Supplement of Atmos. Chem. Phys., 22, 16073–16090, 2022
https://doi.org/10.5194/acp-22-16073-2022-supplement

Atmospheric Chemistry and Physics
Open Access

Supplement of

Decoupling impacts of weather conditions on interannual variations in concentrations of criteria air pollutants in South China – constraining analysis uncertainties by using multiple analysis tools

Yu Lin et al.

Correspondence to: Leiming Zhang (leiming.zhang@ec.gc.ca) and Xiaohong Yao (xhyao@ouc.edu.cn)

Text S1. Description of the method calculating the range of the deweathered percentage change

To calculate the range of the deweathered percentage change (DePC) of an air pollutant in any given two years, five steps were designed. The hourly average $PM_{2.5}$ concentrations in Guangzhou from 2014 to 2020 were used as an example below (the code can be downloaded from https://pypi.org/project/DePC/).

Step 1: Construction of a dataset with equal size in each year

In each year from 2014 to 2020 (with i=1, 2, … and 7 representing the first, second, … and the seventh year), there are N_i data points of hourly $PM_{2.5}$ mass concentration. Rearrange all data points in each year from the smallest to the largest values to generate an array of data, $A_i(j)$, with j=1 to N_i.

For any two selected years:

$A_i(j)$ with j=1 to N_i and A_i' (j) with j=1 to N_i', set N_s = Min (N_i, N_i').

Then convert $A_i(j)$ with j=1 to N_i into $B_i(j)$ with j=1 to N_s, and convert A_i' (j) with j=1 to N_i' into $B_{i'}(j)$ with j=1 to N_s.

If $N_i= N_i'$, $B_i(j)=A_i(j)$ and $B_{i'}(j)=A_{i'}(j)$ with j=1 to N_s.

If $N_i > N_i'$ ($N_i'=N_s$), the difference (n) of N_i-N_s is calculated together with the quotient of N_i/n for filtering data.

If N_i is divisible by n, N_i/n equals to an integer as k. Then $B_i(j)$ is calculated as below:

1) with j=1 to k-2, $B_i(j) = A_i(j)$,

and with j=k-1, $B_i(k-1) = (A_i(k-1) + A_i(k))/2$;

2) with j=k to 2*k-3, $B_i(j) = A_i(j+1)$,

and with j=2*k-2, $B_i(2*k-2) = (A_i(2*k-1) + A_i(2*k))/2$;

…

n) with j=(n-1) *k-n+2 to n*k-n-1, $B_i(j) = A_i(j+n-1)$,

and with j=n*k-n. $B_i(n*k-n) = (A_i(n*k-1) + A_i(n*k))/2$.

Note that if N_i is not divisible by n, round it down to an integer as k, i.e., $k = [N_i/n]$.

Then use k_1=k and k_2=k+1 to calculate corresponding n_1 and n_2 to meet

a) $n_1+ n_2=n$;

b) $(k_1-1) *n_1+ (k_2-1) *n_2=N_s$.

If $n_1=n_2$, use k=k_1 in 1) and k=k_2 in 2) and repeat the replacement through the conversion, else choose $n'=|n_1-n_2|$.

There are two scenarios,

a) when $n_1>n_2$, use k=k_1 in 1), 2), …, n'), n'+1), n'+3), …, n-1) and k=k_2 in n'+2), n'+4), …, n) to process the replacement;

b) when $n_1<n_2$, use k=k_2 in 1), 2), …, n'), n'+1), n'+3), …, n-1) and k=k_1 in n'+2), n'+4), …, n) to process the replacement.

For example, N_1=N_s=7719, B_1(j)=A_1(j) with j=1 to 7719. Convert A_7(j) with j=1 to N_7 (8391) into B_7(j) with j=1 to 7719. Since N_7>N_s, the difference (n=672) of N_7-N_s is calculated with the quotient of N_7/n (≈12.5). Therefore, two parameters, i.e., k_1=k=12 and k_2=k+1=13, are used to process the replacement. Since (12-1) *345+(13-1) *327=7719 and 345+327=672, i.e., n_1=327, n_2=345, n'=18, corresponding B_7(j) can be gained as below:

1) with j=1 to 10, B_7(j) = A_7(j),

and with j=11, B_7 (11) = (A_7 (11) + A_7 (12))/2;

2) with j=12 to 22, B_7(j) = A_7(j+1),

and with j=23, B_7 (23) = (A_7 (24) + A_7 (25))/2;

…

37) with j=415 to 424, B_7(j) = A_7(j+36),

and with j=425, B_7 (425) = (A_7 (461) + A_7 (462))/2;

38) with j=426 to 435, B_7(j) = A_7(j+37),

and with j=436, B_7 (436) = (A_7 (473) + A_7 (474))/2;

39) with j=437 to 447, B_7(j) = A_7(j+38),

and with j=448, B_7 (448) = (A_7 (486) + A_7 (487))/2;

…

671) with j=7697 to 7706, B_7(j) = A_7(j+670),

and with j=7707, B_7 (7707) = (A_7 (8377) + A_7 (8378))/2;

672) with j=7708 to 7718, B_7(j) = A_7(j+671),

and with j=7719. B_7 (7719) = (A_7 (8390) + A_7 (8391))/2.

The above approach of adjusting the dataset size would not significantly alter the statistical metrics of the original data set. For example, Table S1 lists the comparison of statistics metrics between A_7 (with 8391 original hourly data points) and B_7 (with 7719 adjusted data points), which are highly consistent.

Step 2: Preliminary identification of outlier data points for $B_i(j)$ against $B_i'(j)$

Linear regression (LR) is applied between any two years of a data array, i.e., B_i(j) against B_i'(j). The infection point is visibly identified from the regression curve as the first guess and all data point having a value larger than the inflection point are considered as outliers. Thus, the data volume of B_i(j) changes from N_s to $N_{first\ guess}$. In this case, we continue to use B_7 against B_1 of $PM_{2.5}$ concentration in Guangzhou as an example.

Fig. S1 shows the regression curve of $PM_{2.5}$ data in 2020 against 2014, with the regression equation of y=0.71*x-3.57, R^2=0.981, N_s=7719. The infection point can be visibly identified at x≈119 μg·m^{-3}, corresponding to $N_{first\ guess}$=7637.

The data larger than infection point does not follow the general trend. The part of the data only account for <5% of the total data sets, occur less frequently and thereby cannot repeatedly occur between the two years. They should be statistically treated as outliers caused by occasional factors such as poor dispersion conditions and/or accident emissions since they deviate from the general trend in data comparison between the two years.

Step 3: Secondary identification of outlier data points for $B_i(j)$ against $B_i'(j)$

LR with zero interception is conducted using data in $B_i(j)$ against $B_i'(j)$ with j=1 to $N_{first\ guess}$. The regression equation is derived as $B_i(j)=k_a*B_i'(j)$, in which k_a is a LR slope. The regression equation is used to predict concentrations in different percentiles and reconstructed $C_i(j)=k_a*B_i'(j)$, j=1 to $N_{first\ guess}$. The absolute values of the quotient of $(C_i(j)-B_i(j))/B_i(j))$ are then calculated. Set j from $N_{first\ guess}$ to 1 and search for the absolute quotient of $(C_i(j)-B_i(j))/B_i(j))$ ≤5% to be first gained, and the corresponding j is defined as N_{final}. Here 5% is set as the threshold since analytic errors of these on-line instruments are at least 5% or even larger.

For the data of the aforementioned example, the regression equation is derived as y=0.63*x, R^2=0.998. The regression equation is used to calculate concentrations in different percentiles and reconstructed $C_7(j)=0.63*B_1(j)$, j=1 to $N_{first\ guess}$ (7637). Set j from 7637 to 1 and search for the absolute quotient of $(C_7(j)-B_7(j))/B_7(j))$ ≤5% to be first gained. Mark the corresponding j as N_{final}, i.e., j=N_{final}=7599 and $B_7(7599)$ ≈73 μg·m^{-3}.

Throughout Steps 1–3, 84%–99% of the observational data were left to reconstruct the final arrays for a given pollutant between any two consecutive years of the same size. The remaining data size are large enough to represent the statistic property for a given pollutant in each year.

Step 4: Calculation of the closed interval of the DePC

LR with zero interception is conducted using data in $B_i(j)$ array against $B_i'(j)$ with j=1 to N_{final}. The regression equation is derived as $B_i(j)=k_b*B_i'(j)$, where k_b is the LR slope. The $DePC_i$ is set as k_b minus the one. The $DePC_i$ is referred as the primary $DePC_i$ ($DePC_i$-primary) in this study.

Again, for the data of the aforementioned example, the regression equation is derived as y=0.63*x, R^2=0.998. The primary $DePC_{2020\text{-}2014}$ is calculated as -37%, with 7599 out of the total of 7719 data to be used.

LR with zero interception is conducted using data in $B_i(j)$ array against $B_i'(j)$ with j=1 to N_{final}. The regression equation is derived as $B_i(j)=k_c*B_i'(j)+b_i$, in which k_c and b_i are the LR slope and interception, respectively. Set k_c minus one as the secondary $DePC_i$ ($DePC_{i-secodnary}$).

For the data of the aforementioned example, LR with non-zero interception is derived as y=0.67*x-2.06, R^2=0.994. The secondary $DePC_{2020-2014}$ is calculated as -33%. Thus, the estimated DePC in 2020 against 2014 is in the range of -37% (primary) and -33% (secondary). The small difference between the primary DePC and the secondary DePC indicates that the perturbation from varying weather conditions have mostly been removed. The R^2 values of the LR analysis for calculating both $DePC_{secondary}$ and $DePC_{primary}$ or calculating $DePC_{secondary}$ were mostly larger than 0.99. The true DePC should be between $DePC_{primary}$ and $DePC_{secondary}$, although this must be confirmed mathematically.

Step 5: Evaluation of residual perturbation by varying weather conditions

Theoretically, the intercept (b_i) should infinitely approach zero when the perturbation by varying weather conditions is eliminated. If $DePC_{i-secondary}$ is smaller than $DePC_{i-primary}$, the perturbation incompletely removed by Step 3 likely disfavors the accumulation and/or formation of the pollutant. The reverse would be true if $DePC_{i-secondary}$ was larger than $DePC_{i-primary}$. Thus, both $DePC_{i-primary}$ and $DePC_{i-secondary}$ were used in this study to construct the closed interval of DePC, i.e., [$DePC_{i-primary}$, $DePC_{i-secondary}$]. In the example, the closed interval range was [-37%, -33%].

Table S1. Information of the air quality monitoring stations and meteorological monitoring sites used in this study (/ represents elevations of air quality monitoring stations were unavailable).

City	ID	Site name	Site type	Latitude	Longitude	Elevation(m)
Sanya	1876A	Hedong Station	urban	18.249	109.508	/
	1877A	Hexi Station	urban	18.268	109.496	/
	574941	SANYA PHOENIX INTL	meteorological	18.303	109.412	28.04
Zhuhai	1367A	Jida	urban	22.262	113.581	/
	1368A	Qianshan	urban	22.234	113.488	/
	1369A	Tangjia	suburban	22.425	113.628	/
	1370A	Doumen	rural	22.228	113.299	/
	594880	ZHUHAI SANZAO	meteorological	22.017	113.383	3.00
Haikou	1409A	Dongzhai Harbor	background	19.951	110.576	/
	1410A	Hainan University	suburban	20.057	110.323	/
	1411A	Xiuying Hainan General hospital	urban	20.0053	110.283	/
	1412A	Hainan Normal University	suburban	19.997	110.338	/
	1413A	Environmental Protection Agency Dormitory of Longhua Road	urban	20.036	110.330	/
	470311	MEILAN	meteorological	19.935	110.459	22.86
Zhanjiang	1680A	Cinema of Zhanjiang	urban	21.271	110.354	/
	1681A	Municipal Environmental Monitoring Station	urban	21.223	110.393	/
	1682A	Environmental Protection Agency Dormitory	urban	21.200	110.402	/
	1683A	Xiashan Swimming Pool	suburban	21.203	110.411	/
	1684A	Mazhang District Environmental Protection Bureau	suburban	21.268	110.332	/
	1685A	Potou District Environmental Protection Bureau	suburban	21.257	110.456	/
	596580	ZHANJIANG	meteorological	21.217	110.400	28.00

Guangzhou	1345A	Guangya Middle School	urban	23.142	113.235	/
	1346A	No. 5 Middle School	urban	23.105	113.261	/
	1347A	Tianhe Vocational School	urban	23.147	113.329	/
	1348A	Guangdong Business School	urban	23.092	113.348	/
	1349A	No. 86 Middle School	urban	23.105	113.433	/
	1350A	Panyu Middle School	suburban	22.948	113.351	/
	1351A	Huadu Normal University	urban	23.392	113.215	/
	1352A	City Monitoring Station	urban	23.133	113.260	/
	1353A	Zhenlong of Kowloon Town	suburban	23.312	113.5618	/
	1354A	Luxelakes	suburban	23.154	113.277	/
	1355A	Maofengshan Forest Park	background	23.3035	113.443	/
	592870	BAIYUN INTL	meteorological	23.392	113.299	15.24
Shenzhen	1356A	Tongxinling Station	urban	22.555	114.106	/
	1357A	Honghu	urban	22.563	114.117	/
	1358A	Overseas Chinese Town	urban	22.542	113.987	/
	1359A	Nanhai Station	urban	22.517	113.918	/
	1360A	Yantian	urban	22.569	114.246	/
	1361A	Longgang	suburban	22.724	114.228	/
	1362A	Xixiang	suburban	22.586	113.895	/
	1363A	Nan’ao	rural	22.542	114.494	/
	1364A	Kuiyong	suburban	22.634	114.410	/
	1365A	Meisha	urban	22.598	114.297	/
	1366A	Guanlan	suburban	22.738	114.068	/
	594930	BAOAN INTL	meteorological	22.639	113.811	3.96

Table S2. Summary of collected observational dataset, the order of the cities was same as Table 1.

Pollutant	City	Data number	Missing
$PM_{2.5}$	Guangzhou	58677 (96%)	2379 (4%)
	Shenzhen	58695 (96%)	2361 (4%)
	Zhanjiang	57780 (95%)	3276 (5%)
	Zhuhai	57872 (95%)	3184 (5%)
	Haikou	58680 (96%)	2376 (4%)
	Sanya	57497 (94%)	3559 (6%)
PM_{10}	Guangzhou	58669 (96%)	2387 (4%)
	Shenzhen	58673 (96%)	2383 (4%)
	Zhuhai	54626 (89%)	6430 (11%)
	Zhanjiang	52024 (85%)	9032 (15%)
	Haikou	58613 (96%)	2443 (4%)
	Sanya	56070 (92%)	4986 (8%)
O_3	Zhuhai	57789 (95%)	3267 (5%)
	Zhanjiang	57729 (95%)	3327 (5%)
	Shenzhen	58694 (96%)	2362 (4%)
	Haikou	58676 (96%)	2380 (4%)
	Sanya	57497 (94%)	3559 (6%)
	Guangzhou	58675 (96%)	2381 (4%)
NO_2	Guangzhou	58676 (96%)	2380 (4%)
	Shenzhen	58694 (96%)	2362 (4%)
	Zhuhai	57809 (95%)	3247 (5%)
	Zhanjiang	56329 (92%)	4727 (8%)
	Haikou	58679 (96%)	2377 (4%)
	Sanya	57569 (94%)	3487 (6%)
CO	Guangzhou	58676 (96%)	2380 (4%)
	Zhanjiang	57791 (95%)	3265 (5%)
	Shenzhen	58694 (96%)	2362 (4%)
	Zhuhai	57702 (95%)	3354 (5%)
	Haikou	58679 (96%)	2377 (4%)
	Sanya	57678 (94%)	3378 (6%)
SO_2	Guangzhou	58676 (96%)	2380 (4%)
	Zhanjiang	57818 (95%)	3238 (5%)
	Shenzhen	58694 (96%)	2362 (4%)
	Zhuhai	57819 (95%)	3237 (5%)
	Haikou	58678 (96%)	2378 (4%)
	Sanya	57695 (94%)	3361 (6%)
Average		57869 (95%)	3187 (5%)
Summary		2083292 (95%)	114724 (5%)

Table S3. The top three influential meteorological factors in each RF and BRTs model.

city	pollutant / method	$PM_{2.5}$	PM_{10}	O_3	NO_2+O_3	NO_2	CO	SO_2
Guangzhou	RF	dp, sp, at	sp, at, rh	ssr, at, rh	ssr, rh, at	at, ap, sp	dp, at, sp	rh, ssr, at
	BRTs	sp, rh, dp	sp, rh, blh	ssr, at, rh	blh, ssr, rh	blh, dp, sp	sp, dp, blh	rh, ssr, wd
Shenzhen	RF	sp, dp, rh	sp, dp, rh	ssr, dp, at	ssr, at, sp	at, blh, dp	dp, sp, at	rh, dp, tp
	BRTs	sp, rh, dp	rh, dp, tp	ssr, tp, dp	ssr, tp, dp	blh, wddp	dp, wd, sp	rh, tp, blh
Zhuhai	RF	sp, dp, rh	sp, dp, rh	dp, ssr, rh	ssr, at, rh	dp, at, sp	dp, at, sp	sp, rh, at
	BRTs	sp, dp, rh	rh,sp,dp	ssr, rh, at	dp, ssr, rh	dp, at, blh	dp, wd, sp	rh, blh, tp
Zhanjiang	RF	sp, tp, ssr	sp, tp, ssr	sp, ssr, tcc	sp, ssr, tp	sp, ssr, blh	sp, tp, dp	ssr, sp, tp
	BRTs	sp, tp, rh	sp, tp, blh	sp, ssr, tcc	sp, tcc, tp	sp, blh, tcc	sp, blh, tcc	blh, sp, tp
Haikou	RF	at, dp, sp	at, dp, sp	sp, dp, ssr	sp, dp, at	sp, dp, blh	ssr, sp, dp	ssr, rh, dp
	BRTs	rh, dp, tp	tp, dp, rh	sp, dp, ssr	sp, dp, ssr	dp, wd, blh	ssr, sp, wd	dp, blh, sp
Sanya	RF	rh, dp, tp	sp, rh, dp	dp, sp, ssr	dp, sp, rh	ssr, at, sp	dp, sp, at	sp, ssr, rh
	BRTs	rh, dp, tp	tp, blh, dp	dp, blh, rh	dp, blh, ssr	wd, at, ssr	dp, at, sp	blh, sp, tcc

at: air temperature, blh: boundary layer height, dp: dew point, rh: relative humidity,
ssr: surface net solar radiation, sp: surface pressure, tcc: total cloud cover, tp: total precipitation, wd: wind direction

Table S4. The performance of RF and BRTs methods, the order of the cities was same as Table 1.

indices		R^2		RMSE		MB		MFB		MFE	
pollutant	city	RF	BRTs	RF	BRTs	RF	BRTs	RF	BRTs	RF	BRTs
$PM_{2.5}$	Guangzhou	0.93	0.91	12.10	12.29	-0.22	0.01	7%	7%	23%	29%
	Shenzhen	0.95	0.91	7.78	9.18	0.08	-0.13	6%	6%	20%	27%
	Zhanjiang	0.88	0.85	11.63	12.79	0.14	-0.24	13%	10%	29%	35%
	Zhuhai	0.94	0.86	9.66	12.29	0.02	-0.13	10%	12%	25%	38%
	Haikou	0.93	0.88	7.78	8.21	0.26	-0.01	8%	8%	20%	30%
	Sanya	0.93	0.87	5.72	6.48	-0.09	0.00	9%	11%	27%	34%
PM_{10}	Guangzhou	0.93	0.91	17.72	19.32	-0.26	0.26	6%	7%	22%	28%
	Shenzhen	0.95	0.92	12.94	14.55	-0.09	-0.22	5%	5%	19%	25%
	Zhuhai	0.92	0.90	17.84	16.24	-0.42	0.16	8%	9%	27%	30%
	Zhanjiang	0.88	0.87	19.57	18.24	-0.09	-0.17	10%	9%	32%	33%
	Haikou	0.94	0.91	11.12	12.30	0.34	0.21	7%	7%	20%	27%
	Sanya	0.92	0.89	10.00	11.11	-0.02	-0.05	7%	8%	25%	30%
O_3	Zhuhai	0.93	0.90	19.51	25.41	0.04	-0.12	13%	8%	26%	30%
	Zhanjiang	0.90	0.89	23.02	26.13	0.27	0.26	16%	10%	35%	34%
	Shenzhen	0.94	0.91	15.02	21.07	0.46	0.02	10%	7%	22%	28%
	Haikou	0.95	0.93	12.89	18.60	0.34	0.32	9%	7%	22%	26%
	Sanya	0.95	0.92	15.29	18.97	-0.07	0.05	11%	10%	27%	31%
	Guangzhou	0.90	0.88	20.29	24.60	0.68	0.19	23%	13%	42%	51%
NO_2+O_3	Zhuhai	0.97	0.96	18.63	22.22	0.55	-0.42	6%	3%	14%	16%
	Guangzhou	0.96	0.95	21.48	25.36	0.83	0.01	6%	3%	16%	18%
	Shenzhen	0.98	0.96	14.72	19.72	0.71	0.00	4%	3%	12%	16%
	Zhanjiang	0.94	0.92	22.84	26.01	0.55	-0.42	10%	6%	25%	25%
	Haikou	0.97	0.95	11.99	18.71	0.15	-0.20	5%	4%	15%	20%
	Sanya	0.95	0.92	13.31	16.79	0.17	-0.05	4%	4%	15%	20%
NO_2	Guangzhou	0.87	0.88	24.66	23.98	1.24	-0.06	14%	11%	39%	38%
	Shenzhen	0.89	0.89	14.50	14.33	0.88	0.08	13%	10%	36%	35%
	Zhuhai	0.80	0.81	19.92	19.66	1.06	-0.32	23%	18%	52%	51%
	Zhanjiang	0.81	0.83	11.21	10.82	1.17	-0.06	27%	20%	52%	49%
	Haikou	0.86	0.89	9.01	8.64	0.43	-0.19	18%	13%	43%	40%
	Sanya	0.87	0.89	7.77	7.64	0.45	0.00	22%	18%	51%	50%
CO	Guangzhou	0.95	0.96	0.25	0.25	0.02	0.00	5%	4%	20%	19%
	Zhanjiang	0.92	0.94	0.26	0.23	0.01	0.00	7%	5%	24%	22%
	Shenzhen	0.97	0.97	0.16	0.15	0.01	0.00	4%	2%	16%	15%
	Zhuhai	0.93	0.95	0.20	0.19	0.01	0.00	7%	5%	24%	22%
	Haikou	0.92	0.95	0.44	0.42	0.05	0.00	13%	6%	26%	22%
	Sanya	0.95	0.96	0.14	0.13	0.00	0.00	4%	3%	20%	19%

SO_2	Guangzhou	0.91	0.92	4.88	4.74	0.35	0.07	10%	8%	30%	29%
	Zhanjiang	0.78	0.82	7.53	7.23	0.56	-0.04	26%	20%	51%	46%
	Shenzhen	0.94	0.95	2.40	2.34	0.11	0.03	6%	4%	21%	21%
	Zhuhai	0.83	0.85	4.93	4.77	0.40	0.03	22%	16%	47%	44%
	Haikou	0.89	0.91	2.96	2.86	0.17	0.04	11%	8%	29%	27%
	Sanya	0.86	0.88	4.59	4.52	0.17	0.03	18%	13%	34%	32%

Table S5. Statistical comparison between raw array (A_7) and reconstructed array (B_7) for hourly average $PM_{2.5}$ concentrations in Guangzhou in 2020.

Statistical parameters		Raw array A_7	Reconstructed array B_7
Volume		8391	7719
Mean		28	28
Median		24	24
Mode		11	11
Standard deviation		18	18
Skewness		1.80	1.77
Kurtosis		6.64	6.43
Minimum		1	1
Maximum		153	153
Percentile	10	10	10
	25	15	15
	75	38	38
	90	50	50

Table S6. The exceedance in percentages of six criteria air pollutants in six cities in 2020 according to the Class-I levels of AAQS in China. * represents the annual averages of air pollutants concentrations exceed the corresponding values of the Chinese AAQS (Class-I levels), else were represented by -.

Pollutant	Averaging time	Chinese AAQS (Class-I levels)	Exceed percentage (%)					
			Guangzhou	Zhanjiang	Shenzhen	Zhuhai	Haikou	Sanya
$PM_{2.5}$	Annual	15($\mu g \cdot m^{-3}$)	*	*	*	*	*	-
	24-hour	35($\mu g \cdot m^{-3}$)	31	25	16	15	10	2
PM_{10}	Annual	40($\mu g \cdot m^{-3}$)	*	*	*	*	-	-
	24-hour	50($\mu g \cdot m^{-3}$)	45	27	33	32	14	3
O_3	8-hour	100($\mu g \cdot m^{-3}$)	44	19	15	17	18	8
	1-hour	160($\mu g \cdot m^{-3}$)	7	2	2	3	2	0
NO_2	Annual	40($\mu g \cdot m^{-3}$)	*	-	-	-	-	-
	24-hour	80($\mu g \cdot m^{-3}$)	3	0	1	2	0	0
CO	24-hour	4($mg \cdot m^{-3}$)	0	0	0	0	0	0
	1-hour	10($mg \cdot m^{-3}$)	0	0	0	0	0	0
SO_2	24-hour	50($\mu g \cdot m^{-3}$)	0	0	0	0	0	0
	1-hour	150($\mu g \cdot m^{-3}$)	0	0	0	0	0	0

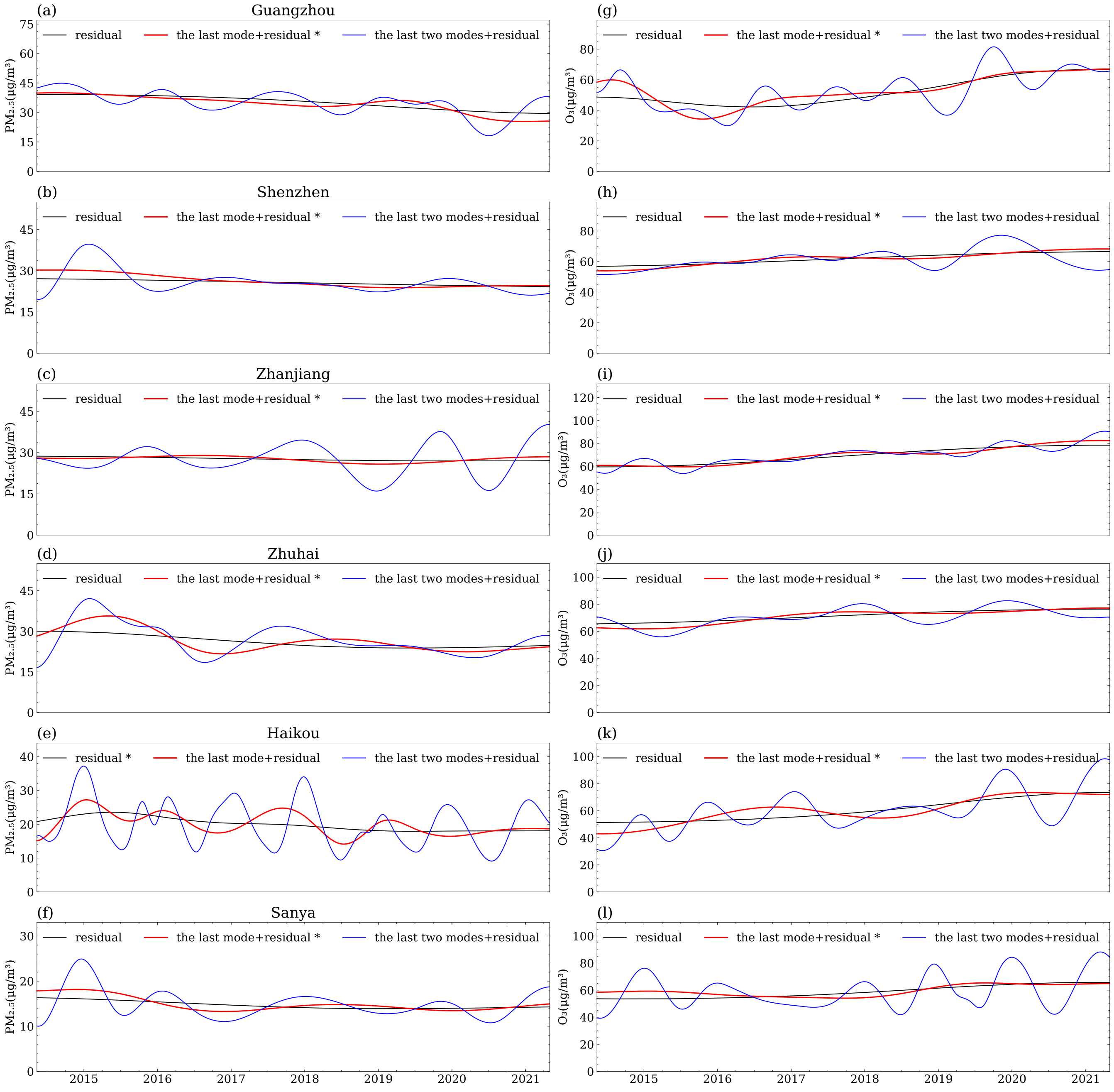

Figure S1: The ICEEMDAN-decomposed residuals, (the last mode + residual) and (the last two modes + residual) of $PM_{2.5}$ (a-f) and O_3 (g-l) from May 2014 to April 2021 in the six cities; * represents the time series of values to be used to calculate PC; the order of the cities was same as Figure 4.

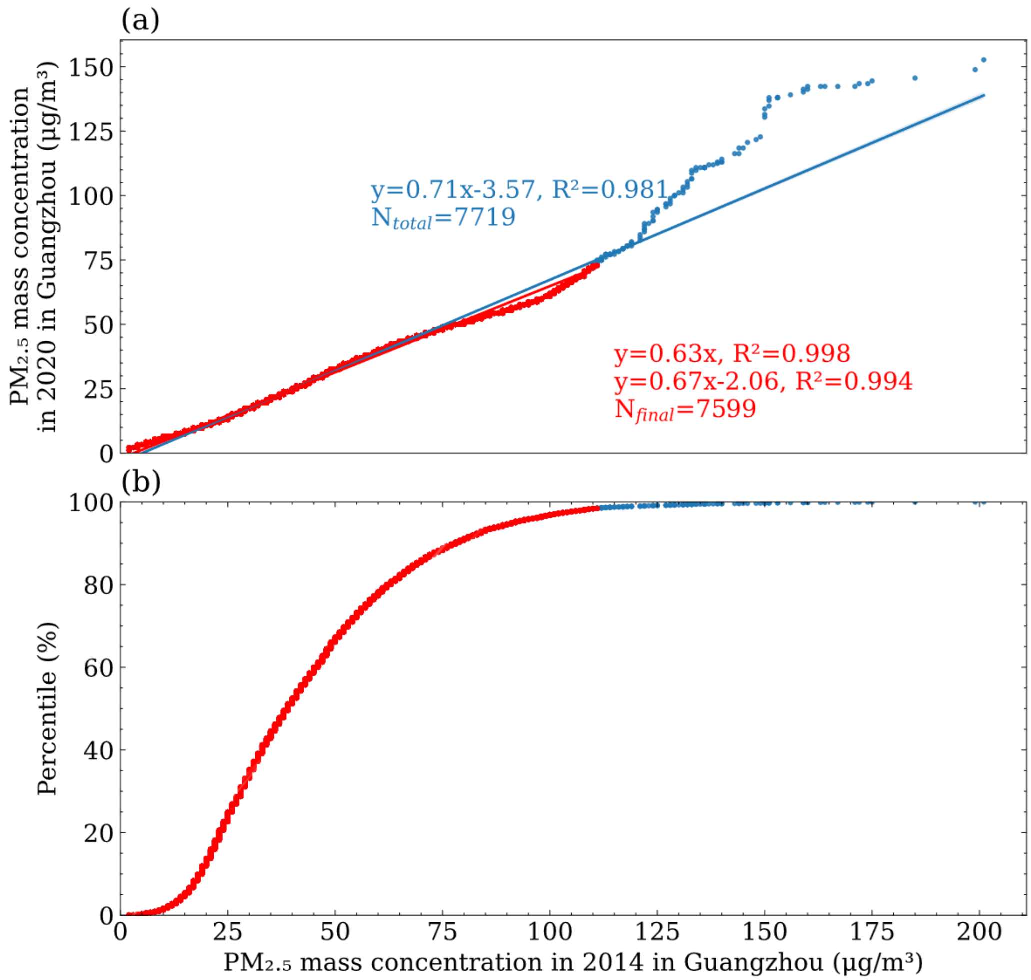

Figure S2: Analysis results of $PM_{2.5}$ mass concentrations in Guangzhou (2020 vs. 2014) in pairs of arrays reconstructed by Step 1 (a: reconstructed $PM_{2.5}$ concentrations in 2020 vs 2014; b: accumulation percentile of concentration in 2014; blue markers and regression curves use all data points, and red ones use the selected data by excluding data points suffered from severe perturbations from the anomalies).

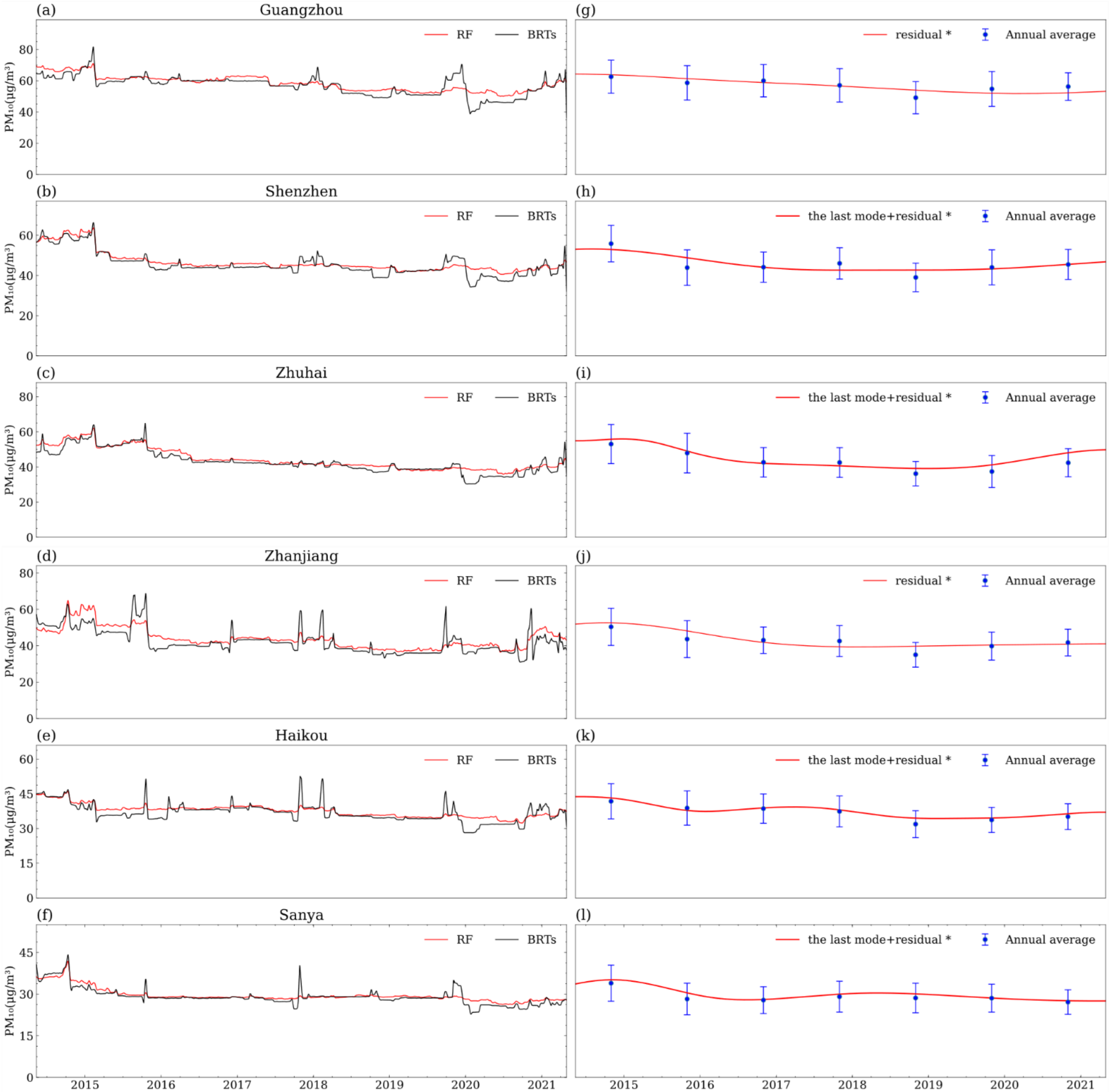

Figure S3. The RF-deweathered and BRTs-deweathered concentrations, ICEEMDAN-decomposed residuals (or mode + residuals) of PM_{10} and annual averages from May 2014 to April 2021 (a-f: deweathered concentrations in the six cities (the order of the cities was same as listed in Table 1); g-l: decomposed residual or (the last mode + residual) and annual averages plus one-third standard deviation in the six cities; * in g-l represents the time series of values to be used to calculate the trend and PC).

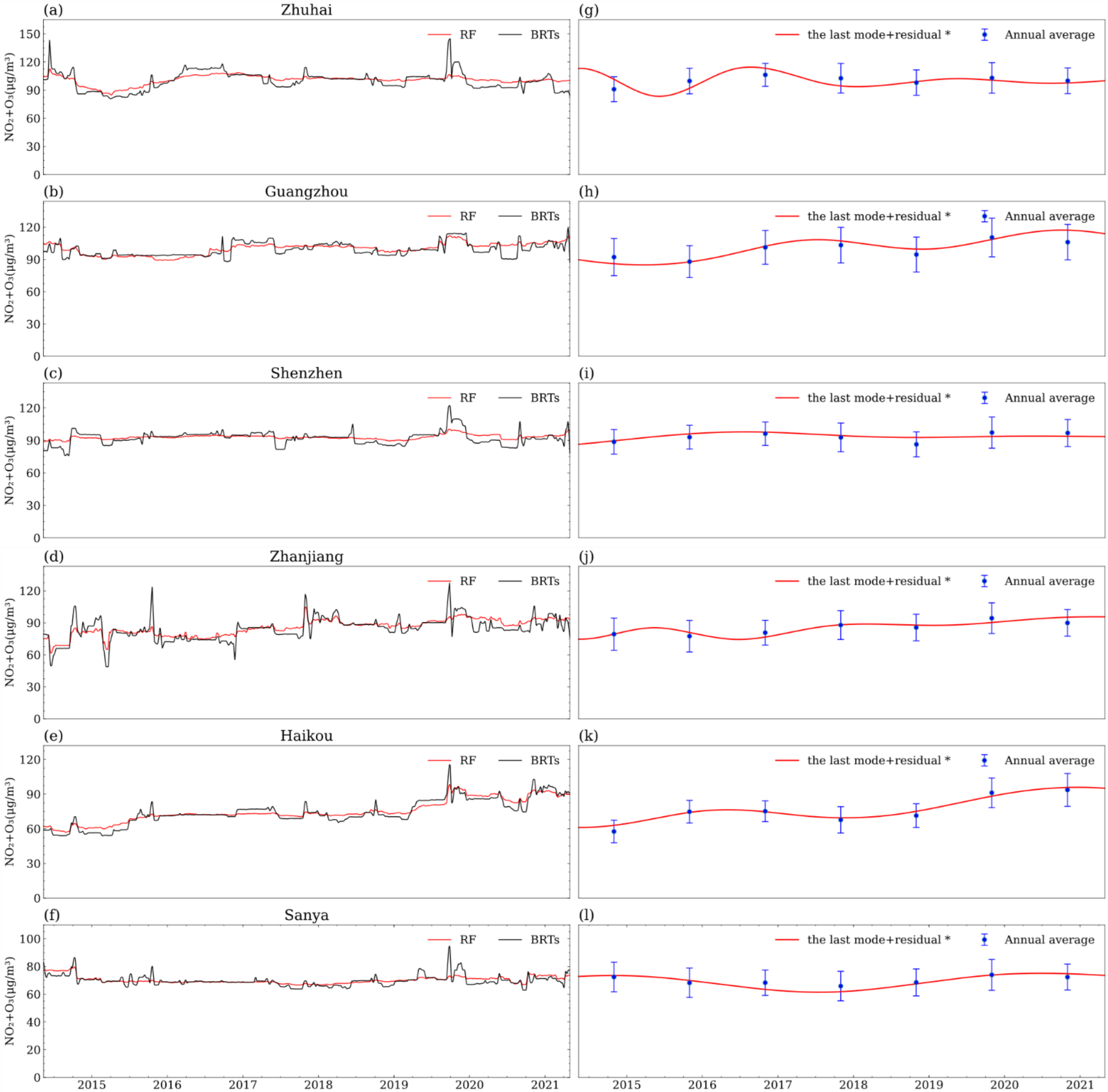

Figure S4. Same as Figure S3, except the pollutant to be (NO_2+O_3).

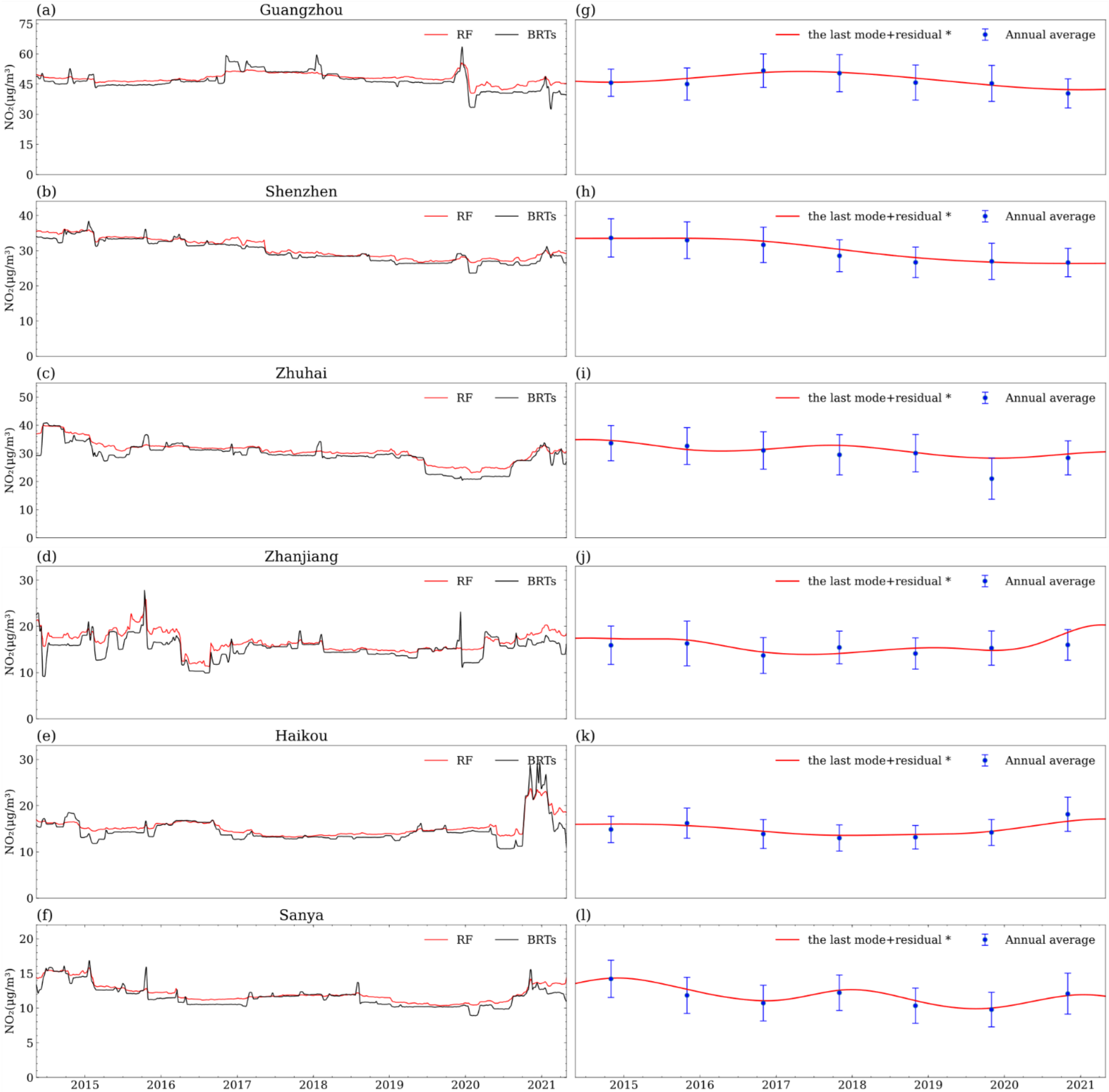

Figure S5. Same as Figure S3, except the pollutant to be NO_2.

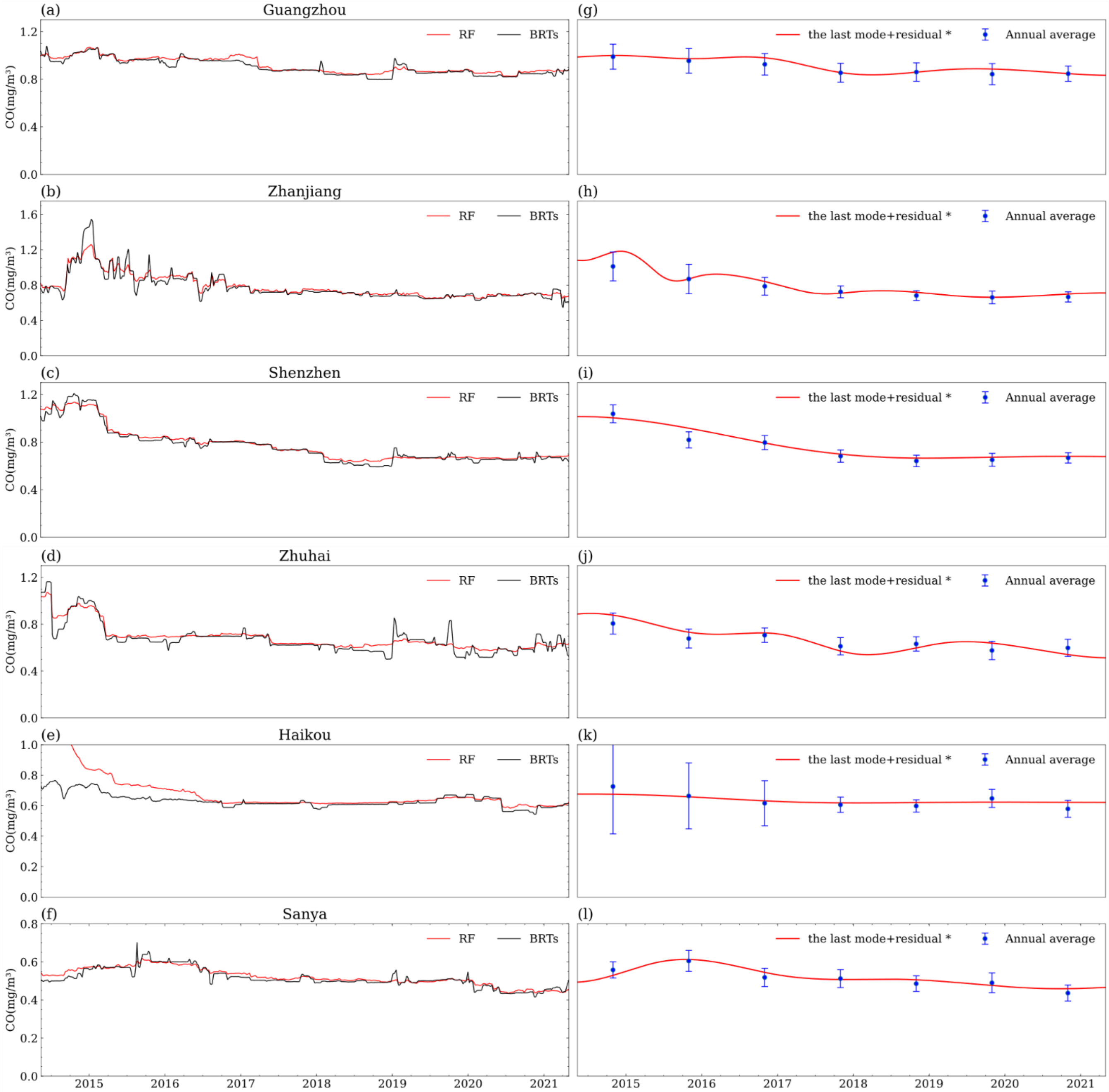

Figure S6. Same as Figure S3, except the pollutant to be CO.

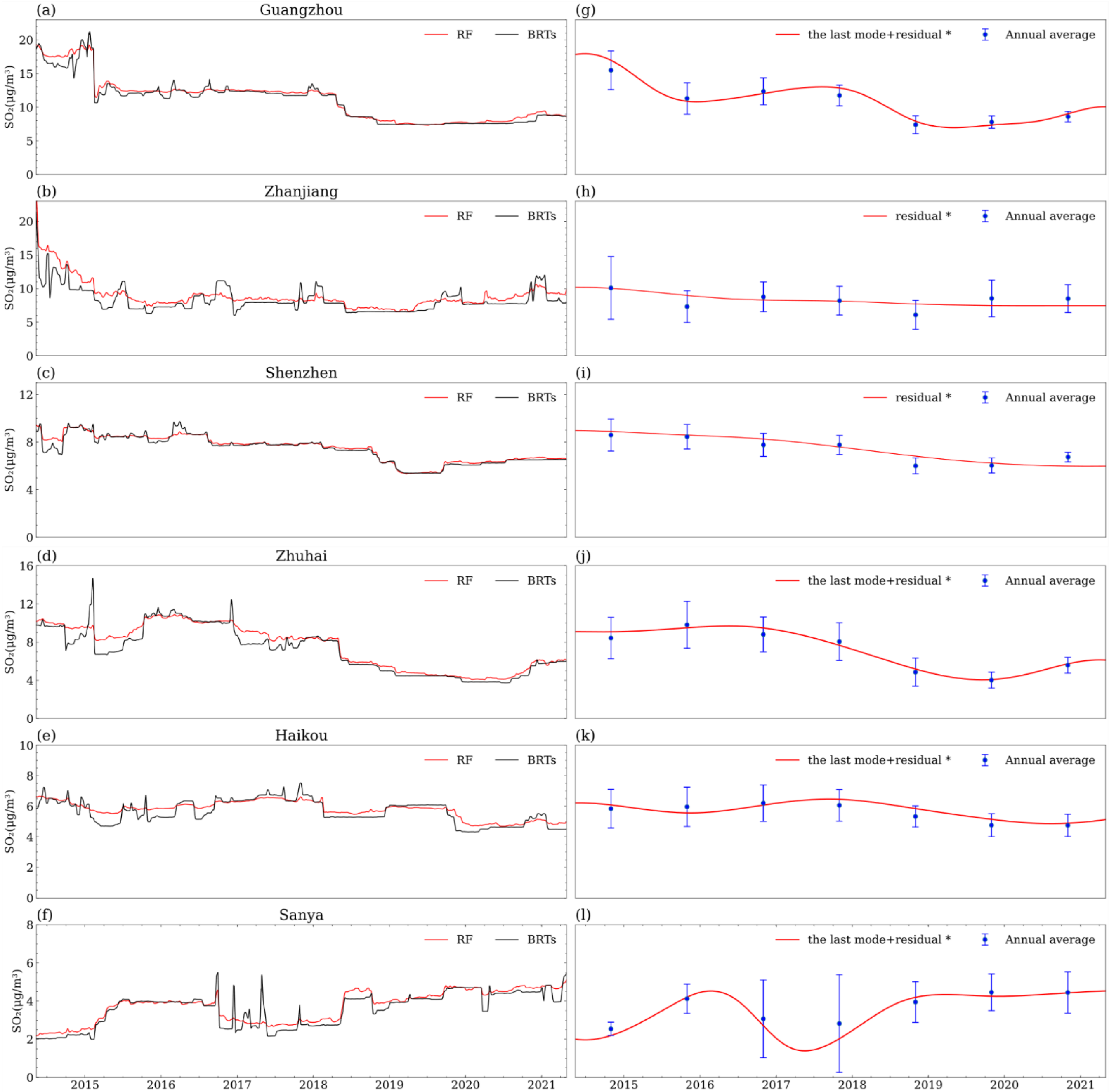

Figure S7. Same as Figure S3, except the pollutant to be SO_2.