical Statistics and Epidemiology & Health Information Research Center & Guangdong Key Laboratory of Medicine, School of Public Health, Sun Yat-sen University, Guangzhou, 510080, Guangdong, China.

^bGuangzhou Center for Disease Control and Prevention, Guangzhou, 510440, Guangdong, China.

^cGuangdong Provincial Center for Disease Control and Prevention, Guangzhou, 511430, Guangdong, China.

^dSun Yat-sen Global Health Institute, Sun Yat-sen University, Guangzhou, 510080, Guangdong, China. Professor and dean of School of Public Health.

¹ Xiao Lin and Hang Dong contributed equally to this work.

² Zhicong Yang and Yuantao Hao also contributed equally to this work.

***Address correspondence to** Prof Yuantao Hao, Sun Yat-sen University, 74 Zhongshan 2nd Rd, Guangzhou, 510080, Guangdong, China. Telephone: +86-13610271412. E-mail: haoyt@mail.sysu.edu.cn. or Director Zhicong Yang, Guangzhou Center for Disease Control and Prevention, Guangzhou, 510440, Guangdong, China. E-mail: yangzc@gzcdc.org.cn.

Declarations of interest: none.

Contents

Table S1 3

Table S2 8

Table S3 9

Figure S1..... 11

Figure S2..... 12

Figure S3..... 14

Reference 15

Table S1. Annual age-specific population counts for permanent registered residents in Guangzhou, in 2010-2013, by administrative district.

Age group (years)	2010	2011	2012	2013
Liwan				
0-4	29632	23153	24204	25123
5-14	67868	48312	47738	47379
15-29	184809	131714	129081	126551
30-44	176593	149164	149697	148290
45-59	147423	211756	209609	209103
60-69	50627	69720	75057	82229
70-79	35418	50296	48864	46984
80+	15152	26169	27415	29620
Yuexiu				
0-4	48882	42993	43241	44040
5-14	111959	86418	86456	86972
15-29	304948	228166	221996	216732
30-44	291457	276764	277043	272388
45-59	243291	314295	311941	313489
60-69	83565	102880	109699	119237
70-79	58476	80746	79220	76012
80+	25039	38987	41627	45422
Haizhu				
0-4	39598	39900	41478	42506
5-14	90659	73784	74903	75833
15-29	246809	189356	183027	178526
30-44	235771	232617	236862	236556
45-59	196862	250851	250977	254138
60-69	67597	90186	96701	105881

Age group (years)	2010	2011	2012	2013
70-79	47267	64019	64279	63056
80+	20200	26530	28506	31619
Baiyun				
0-4	34117	45116	47310	50368
5-14	78050	79925	81106	83232
15-29	212612	223801	218642	213067
30-44	203210	217160	224605	229935
45-59	169662	155539	157530	163198
60-69	58286	69480	74235	79386
70-79	40770	35596	36933	37737
80+	17463	16867	17460	18627
Huangpu				
0-4	8356	9405	9802	10114
5-14	19144	19813	19733	19509
15-29	52002	42781	42427	42212
30-44	49572	57204	57663	56977
45-59	41421	40488	41284	43048
60-69	14200	17541	18256	18973
70-79	9907	10777	11224	11479
80+	4198	3281	3463	3803
Panyu				
0-4	41811	48116	41492	45524
5-14	95752	105194	83768	83558
15-29	260848	285760	222232	214262
30-44	249346	259926	215537	218455

Age group (years)	2010	2011	2012	2013
45-59	208130	185308	144673	151914
60-69	71497	68813	56099	59801
70-79	50038	36815	29324	30266
80+	21438	16281	12773	13518
Tianhe				
0-4	31818	39802	41320	42820
5-14	72815	67821	71581	75224
15-29	198051	251156	238960	229012
30-44	189022	215829	225290	229626
45-59	157897	122584	126454	134687
60-69	54186	44420	46860	50071
70-79	37844	33415	34495	34987
80+	16117	9820	10726	12374
Huadu				
0-4	27463	35838	37837	40401
5-14	62870	61479	61628	62708
15-29	171213	190308	184630	178773
30-44	163608	170404	175613	178645
45-59	136593	120650	122148	126078
60-69	46913	49856	53606	57383
70-79	32815	25626	25936	26374
80+	14042	13744	13602	14359
Zengcheng				
0-4	34878	39620	39355	41418
5-14	79879	90008	84664	81447

Age group (years)	2010	2011	2012	2013
15-29	217442	246493	243386	239341
30-44	207705	210261	212959	213208
45-59	173421	152987	157158	163554
60-69	59542	52091	55308	59143
70-79	41638	32828	32652	32922
80+	17789	18739	18371	20330
Chonghua				
0-4	23697	30232	30846	32803
5-14	54221	68934	67710	65323
15-29	147636	170392	169888	169005
30-44	141051	149439	153950	154400
45-59	117781	99222	102050	106406
60-69	40450	33468	35568	38179
70-79	28282	20548	20455	20459
80+	12094	10258	10495	11124
Nansha				
0-4	6392	7078	17608	19017
5-14	14637	15494	34869	33476
15-29	39857	40084	93340	93942
30-44	38083	38657	87657	87114
45-59	31795	32632	78559	80687
60-69	10919	12076	30469	32576
70-79	7638	6775	15847	16162
80+	3267	3455	7721	8165
Luogang				

Age group (years)	2010	2011	2012	2013
0-4	7760	12084	13285	14651
5-14	17753	19954	20267	20960
15-29	48320	60775	59867	58861
30-44	46148	51250	54982	58154
45-59	38541	30735	31621	33318
60-69	13233	10873	11717	12533
70-79	9248	6107	6141	6267
80+	3949	3363	3209	3354
Total	7980045	8129427	8199852	8298470

Note: Data on permanent registered residents for 2010-2013 were collected from the population registry system at the Guangzhou Center for Disease Control and Prevention (Luo et al., 2019).

Table S2. Convergence diagnostics of IER model.

Items	Random initialization			Fixed initialization ^a		
	Coefficients (95% UI)	N_eff ^b	$\hat{R}^c$	Coefficients (95% UI)	N_eff	$\hat{R}$
α	2589.49 (534.08, 5791.18)	4475.17	1.0000	2600.79 (545.46, 5837.89)	3445.3	1.0001
β	0.00 (0.00, 0.00)	4871.36	0.9998	0.00 (0.00, 0.00)	1844.41	0.9998
γ	1.47 (1.20, 1.79)	4793.77	0.9999	1.47 (1.19, 1.80)	1937.25	0.9999
δ_1	0.87 (0.62, 1.25)	6000.00	0.9999	0.88 (0.62, 1.26)	2605.73	0.9999
δ_2	3.48 (0.14, 19.09)	3100.55	1.0000	6.73 (0.16, 33.82)	2208.31	1.0002
δ_3	0.12 (0.06, 0.20)	2506.75	0.9999	0.11 (0.06, 0.19)	3434.77	0.9998
δ_4	0.12 (0.01, 0.28)	1937.51	0.9999	0.12 (0.01, 0.28)	3431.32	0.9998
Time ^d						
Burn-in		2325.7		66.64		
Sampling		651.128		73.45		
WAIC ^e						
(SE)		-20980.72 (7.01)		-21059.12 (7.09)		
DIC ^f		-32.65		-32.98		

Note: IER model, integrated exposure-response model; UI, uncertainty interval. To properly estimate the unknown prior parameters, technique of two-step Bayesian fitting was implemented in the study (Foreman et al., 2012). The random-initialization IER models were run for three times (first step) and parameters from the resulting three models were implemented as initial priors in the fixed-initialization model (second step). The latter model was then used as the estimation model for relative risks. In the study, the final IER model was run using a Markov Chain Monte Carlo sampling technique, where we implemented the No-U-Turn Sampler algorithm

(NUTS algorithm) in the STAN Bayesian fitting language, for an iteration of 12000 steps, with 6000 steps as burn-ins, and for a total of three independent Markov Chains (Burnett et al., 2014; Cohen et al., 2017). Further convergence diagnostics of the IER model is presented in Figure S3 below, where traceplots for each parameter were drawn (Carpenter et al., 2017; Lunn et al., 2012).

^aFixed initial parameters at: $\alpha = 2600.79$; $\beta = 0.00$; $\gamma = 1.47$; $\delta_1 = 0.88$; $\delta_2 = 6.73$; $\delta_3 = 0.11$; $\delta_4 = 0.12$.

^bN_eff, number of effective samples (Carpenter et al., 2017).

^c $\hat{R}$, potential scale reduction factor for diagnosing convergence in Markov Chains (Carpenter et al., 2017). If $\hat{R}$ of a parameter approximates one, the multiple Markov Chains are consistent among chains for the corresponding parameter.

^dTime for Monte Carlo simulation in IER models, in seconds (Stan, 2016).

^eWAIC, Watanabe-Akaike information criterion (the smaller the values, the better fit to the data) (Vehtari et al., 2017), generated by the loo package in R (version 3.4.3).

^fDIC, Deviance Information Criterion (the smaller the values, the better fit to the data) (Clements, 2005; Spiegelhalter et al., 2002).

Table S3. Diagnostics of variogram models fitted by universal kriging interpolation.

Dataset and model	Null model ^a			CV ₁₀ ^b Test		Model with PM _{2.5} covariate			CV ₁₀ Test		
	RSS ^c	RMSE ^d	Variance	MSE ^e	MSNE ^f	RSS	RMSE	Variance	MSE	MSNE	
		(%)	explained (%)				(%)	explained (%)			
<i>Dataset for 2010</i>											
Nug ^g	2392.73	62.09	0.05	0.39	1.98	2301.74	60.90	3.75	0.37	1.93	
Gau ^h	323.48	22.83	86.47	0.05	0.99	321.99	22.78	86.54	0.05	0.98	
Wav ⁱ	632.25	31.92	73.56	0.10	1.48	633.11	31.94	73.53	0.10	1.48	
<i>Dataset for 2011</i>											
Nug	2444.03	62.75	0.02	0.39	2.00	2186.03	59.35	10.54	0.35	1.88	
Gau	327.06	22.95	86.62	0.05	1.00	315.82	22.56	87.08	0.05	0.97	
Wav	633.58	31.95	74.07	0.10	1.49	635.52	32.00	73.99	0.10	1.46	
<i>Dataset for 2012</i>											
Nug	1379.98	47.15	0.05	0.22	1.78	1237.02	44.64	10.32	0.20	1.60	
Gau	210.61	18.42	84.73	0.03	1.01	197.90	17.86	85.65	0.03	0.96	
Wav	406.12	25.58	70.56	0.07	1.51	402.51	25.47	70.82	0.06	1.39	

Nug	1414.41	47.74	0.06	0.23	1.81	1216.11	44.26	13.97	0.20	1.60
Gau	206.72	18.25	85.38	0.03	1.00	206.09	18.22	85.42	0.03	0.99
Wav	410.87	25.73	70.93	0.07	1.50	408.92	25.67	71.07	0.07	1.48

Note: CV₁₀, 10-fold cross validation test; PM_{2.5}, particulate matter with aerodynamic diameter <2.5 μm . Prior to choosing specific kernel models for quantitative performance diagnosis, a range of kernels were tested using the eye-fit method (Höfle et al., 2017; Tsutsumi et al., 2011). However, only the Gaussian and Wave kernels stood out in the test and were chosen for further model diagnosis. The diagnosis of model goodness-of-fit suggests that variogram models with Gaussian kernel are better than those with Nugget or Wave kernel because Gaussian-kernel models have advantages in terms of smaller RSS (the sum of squares of the residuals) and RMSE (the root mean squared error) for model goodness-of-fit. Besides, in cross-validation test, Gaussian-kernel models are more likely to yield lower errors in terms of MSE (the mean squared error) and more likely to approximate one in terms of MSNE (the mean square normalized error). Also, the variogram model using PM_{2.5} as covariate outperforms the null model fitted with only intercept. The PM_{2.5}-fitted model has more advantages in explaining the spatial variance, while at the same time, yielding lower modeling errors in terms of RMSE (the root mean squared error).

^aModel was fitted without any covariates.

^bCV₁₀, 10-fold cross validation test (Clements, 2005).

^cRSS, the sum of squares of the residuals from the fitted function (Oliver and Webster, 2014).

^dRMSE, the root mean squared error (the smaller the values, the better fit to the data) (Hijmans and van Etten, 2014).

^eMSE, the mean squared error (the smaller the values, ideally close to zero, the better fit to the data) (Oliver and Webster, 2014).

^fMSNE, the mean square normalized error (the closer to one, the better fit to the data) (Bivand et al., 2008; Oliver and Webster, 2014).

^gNug, the nugget kernel in kriging, which is a baseline variogram model for spatial interpolation (Lam, 1983; Oliver and Webster, 1990).

^hGau, the Gaussian-kernel model (Lam, 1983; Oliver and Webster, 1990).

ⁱWav, the Wave-kernel model (Lam, 1983; Oliver and Webster, 1990).

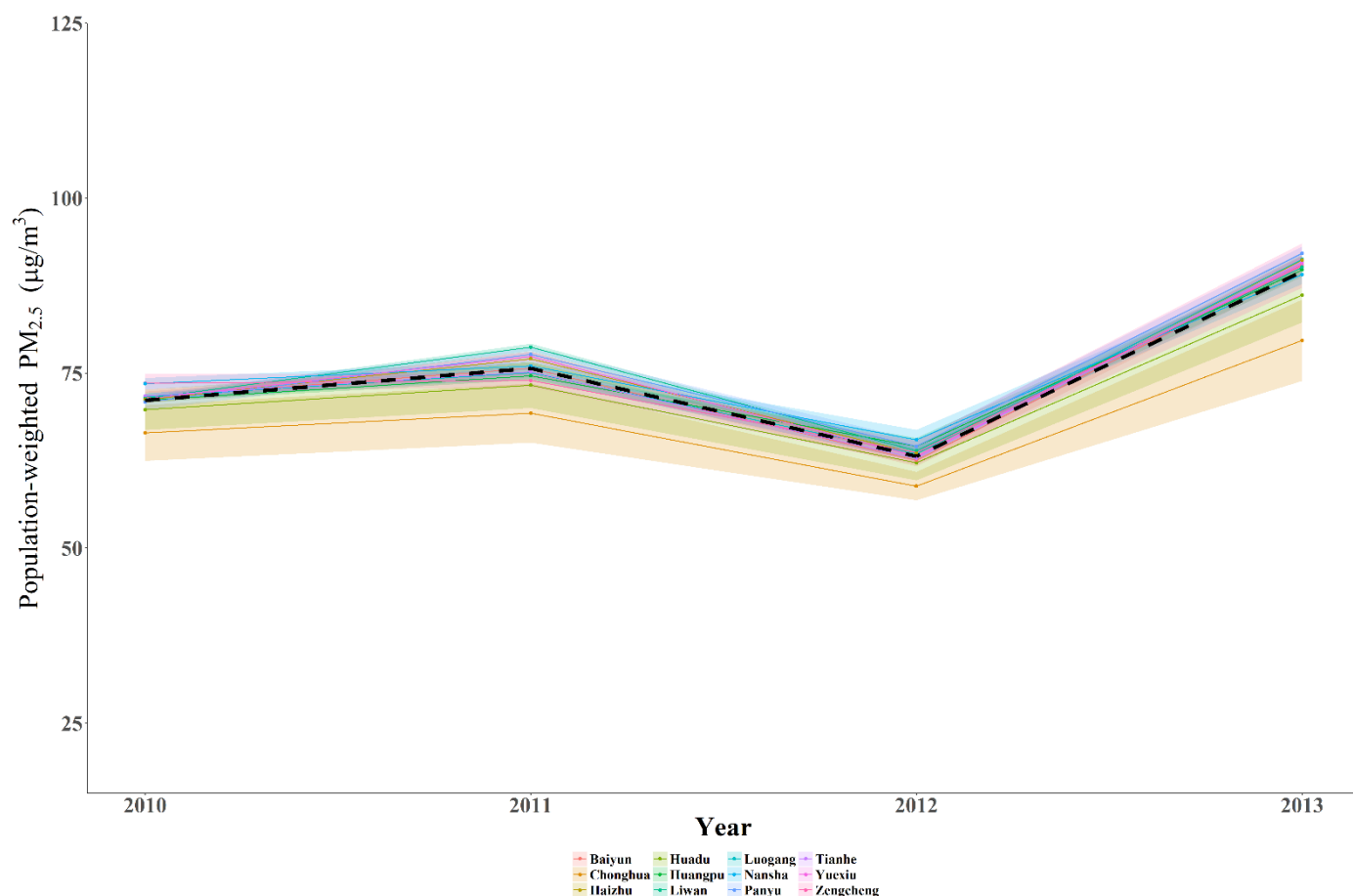


Figure S1. Trends in population-weighted mean concentrations of particulate matter with aerodynamic diameter $<2.5 \mu\text{m}$, by administrative district.

Note: PM_{2.5}, particulate matter with aerodynamic diameter $<2.5 \mu\text{m}$. Shaded areas represent the range of standard deviation (SD) of PM_{2.5} concentrations for each administrative district. Population-weighted mean concentrations of PM_{2.5} ($\mu\text{g}/\text{m}^3$) were calculated based on the Guangzhou population density with a resolution of $0.01 \times 0.01^\circ$ (or approximately 1 km^2 at the equator). The population density was first extracted from the Global Population Projection Grids (Jones and O'Neill, 2016) and provided a continuous surface of population counts covering the 12 districts of Guangzhou, but only with a grid cell resolution of $1/8 \times 1/8^\circ$ (or approximately 14 km^2 at the equator). Detailed information and validation of the population estimates can be extracted elsewhere (Jones and O'Neill, 2017). To obtain population counts at a more refined resolution, the gridded population data was then disaggregated and projected to match the grid cell resolution of 0.01 degree observed in the ambient PM_{2.5} remote image data, using the method of bilinear interpolation (Chang, 2006). Based on the derived population density, district-level as well as the city-level population-weighted mean concentrations of PM_{2.5} were estimated for the period 2010-2013.

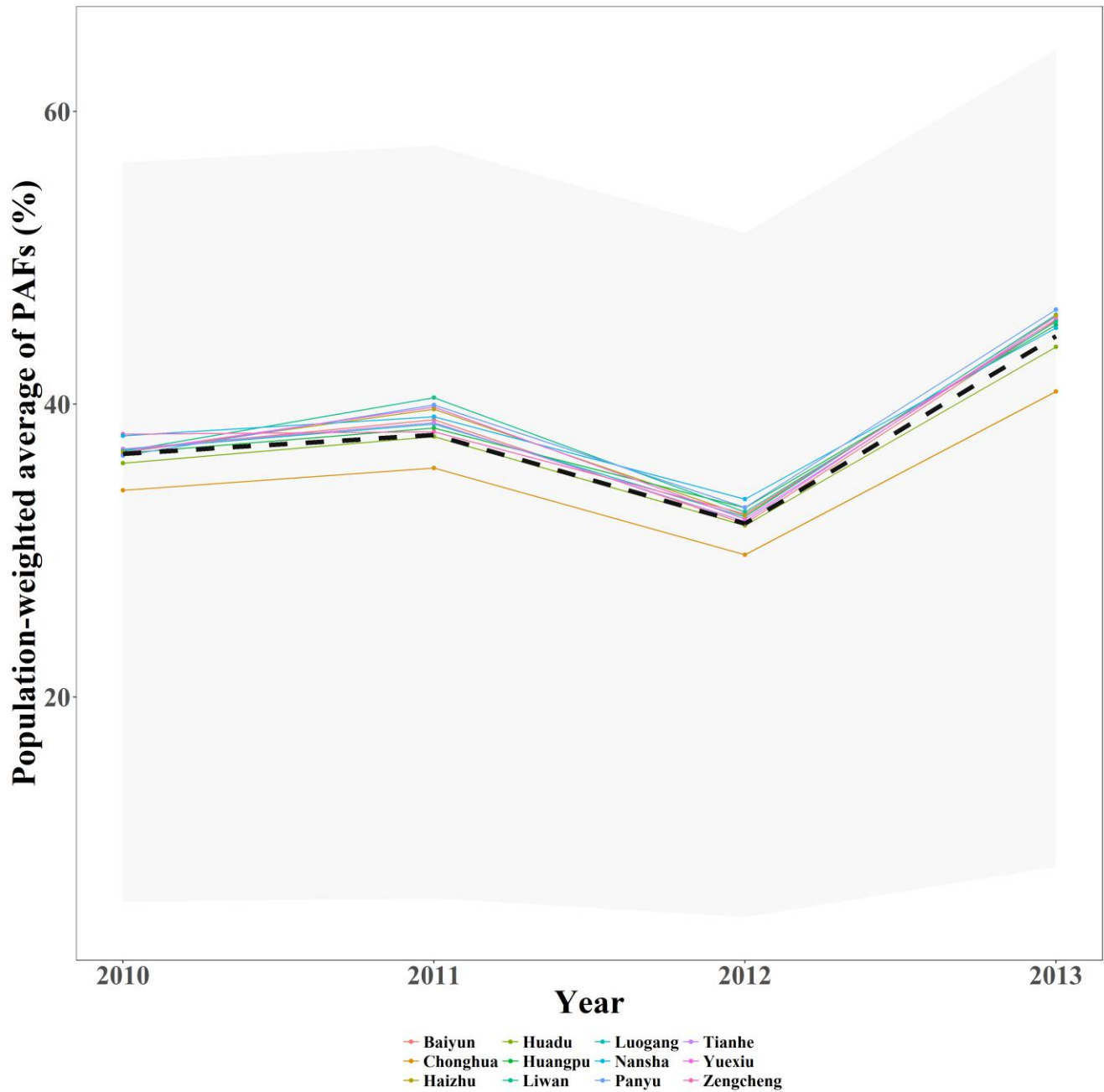


Figure S2. Trends in population-weighted average of PAFs in Guangzhou, by administrative district.

Note: PAF, population attributable fraction (%). It represents the proportion of lung cancer burden that would be reduced in a given year if the exposure to $PM_{2.5}$ in the past were reduced to the counterfactual concentration (defined as the theoretical minimum risk exposure level in the current study) (Gakidou et al., 2017).

In the above figure, black dash line indicates trends in the overall estimates of PAFs for the city of Guangzhou, and the shaded area represents the estimated uncertainty interval of the overall PAFs across the city for the period 2010-2013.

PAFs were weighted by the population counts estimated at $0.01 \times 0.01^\circ$ (or approximately 1 km^2 at the equator) resolution. The gridded data was extracted from the Global Population Projection Grids (Jones and O'Neill, 2016) and provided a continuous surface of population counts with a grid cell resolution of one-eighth degree (or

approximately 14 km^2 at the equator), covering the 12 administrative districts of Guangzhou. Gridded population data was then disaggregated and projected to match the grid cell resolution of 0.01 degree observed in the ambient $\text{PM}_{2.5}$ remote image data, using the method of bilinear interpolation (Chang, 2006). Detailed information and validation of the population estimates can be extracted elsewhere (Jones and O'Neill, 2017).

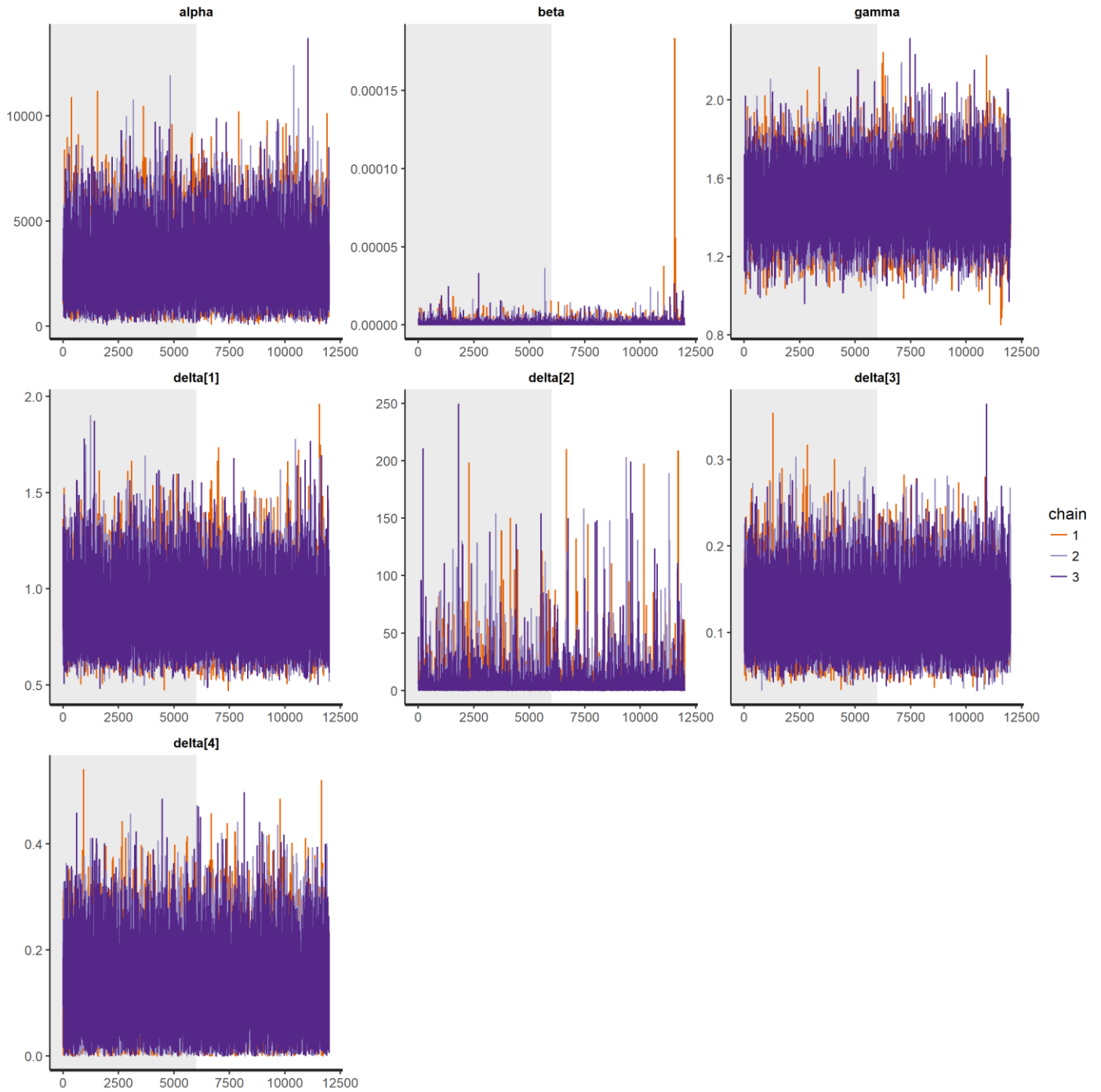


Figure S3. Mixing in the Monte Carlo Chains for the IER model.

Note: IER model, integrated exposure-response model as purposed by Burnett et al. (Burnett et al., 2014). We ran the IER model using the No-U-turn sampler (NUTS) algorithm purposed by Hoffman et al. (Hoffman and Gelman, 2014). To test stationarity (Lunn et al., 2012) of the stochastic IER model, we ran a total of three Markov Chains, each with an iteration of 12000 steps, and then we discarded the first 6000 iterations as burn-in (indicated by the shaded area in plots). The remaining samples reached convergence as indicated by $\hat{R}$ parameters of the initialization-fixed model in Table S2, and all three chains mixed well together.

References

- Bivand RS, Pebesma EJ, Gomez-Rubio V, Pebesma EJ. 2008. Applied spatial data analysis with R. New York, NY:Springer. [ISBN: 978-1-4614-7617-7].
- Burnett RT, Pope CAI, Ezzati M, Olives C, Lim SS, Mehta S, et al. 2014. An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure. *Environ Health Perspect* 122(4):397-403. <https://doi.org/10.1289/ehp.1307049>.
- Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. 2017. Stan: A probabilistic programming language. *J Stat Softw* 76(1). <http://dx.doi.org/10.18637/jss.v076.i01>.
- Chang K. 2006. Introduction to geographic information systems. Boston:McGraw-Hill Higher Education. [ISBN-13: 978-0077805401]
- Clements MS. 2005. Lung cancer rate predictions using generalized additive models. *Biostatistics* 6(4):576-589.
- Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. 2017. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* 389(10082):1907-1918.
- Foreman KJ, Lozano R, Lopez AD, Murray CJ. 2012. Modeling causes of death: an integrated approach using CODEm. *Popul Health Metr* 10:1.
- Gakidou E, Afshin A, Abajobir AA, Abate KH, Abbafati C, Abbas KM, et al. 2017. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* 390(10100):1345-1422.
- Hijmans RJ, van Etten J. 2014. raster: Geographic data analysis and modeling. Available: <https://rdr.io/cran/raster/>. [accessed 18 October 2018]
- Höfle H, Van Damme CJ, Fox C, Lelièvre S, Loots C, Nash RD, et al. 2017. Linking spawning ground extent to environmental factors—patterns and dispersal during the egg phase of four North Sea fishes. *Can J Fish Aquat Sci* 75(3):357-374.
- Hoffman MD, Gelman A. 2014. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 15(1):1593-1623.
- Jones B, O'Neill BC. 2016. Spatially Explicit Global Population Scenarios Consistent with the Shared Socioeconomic Pathways. *Environmental Research Letters*, 11(2016):84003.
- Jones B, O'Neill BC. 2017. Global Population Projection Grids Based on Shared Socioeconomic Pathways (SSPs), 2010-2100. Palisades, NY:NASA. Socioeconomic Data and Applications Center (SEDAC). <https://doi.org/10.7927/H4RF5S0P>.
- Lam NS. 1983. Spatial interpolation methods: a review. *The American Cartographer* 10(2):129-150.
- Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. 2012. The BUGS book: A practical introduction to Bayesian analysis. US:CRC press. [ISBN 9781584888499].
- Luo A, Li K, Li Y, Yang Z, Dong H, Yang Q, et al.. 2019. Spatial distribution of cancer-related burden in Guangzhou from 2010 to 2013. *Chinese Journal of Epidemiology*(10):1262-1268.
- Oliver MA, Webster R. 1990. Kriging: a method of interpolation for geographical information systems. *International Journal of Geographical Information System* 4(3):313-332.
- Oliver MA, Webster R. 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 113:56-69.
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4):583-639.
- Stan. 2016. Stan modeling language: User's guide and reference manual. Available: <http://mc-stan.org>. [accessed 09 September 2018]
- Tsutsumi M, Shimada A, Murakami D. 2011. Land price maps of Tokyo metropolitan area. *Procedia-Social and Behavioral Sciences* 21:193-202.

Vehtari A, Gelman A, Gabry J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 27(5):1413-1432.