Supporting Information for Improving Imputation of Missing $PM_{2.5}$ Speciation Data Using PMF-Informed Source–Receptor Relationships

Wubin Zhu[1], Mingjie Xie[2], Qili Dai[*1,3], Xiaohui Bi[1], Yufen Zhang[1], and Yinchang Feng[1]

[1]State Environmental Protection Key Laboratory of Urban Ambient Air Particulate Matter Pollution Prevention and Control, College of Environmental Science and Engineering, Nankai University, Tianjin 300350, China
[2]Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, School of Environmental Science and Engineering, Nanjing University of Information Science & Technology, 219 Ningliu Road, Nanjing, 210044, China
[3]Tianjin Key Laboratory of Software Experience and Human Computer Interaction, Tianjin 300457, China

*Email: daiql.nankai.edu.cn

Content

*Corresponding author. Email: daiql@nankai.edu.cn

introduction Additional information for generating missing values (Text S1); evaluation of model performance (Texts S2); data treatments of PMF and PMFr (Texts S3); correlation coefficients among species in the $PM_{2.5}$ speciation dataset (Figure S1); results of PMF runs for source-receptor relationship (Figures S2–S10); comparison of imputed values and actual values by different methods (Figures S11–S29); mechanism of generating missing values (Table S1); summary for the missing pattern of the subset of NEPB dataset (Table S2); results of PMF runs for source-receptor relationship (Tables S3 and S4); performance of imputation methods under Cases 1-9 (Tables S5–S12).

Part I. Supplementary Texts

Text S1. The Generation Mechanism of Gaps in Data

Short gaps are generated from an exponential distribution with λ as a parameter, which is determined by the proportion of the short gap[1]. The lengths of short gaps and median gaps are defined by their physical meaning and the frequency of gaps with a length of 6 (Table S2), similar to the lengths of median and large gaps. The generation of large gaps follows a uniform distribution between 7 and 23, and between 23 and 161, respectively. The generation of large gaps follows a uniform distribution between 7 and 23, and between 23 and 161. The parameter 161 is chosen to confine the missing data to gaps lasting more than 5 days, which is the longest gap in the dataset.

Text S2. Evaluation of Model Performance

To evaluate the performance of imputation methods, each imputed value was individually compared against the actual observed value separately. Three indicators were employed to assess the performance of these methods[**BENNETT20131**, 2]: the coefficient of determination (R^2), and mean absolute percentage error, and index of agreement (IoA). R^2 reflects the consistency of the trend between the predicted and observed values. IoA is similar to R but is specifically designed to measure differences in the means and variances between imputed and observed values. Lower MAPE indicate that the imputed values are closer to the actual observations, while higher R^2 and IoAd suggest greater consistency between the imputed values and the observations. The expressions of the indicators are shown as follows:

$$R^2 = \left[\frac{\sum_{i=1}^{n}(y_i - \overline{y})(\hat{y}_i - \tilde{y})}{\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \tilde{y})^2}} \right]^2 \tag{1}$$

$$IoA = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2} \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{3}$$

Where, y_i and $\hat{y}_i$ are the ith observation for the imputed and the original datasets, while $\bar{y}$ and $\tilde{y}$ are the means for the imputed and the original datasets.

Text S3. Details of Data Treatment and Uncertainties Estimation in PMF Analysis for Source Apportionment

For the purpose of selection for the reasonable factor profiles for PMF reconstruction, the species required for imputation have to be included in the PMF run. The simulated dataset includes five bulk species: NH_4^+, SO_4^{2-}, NO_3^-, OC, and EC and fifteen elements: K, Fe, Zn, Ca, Si, Mn, Pb, Cu, Ti, As, V, Ba, Cr, Se, and Ni. These species are all put into PMF run for SA. The dataset was pretreat by excluding samples where one of NH_4^+, SO_4^{2-}, NO_3^- is missing. Then the simulation is conducted on the processed dataset. Both the original and artificially generated missing values are replaced by geometric mean and the corresponding uncertainty

are set as four times as the mean. For observed values, the imputed uncertainty data was calculated as[3]:

$$\text{Uncertainty}=\sqrt{(\text{error fraction} \times \text{concentration})^2 + (0.5 \times \text{detection limit})^2} \quad (1)$$

where the error fraction was estimated as 10% for all chemical species[4–6]. Missing and BDL values of individual species, and their accompanying uncertainties were routinely replaced in a same manner as Polissar et al[7].

In this study, the final factor number of PMF solution was determined based on the interpretability of factor profiles and the model performance.

Part II. Supplementary Figures

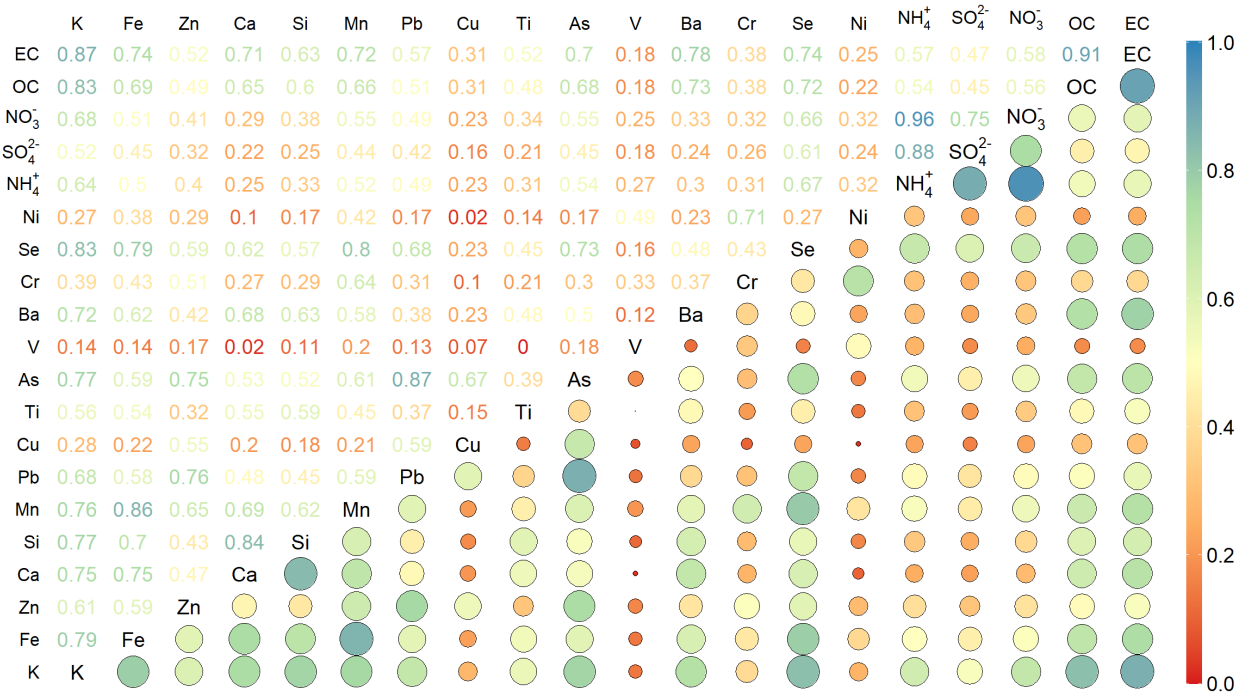

Figure S1: The correlation coefficients among species in the $PM_{2.5}$ speciation dataset.

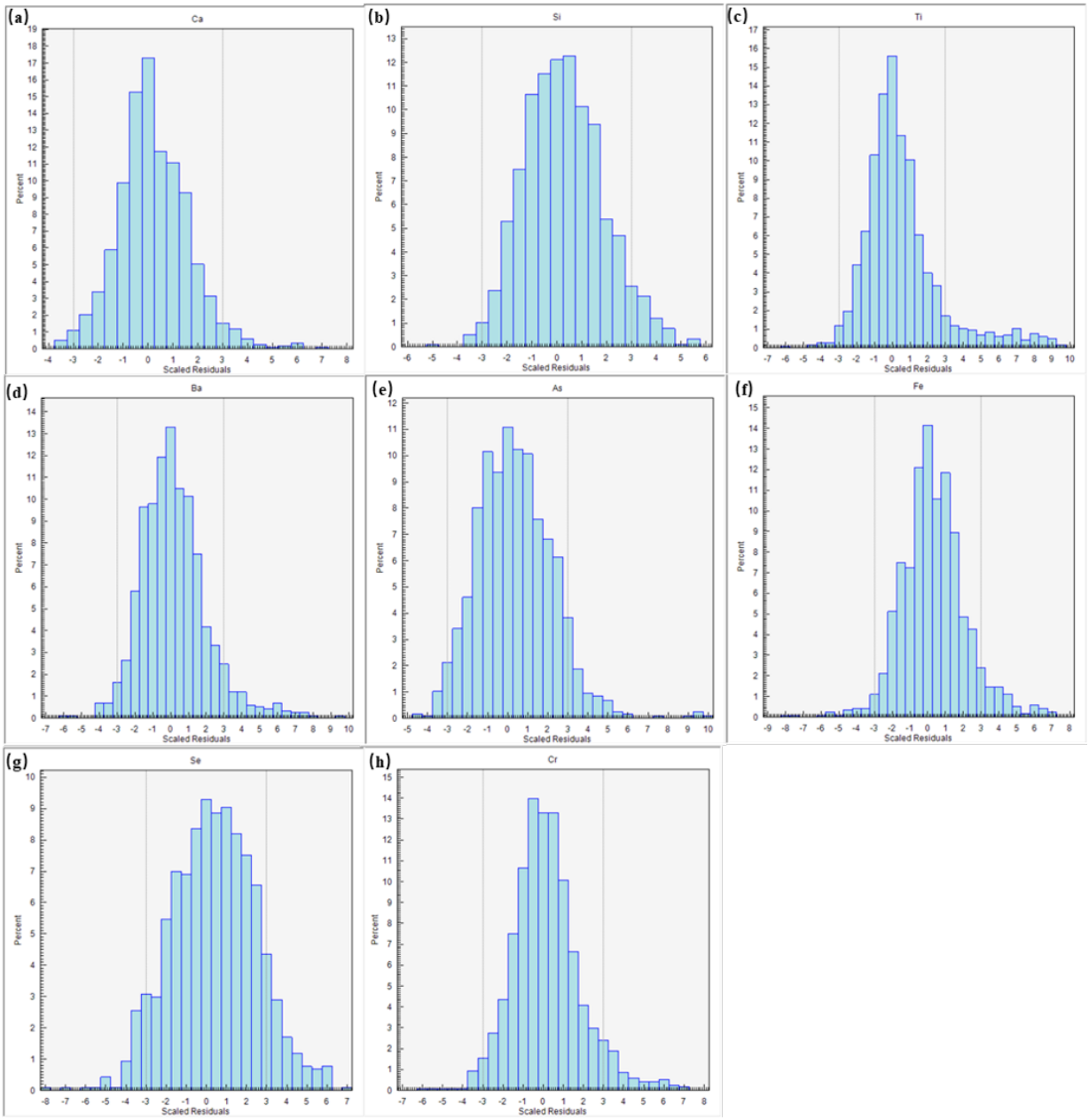

Figure S2: The histogram of scaled residuals of **(a)** Ca; **(b)** Si; **(c)** Ti; **(d)** Ba; **(e)** As; **(f)** Fe; **(g)** Se; and **(h)** Cr of the 7-factor solution.

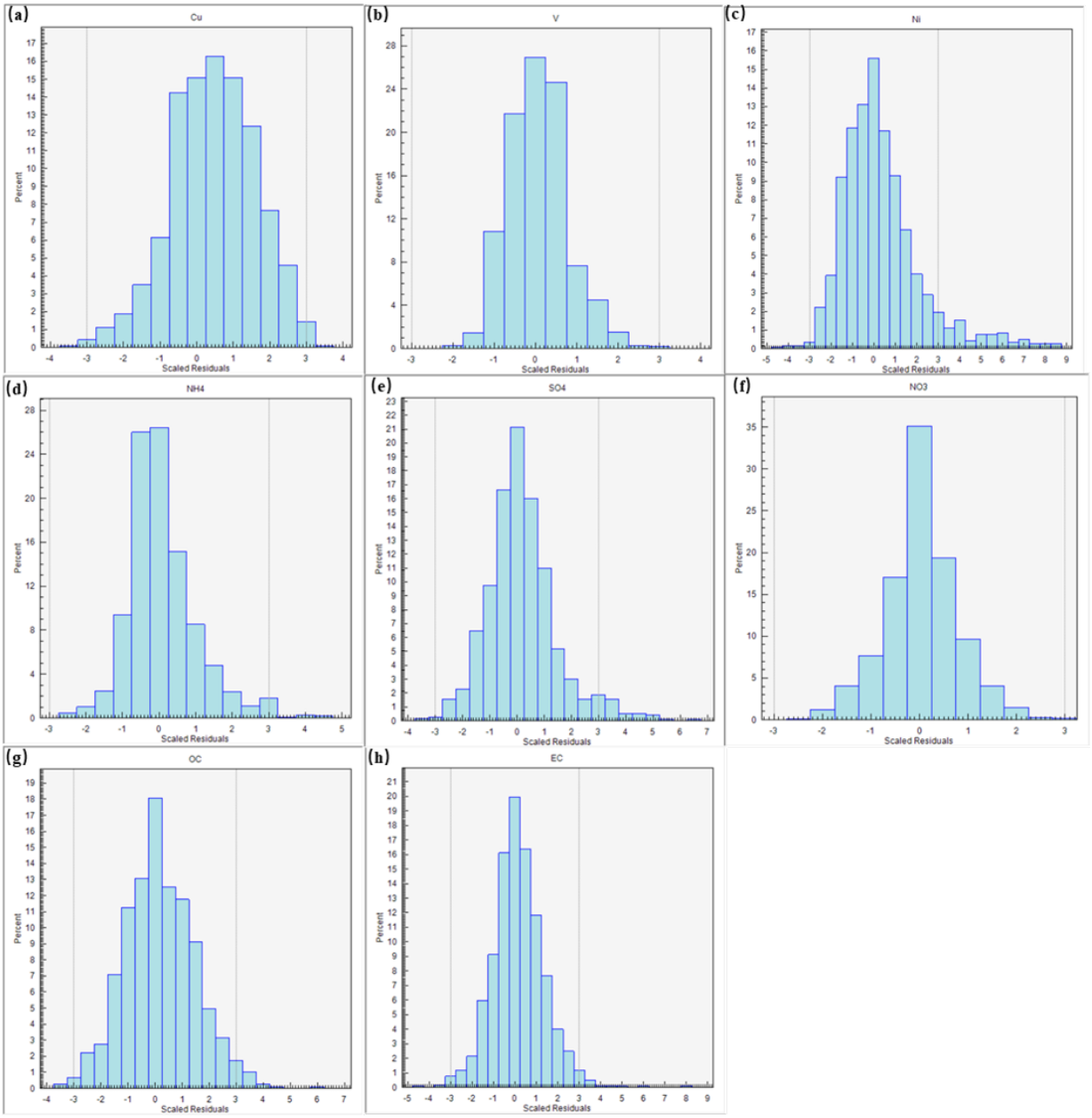

Figure S3: The histogram of scaled residuals of **(a)** Cu; **(b)** V; **(c)** Ni; **(d)** NH_4^+; **(e)** SO_4^{2-}; **(f)** NO_3^-; **(g)** OC; and **(h)** EC of the 7-factor solution.

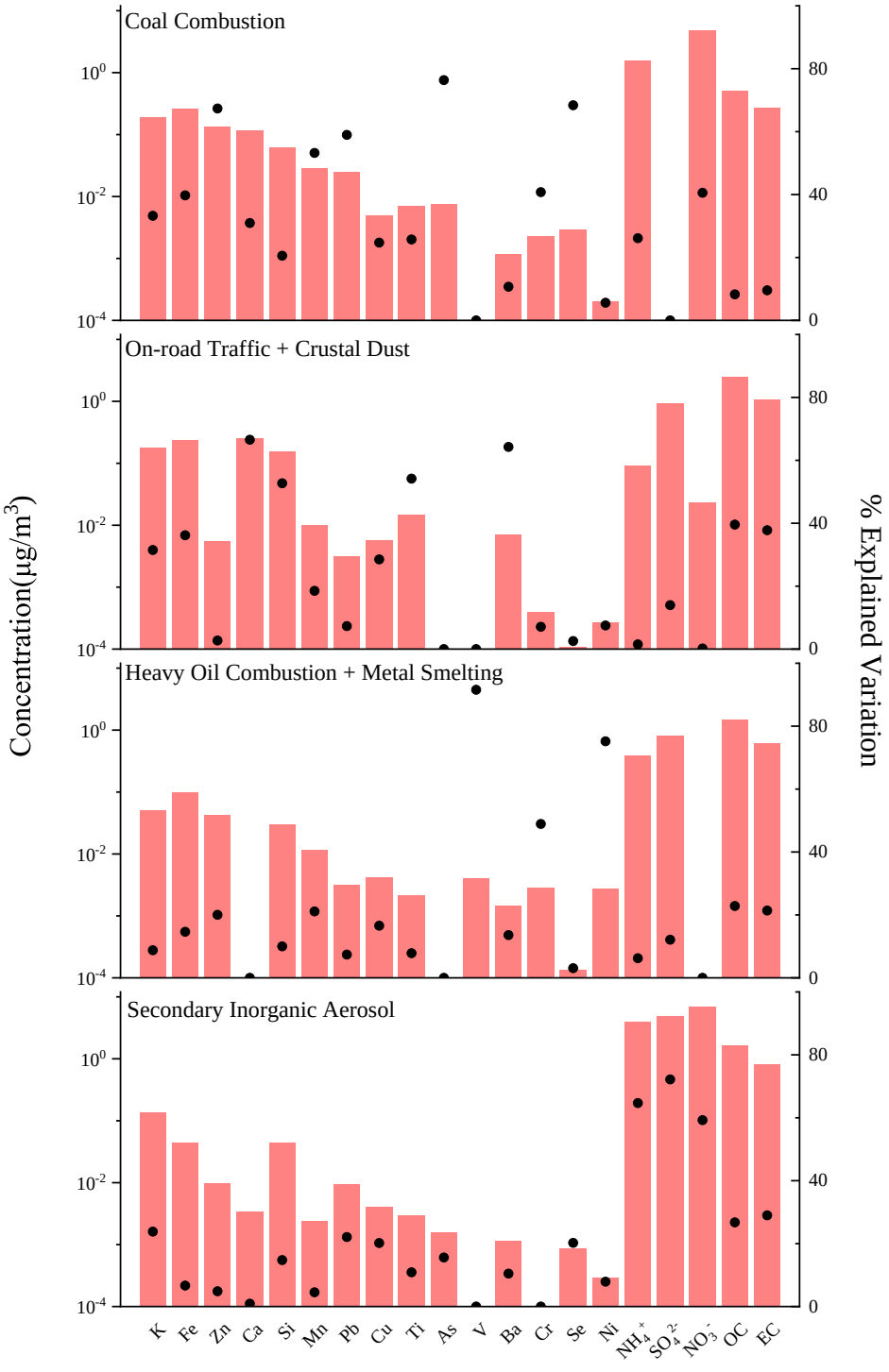

Figure S4: The factor profile of 4-factor solution

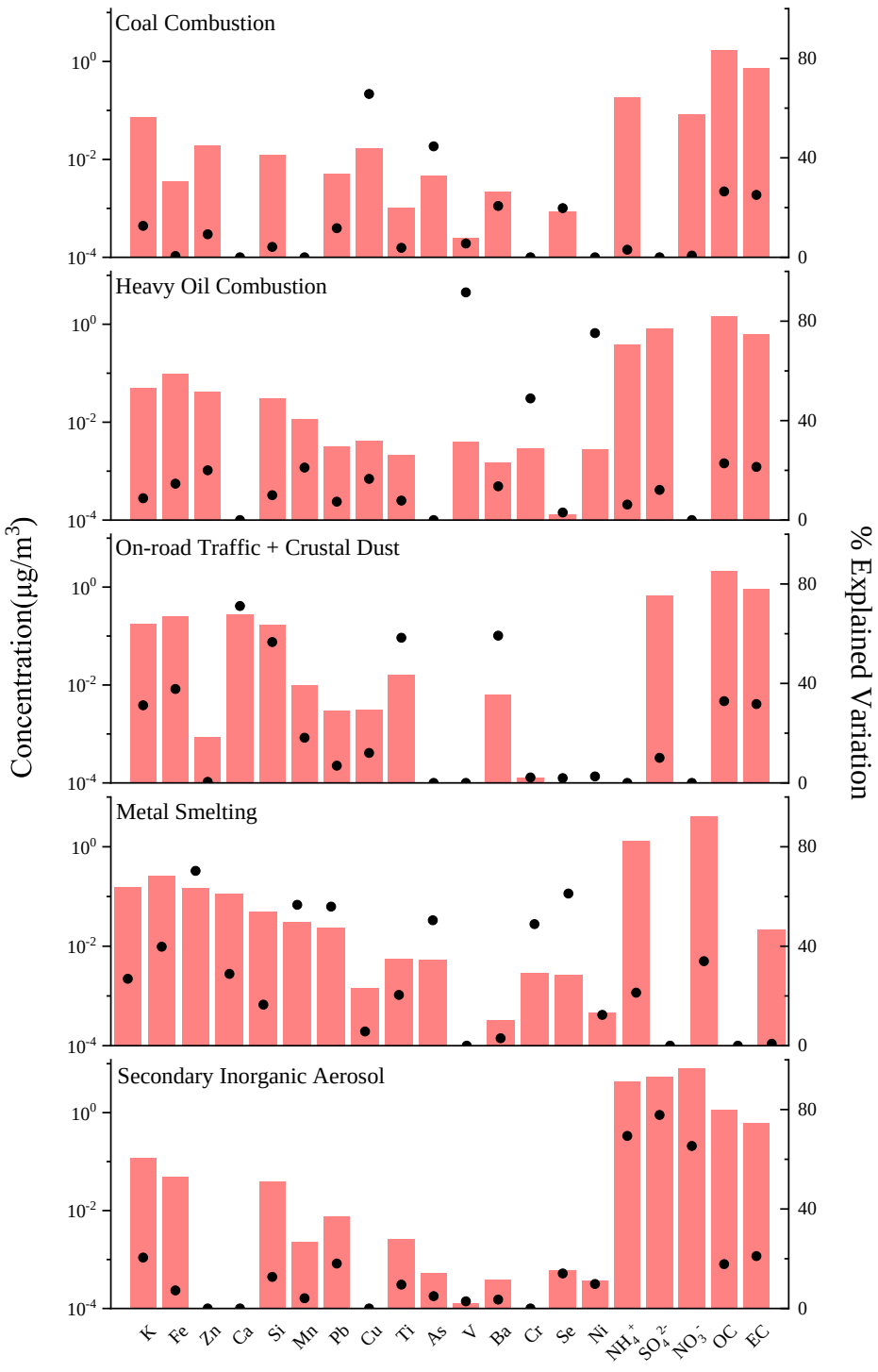

Figure S5: The factor profile of 5-factor solution.

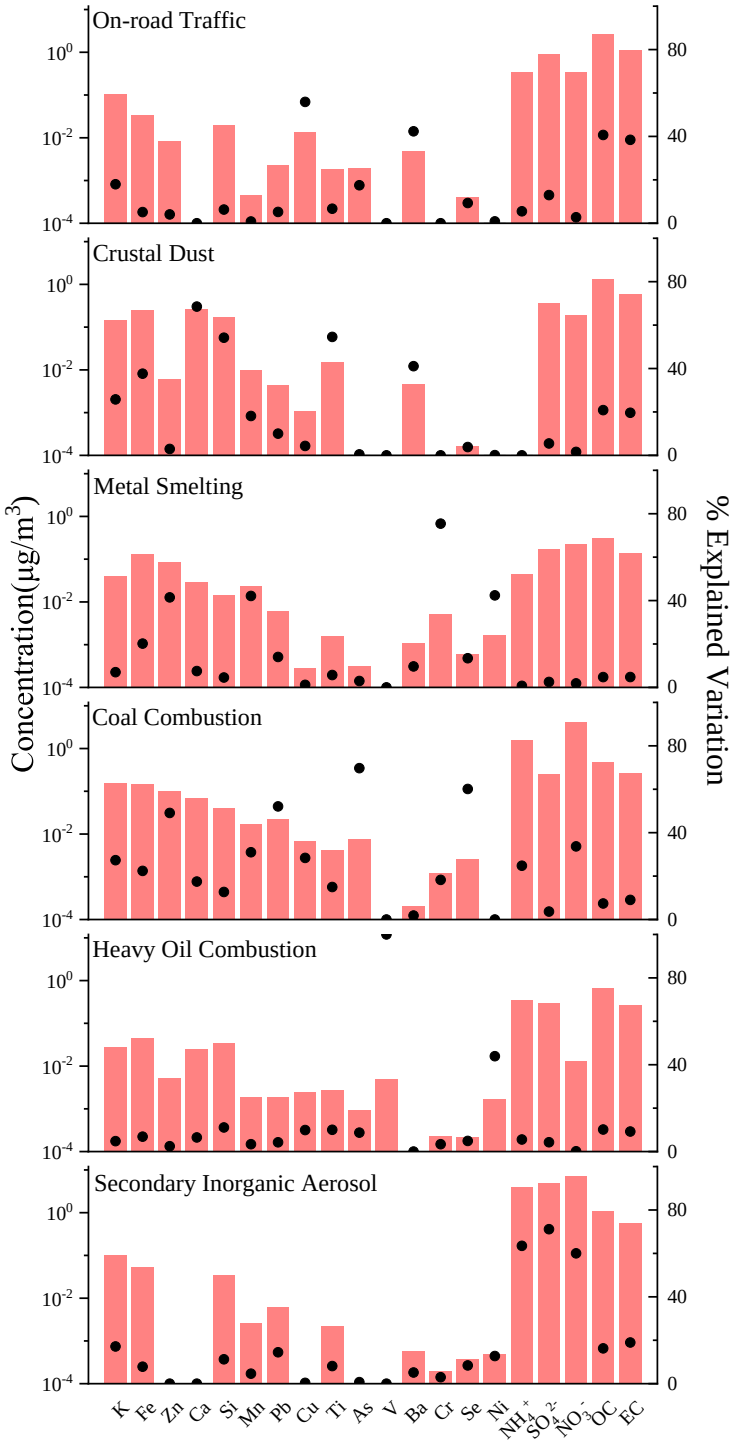

Figure S6: The factor profile of 6-factor solution.

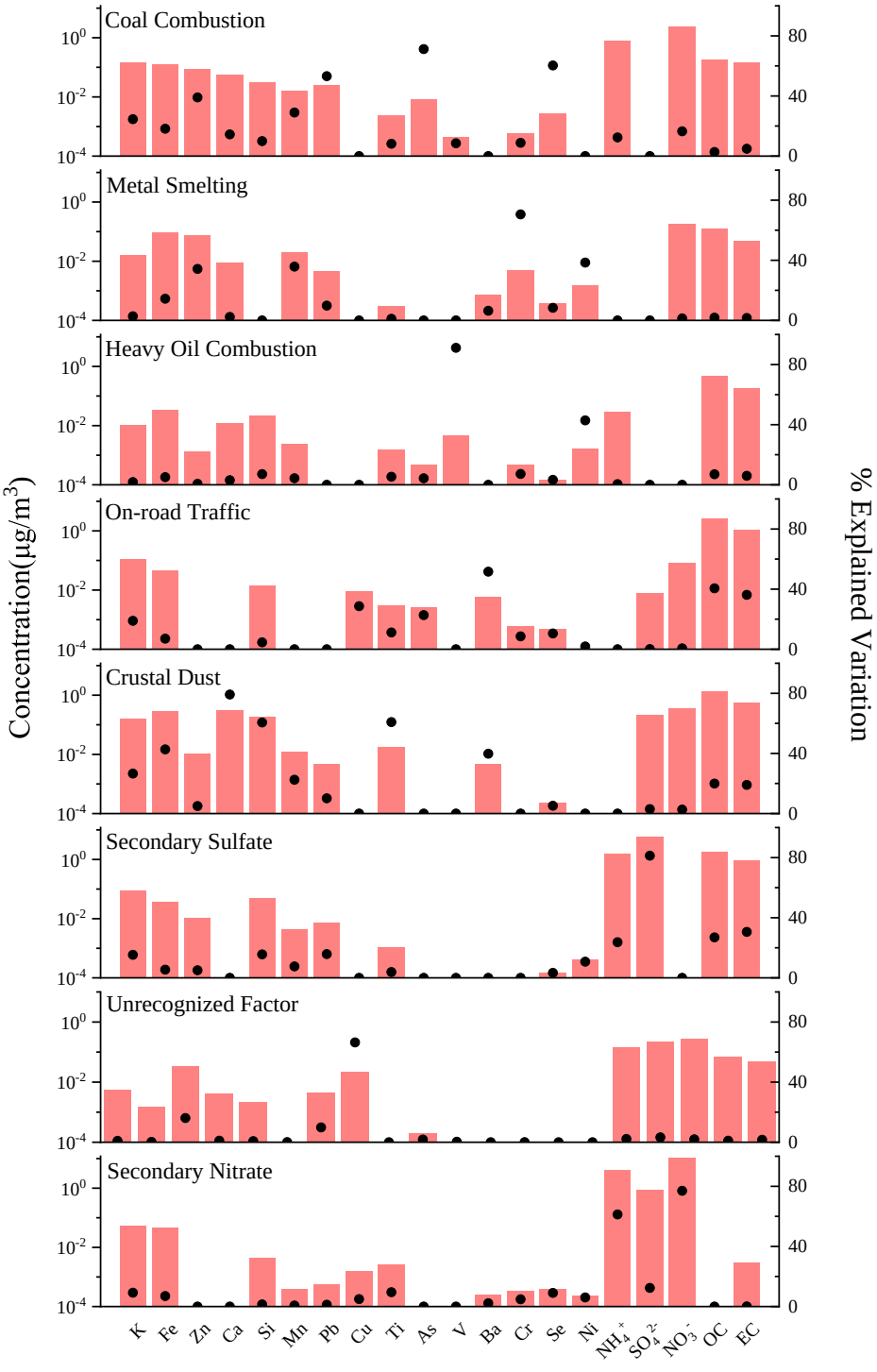

Figure S7: The factor profile of 8-factor solution.

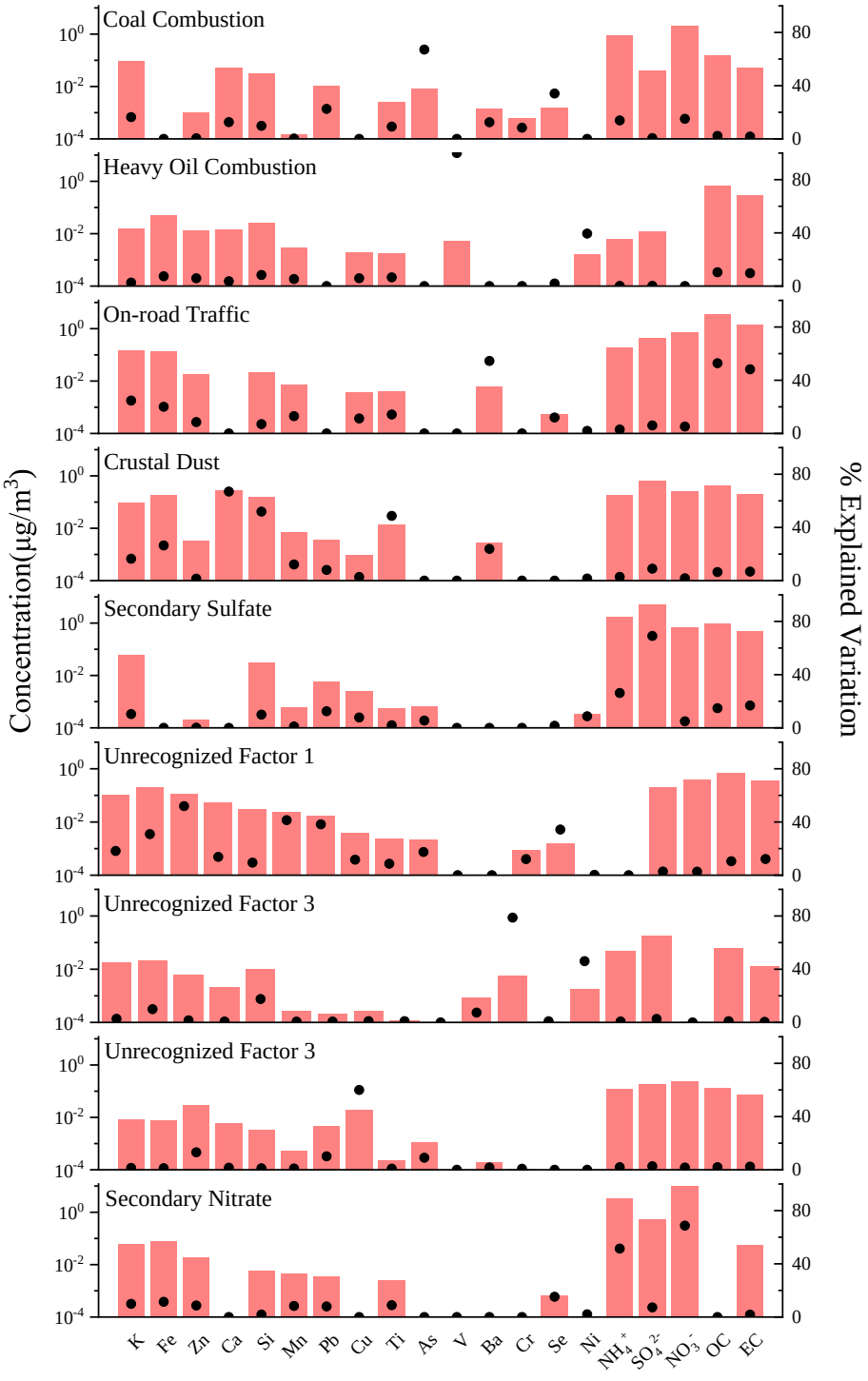

Figure S8: The factor profile of 9-factor solution.

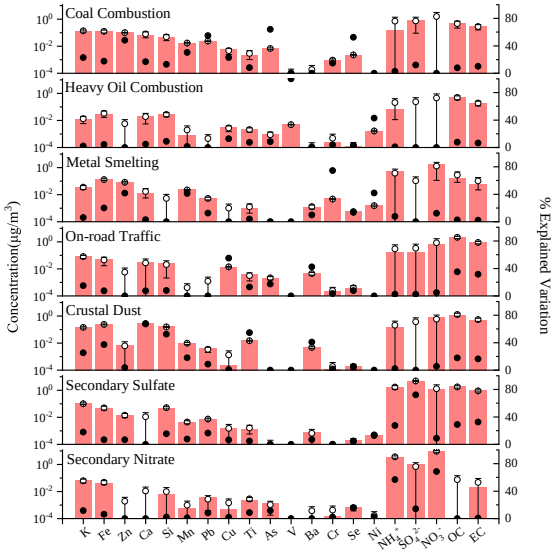

Figure S9: Factor profiles of 7-factor solutions. The bars are the estimated concentrations, the solid points are the explained variations, the asymmetric error bars represent the maximum and minimum DISP values.

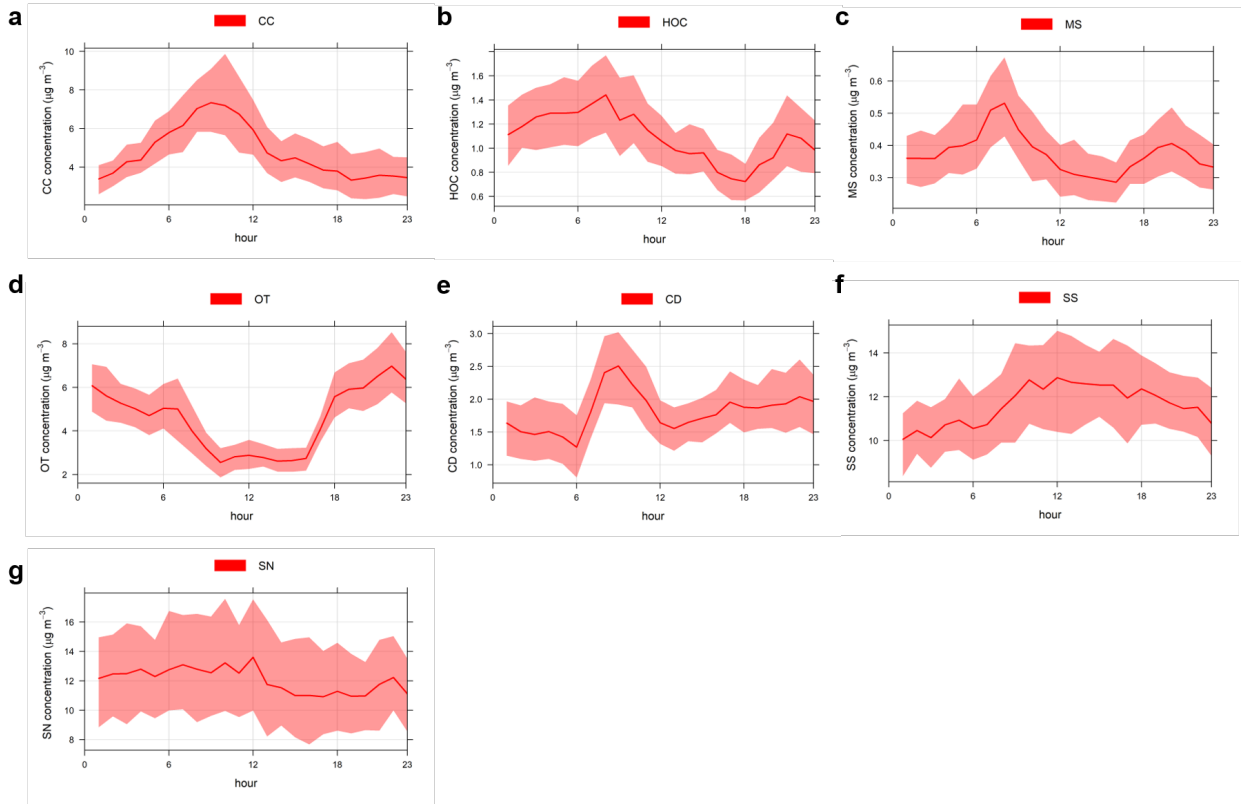

Figure S10: Diurnal variations of the seven factors resolved by PMF: **a** Coal Combustion; **b** Heavy Oil Combustion; **c** Metal Smelting; **d** On-road Traffic; **e** Crustal Dust; **f** Secondary Sulfate; **g** Secondary Nitrate. Shaded areas indicate the 95% confidence intervals.

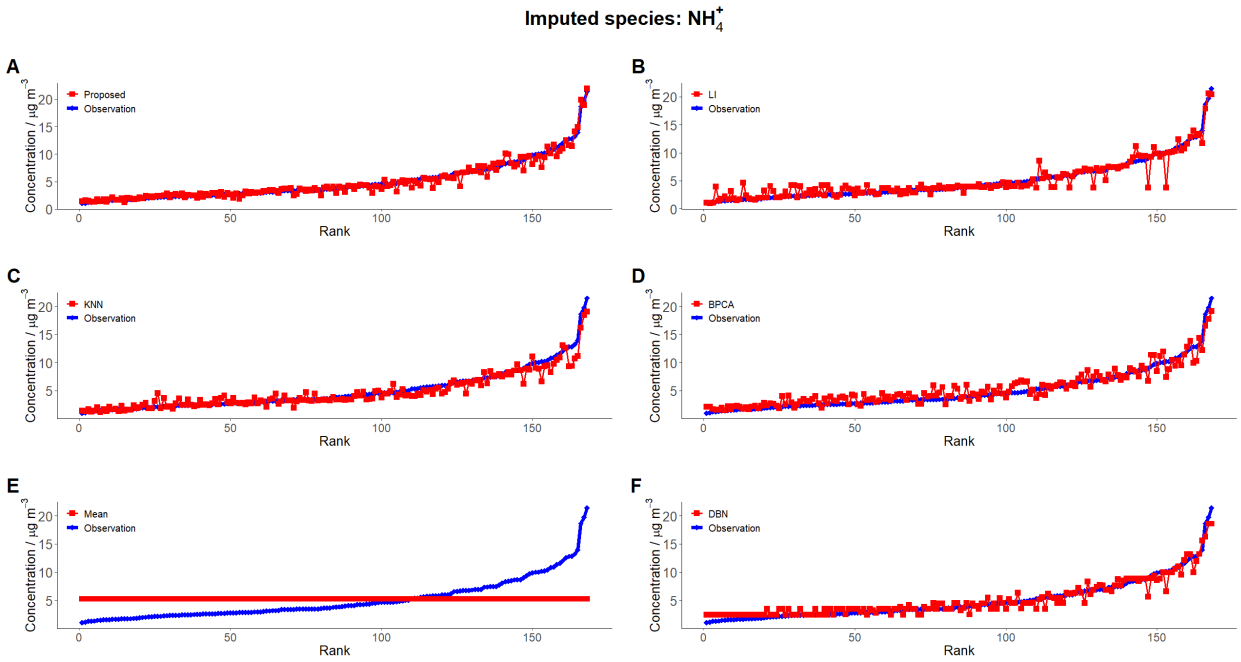

Figure S11: The comparison of imputed values and actual values for NH_4 under Case 1. Actual and imputed values are ranked according to the actual values. **A** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

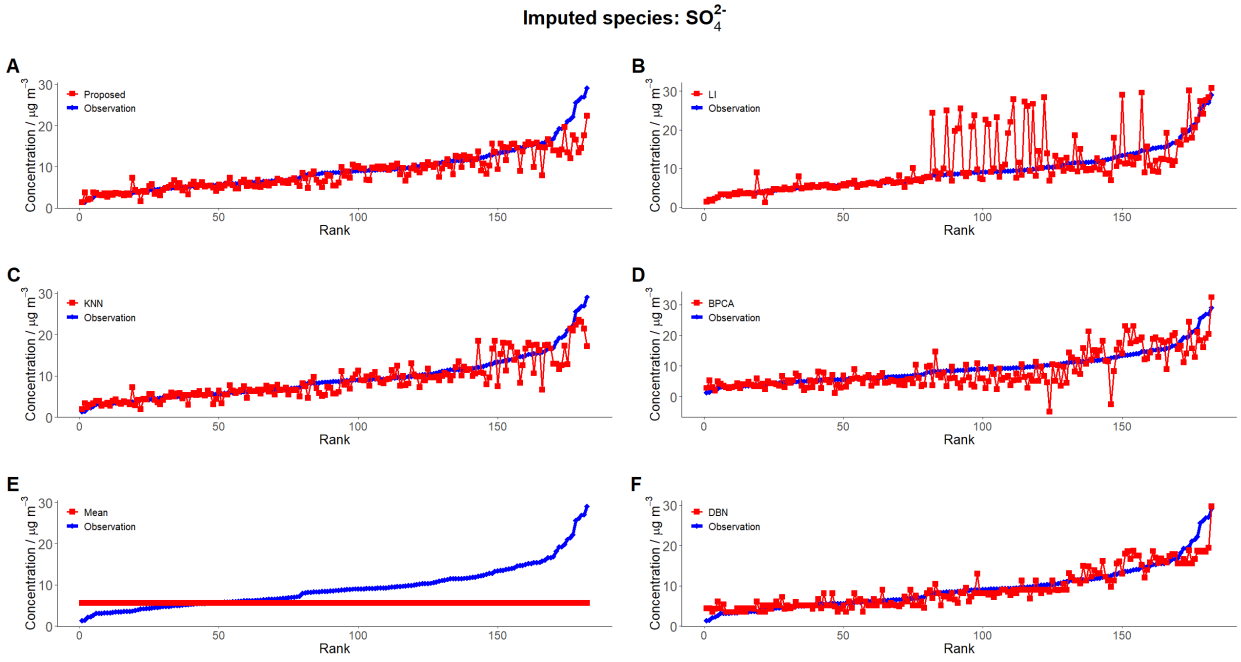

Figure S12: The comparison of imputed values and actual values for SO_4^{2-} under Case 1. Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

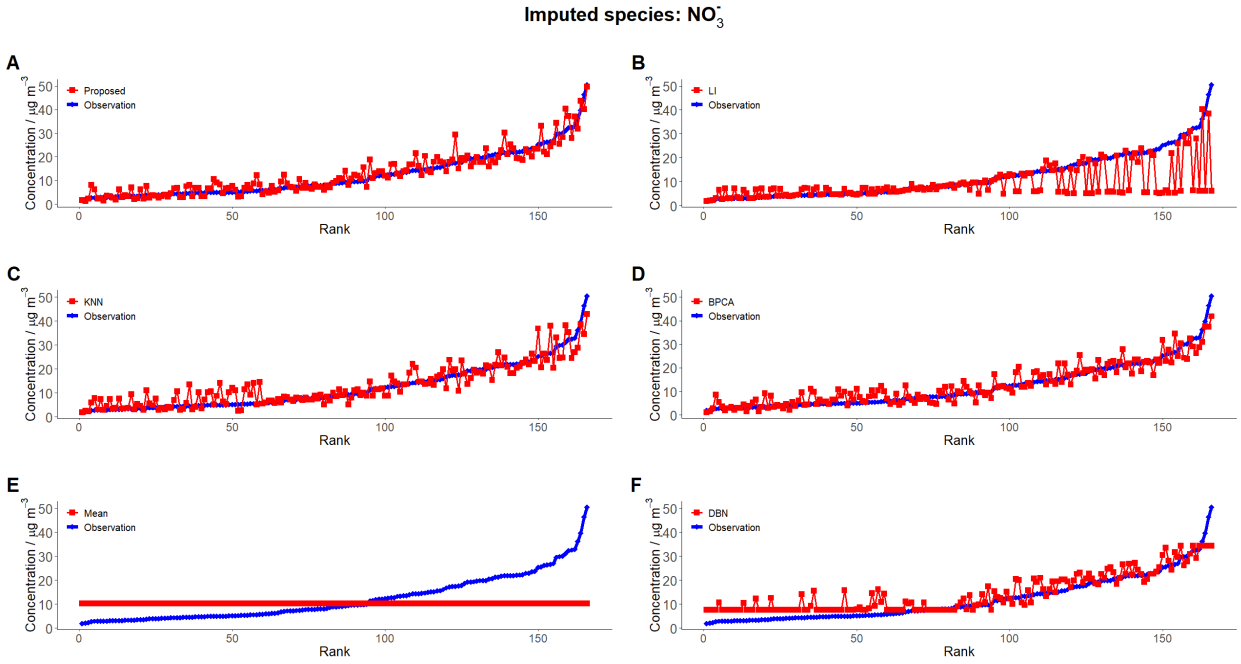

Figure S13: The comparison of imputed values and actual values for NO_3^- under Case 1. Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

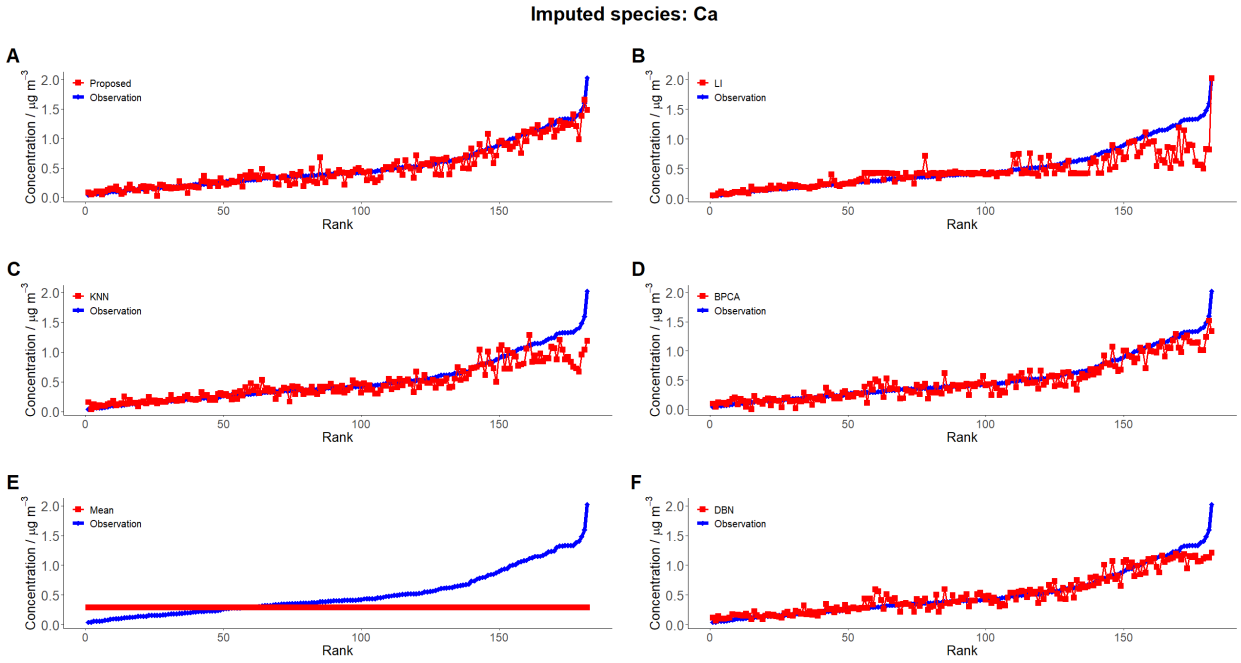

Figure S14: The comparison of imputed values and actual values for Ca under Case 1. Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

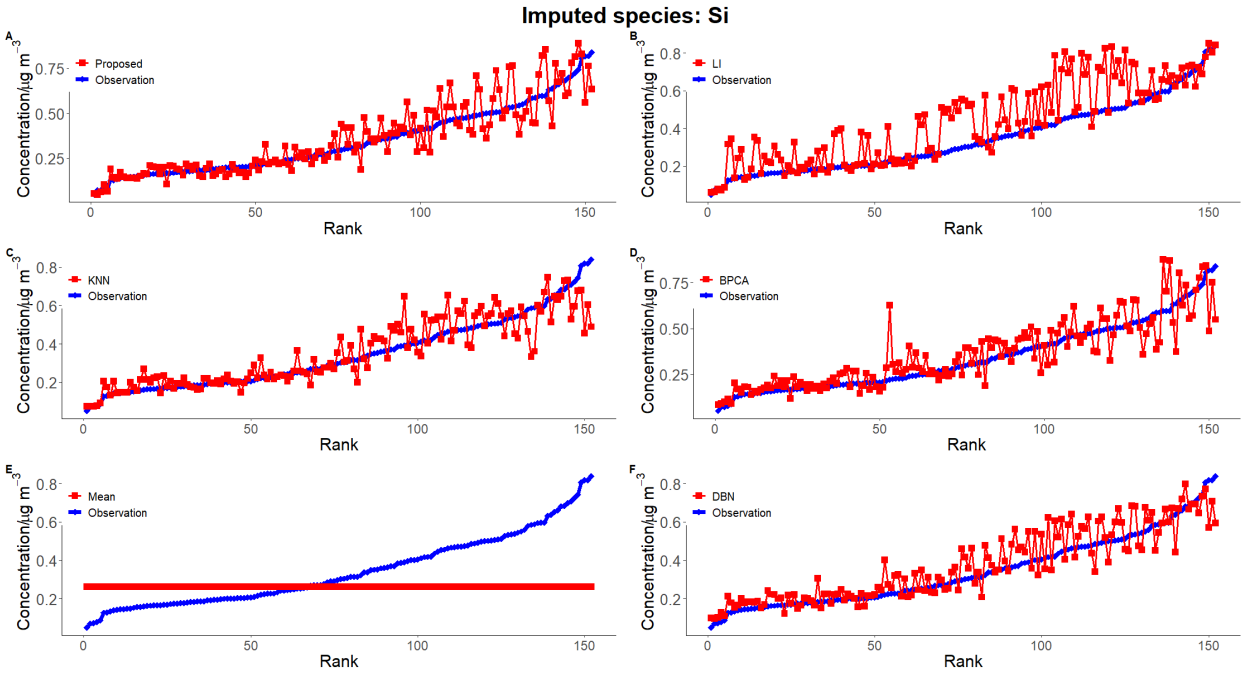

Figure S15: The comparison of imputed values and actual values for Si under Case 1. Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

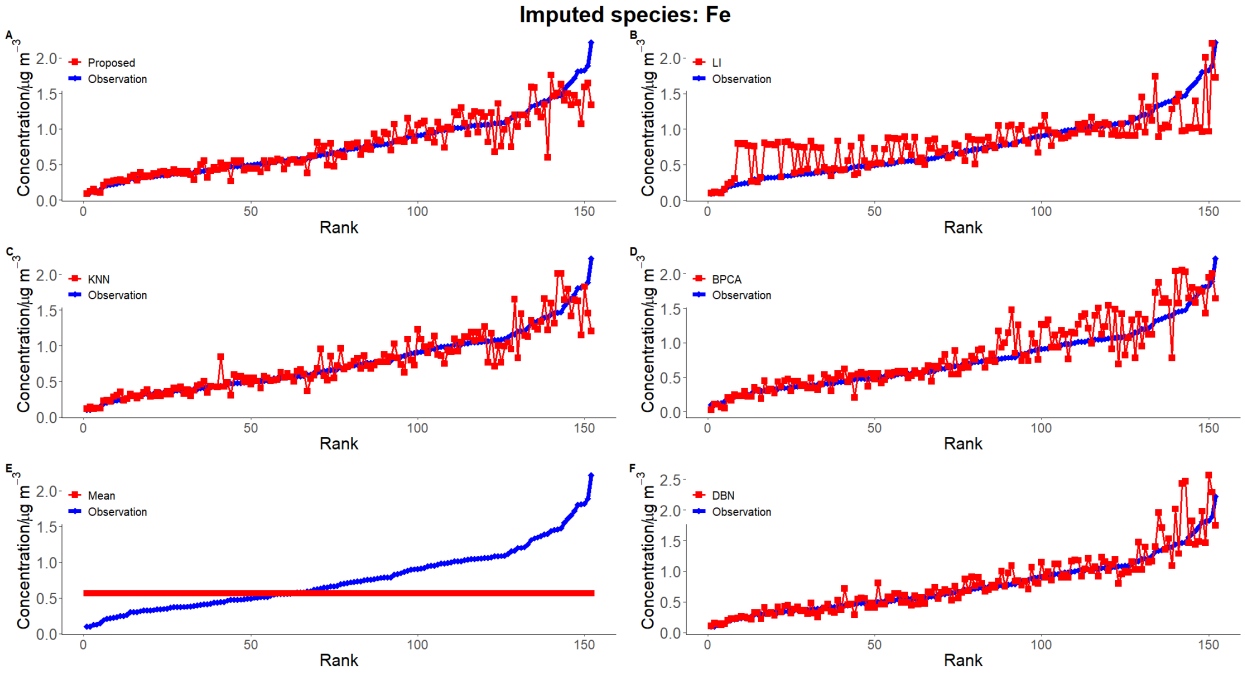

Figure S16: The comparison of imputed values and actual values for NO_3^- under Case 1. Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

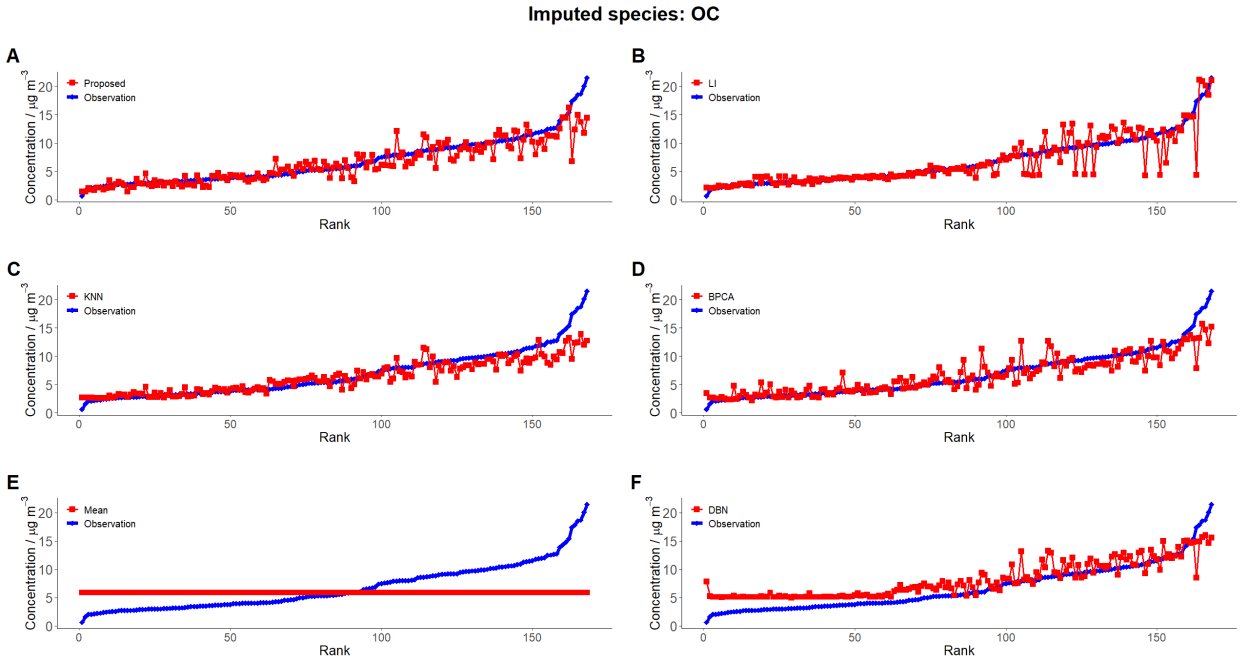

Figure S17: The comparison of imputed values and actual values for OC under Case 1. Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

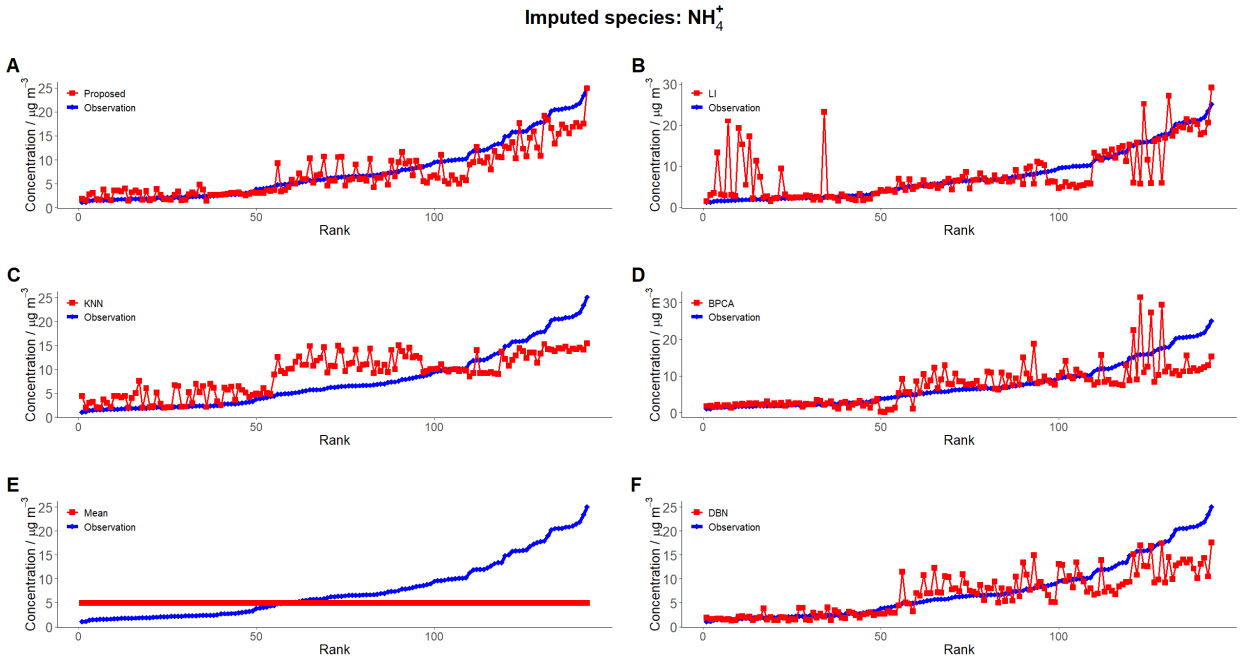

Figure S18: The comparison of imputed values and actual values for NH_4^+ under Case 2(missing proportion: 10%). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

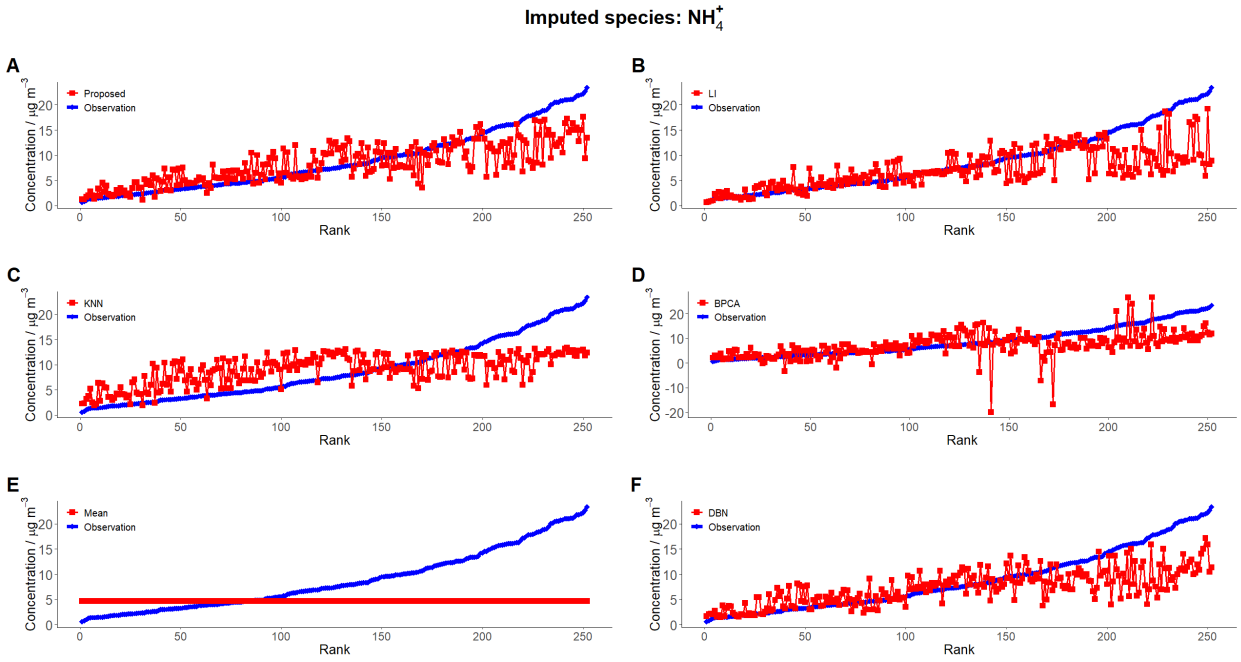

Figure S19: Comparison between the imputed and actual values of NH_4^+ under Case 2 (missing proportion: 20%). Actual and imputed values are ranked according to the actual values. **a** this study; **b** Mean; **c** LI; **d**) KNN; **e** BPCA; **f** DBN.

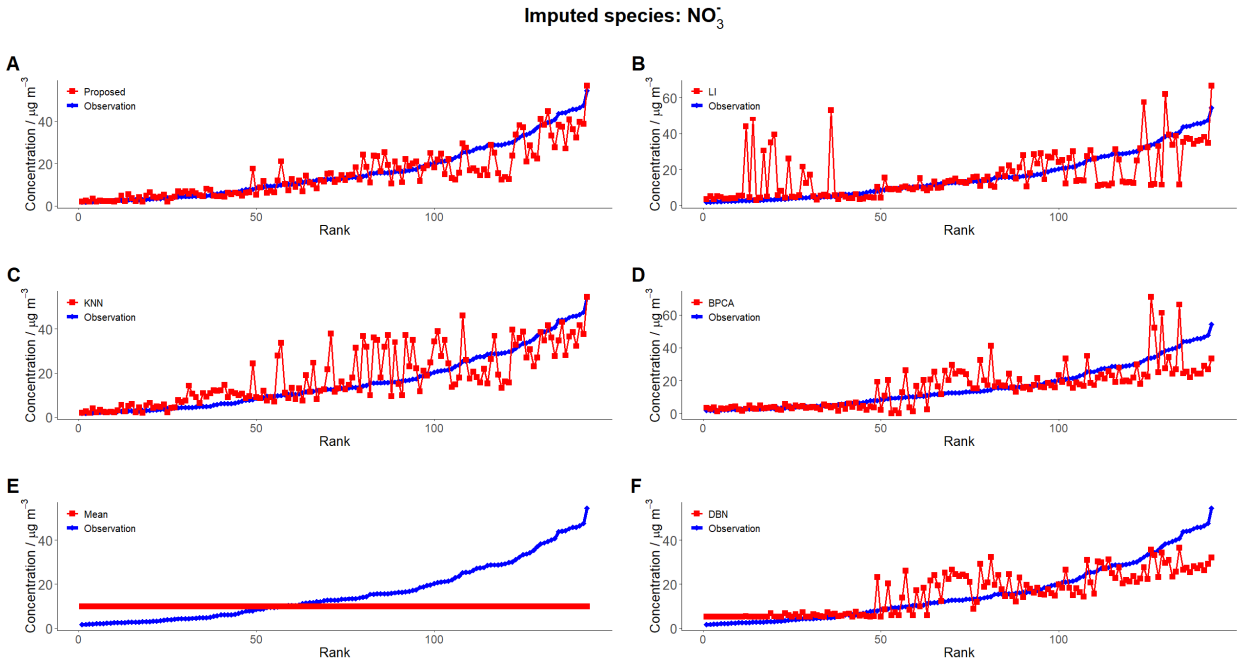

Figure S20: The comparison of imputed values and actual values for NO_3^- under Case 2(missing proportion: 10%). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

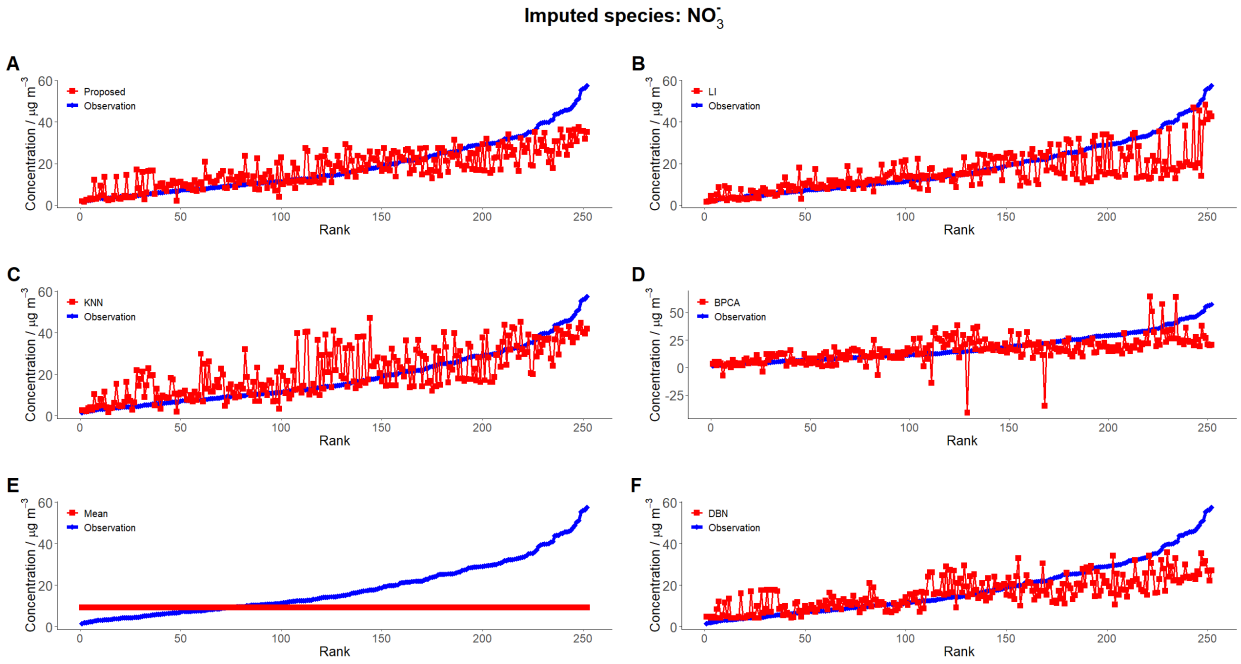

Figure S21: The comparison of imputed values and actual values for NO_3^- under Case 2(missing proportion: 20%). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

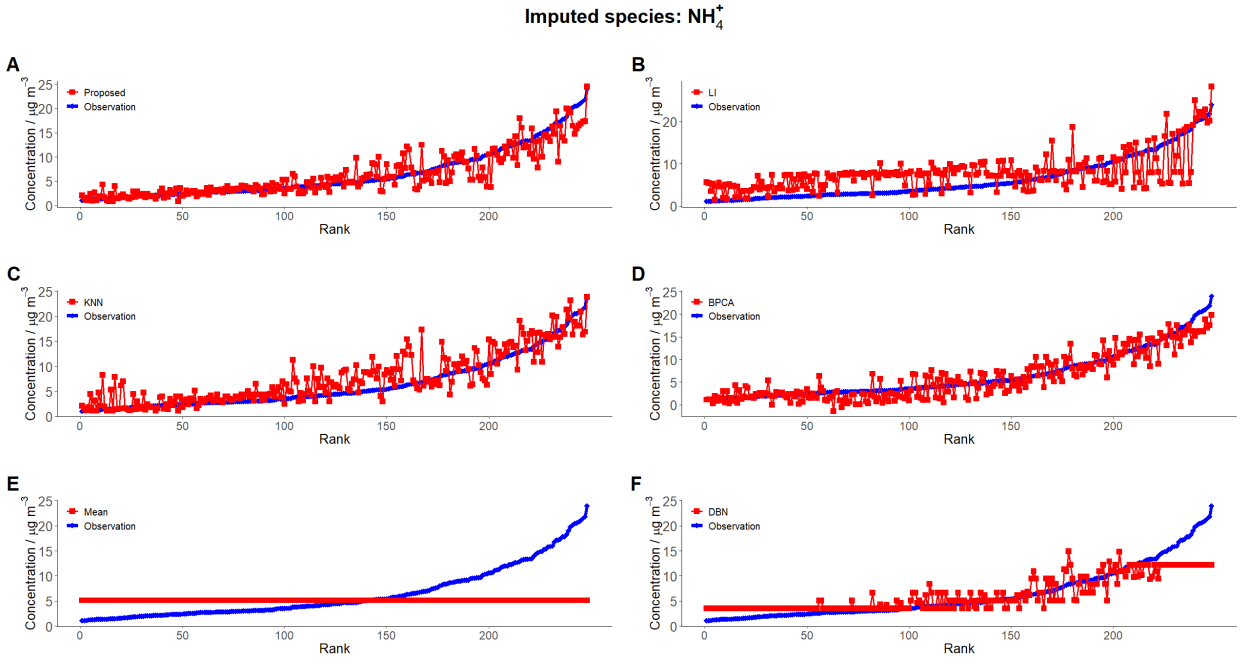

Figure S22: The comparison of imputed values and actual values for NH_4^+ under Case 4(missing proportion: 20%, MCMS). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

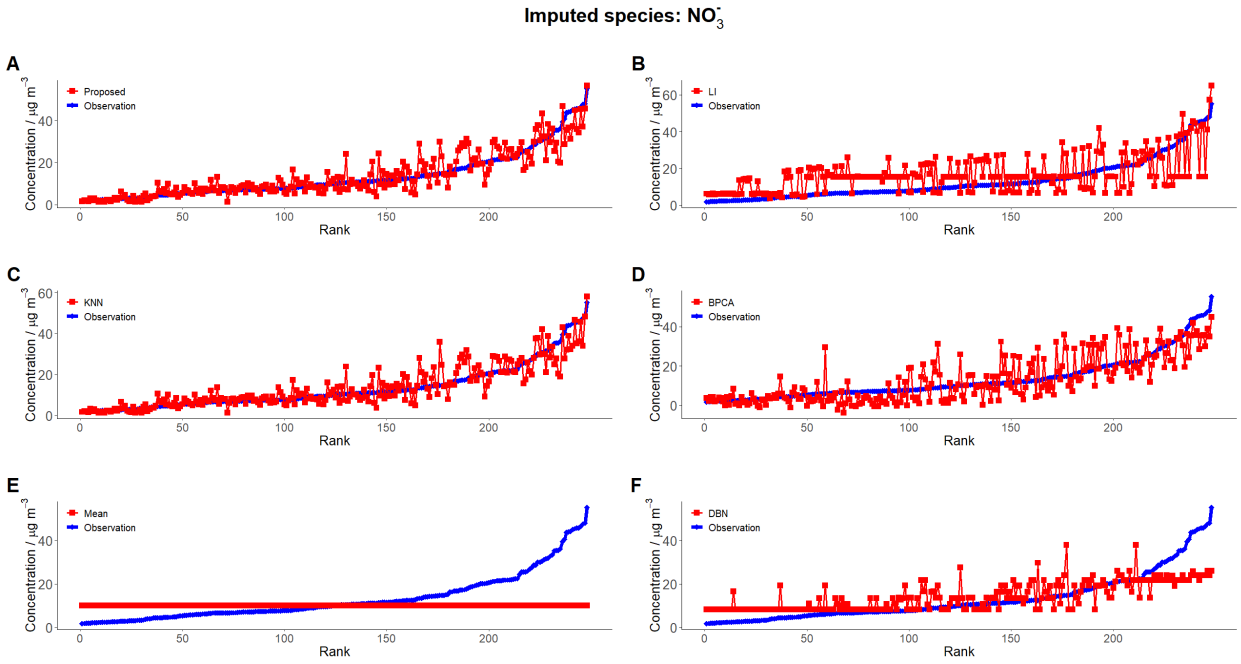

Figure S23: The comparison of imputed values and actual values for NO_3^- under Case 4(missing proportion: 20%, MCMS). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

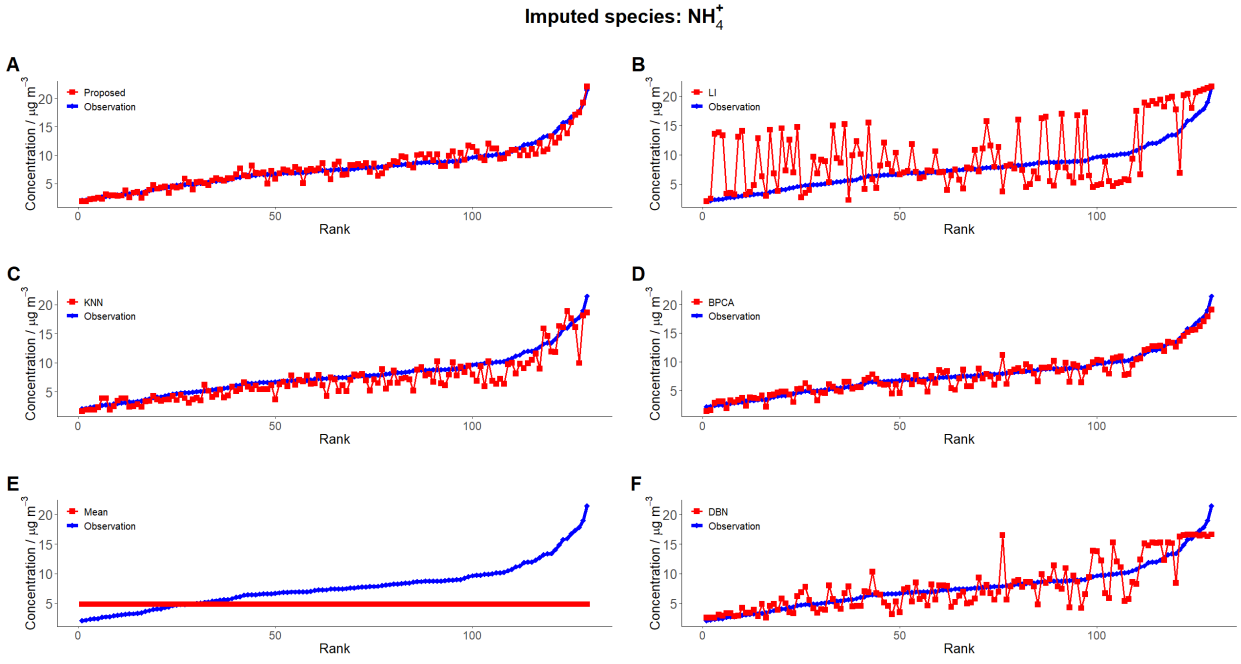

Figure S24: The comparison of imputed values and actual values for NH_4^+ under Case 4(missing proportion: 10%, MCMI). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

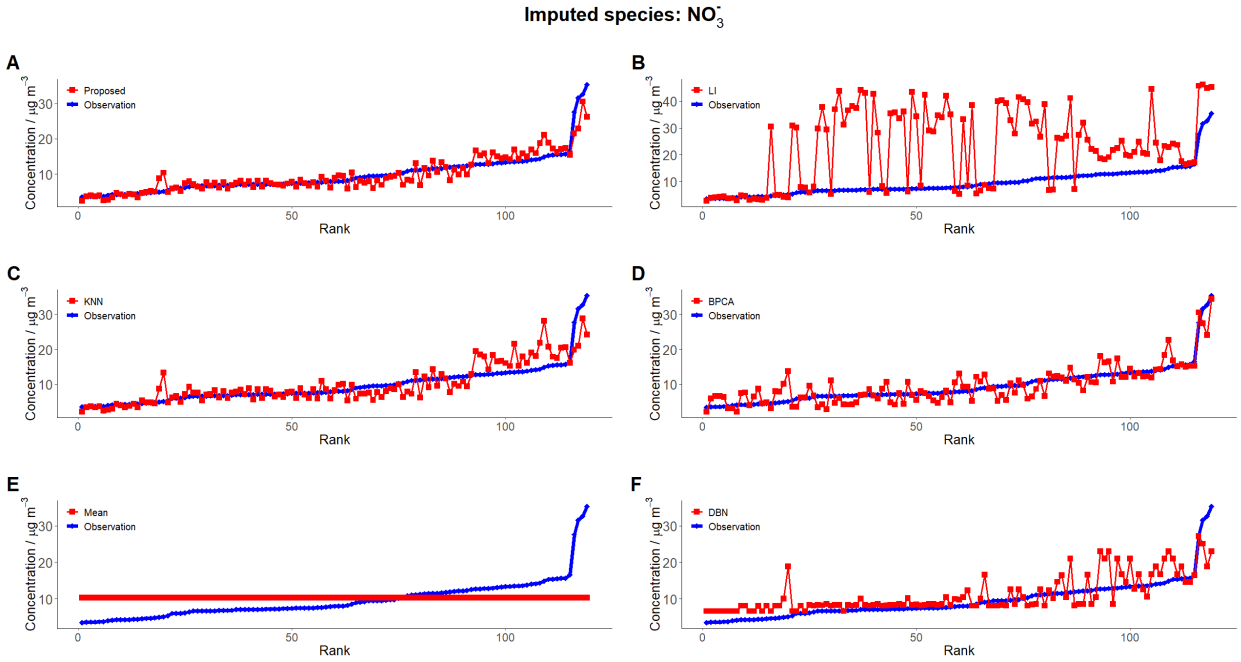

Figure S25: The comparison of imputed values and actual values for NO_3^- in Case 4(missing proportion: 10%, MCMI). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

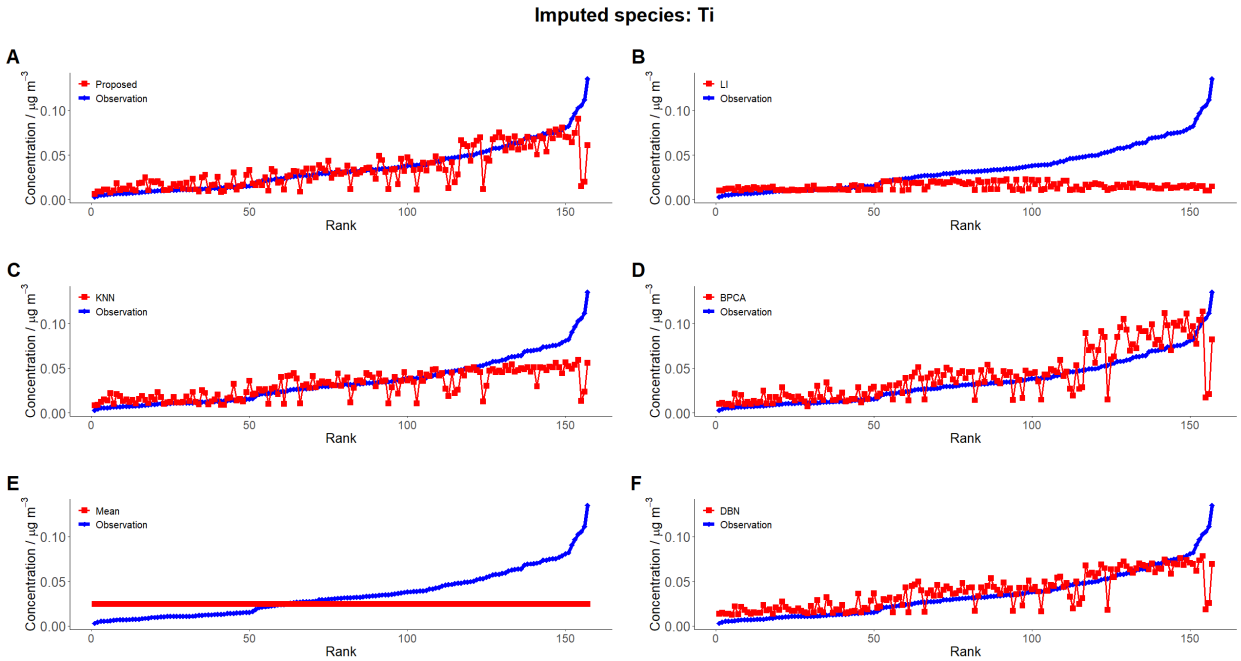

Figure S26: The comparison of imputed values and actual values for Ti in Case 5(missing proportion: 10%, MCMS). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

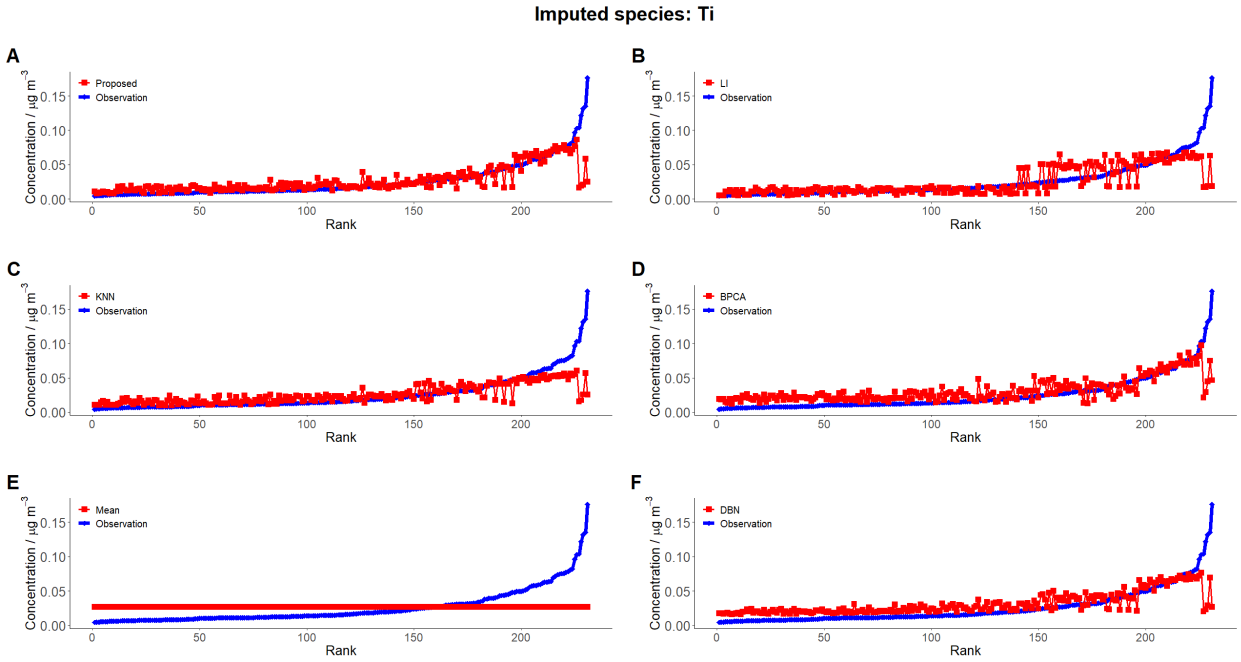

Figure S27: The comparison of imputed values and actual values for Ti in Case 5(missing proportion: 20%, MCMS). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

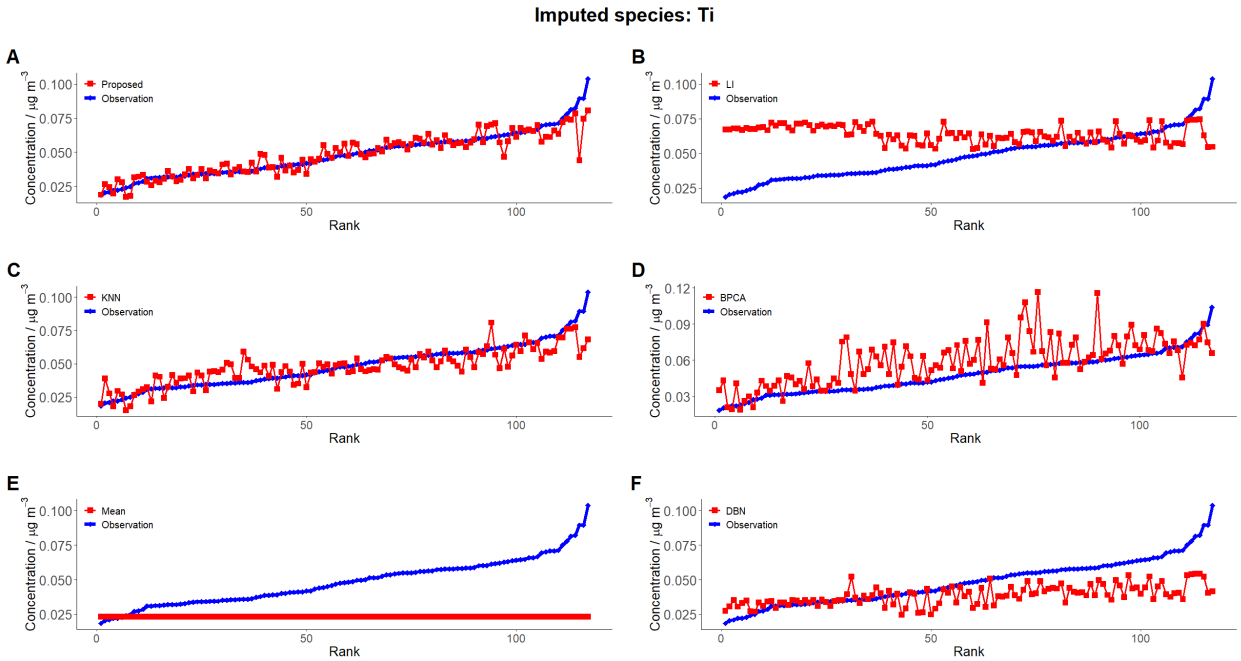

Figure S28: The comparison of imputed values and actual values for Ti in Case 5(missing proportion: 10%, MCMI). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

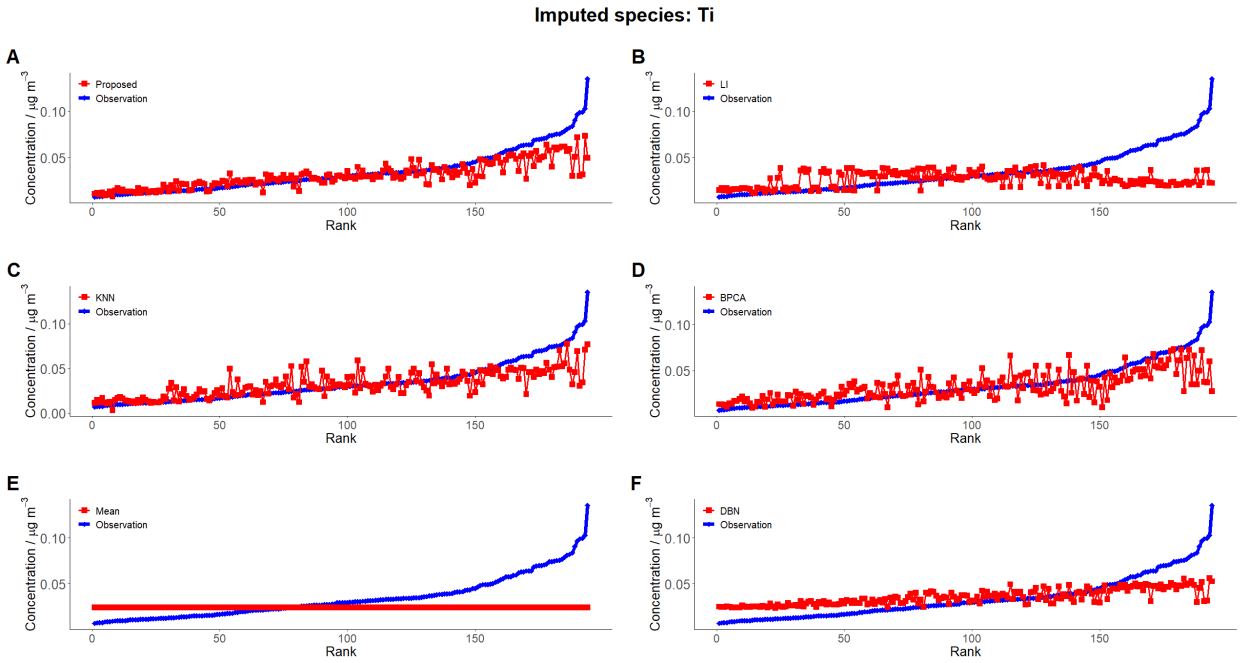

Figure S29: The comparison of imputed values and actual values for Ti in Case 5(missing proportion: 20%, MCMI). Actual and imputed values are ranked according to the actual values. **a** this study; **B** LI; **C** KNN; **D** BPCA; **E** Mean; **F** DBN

Part III. Supplementary Tables

Table S1: The generation mechanism of gap patterns

$l = 1$	$1 < l \leq 6$	$7 < l \leq 23$	$l \geq 24$
short gap		median gap	large gap
$X \sim \text{Exp}(\lambda)$		$X \sim \mathcal{U}(7, 23)$	$X \sim \mathcal{U}(23, 115)$

Table S2: Summary for the missing pattern of the subset of NEPB dataset. n_{total} refers to the total number of missing data in column, namely per species, and % of n_{total} refers to the proportion of the categorized gap accounts for in the species.

gap length	**species**					
	Ca		**Si**		**Ti**	
	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$
1-5h	6	20.0	6	20.0	6	20.0
6h	3		3		3	
7-23h	3	23.4	3	23.4	3	23.4
24h	0	56.6	0	56.6	0	56.6
>24h	1		1		1	
	As		$\mathbf{NH_4^+}$		$\mathbf{SO_4^{2-}}$	
	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$
1-5h	12	21.1	12	25.9	12	26.5
6h	3		3		4	
7-23h	4	26.9	3	22.3	3	22.1
24h	0	52.0	0	51.8	0	51.3
>24h	1		1		1	
	$\mathbf{NO_3^-}$		**OC**		**EC**	
	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$	$\mathbf{n}_{gaps}$	% of $\mathbf{n}_{total}$
1-5h	13	26.9	11	15.2	11	15.2
6h	4		2		2	
7-23h	3	22.0	3	13.4	3	13.4
24h	0	51.1	1	71.5	1	71.5
>24h	1		2		2	

Table S3: Summaries of BS, DISP and BS-DISP error estimation diagnostics and Q/Q_{exp} values of 4- to 9-factor PMF solutions

	4-factor	5-factor	6-factor	7-factor	8-factor	9-factor
BS diagnostics						
Lowest %BS mapping	97	43	85	93	93	83
Highest % unmapped	0	1	1	2	7	1
DISP diagnostics						
Error Code	0	0	0	0	0	0
Largest Decrease in Q	-0.035	0	-0.048	0	0	-0.117
%dQ	0	0	0	0	0	0
Highest swap by factor	0	0	0	0	0	0
BS-DISP diagnostics						
Number of cases accepted	100	92	95	87	99	87
% of cases accepted	100	92	95	87	99	87
Largest Decrease in Q	-12.657	-31.180	-43.780	-22.987	-1.754	-28.813
%dQ	-0.012	-0.03	-0.063	-0.040	-0.004	-0.073
Number of decreases in Q	0	1	3	7	0	2
Number of swaps in best fit	0	0	2	2	1	1
Number of swaps in DISP	0	2	0	4	0	10
Highest swaps by factor	0	4	2	3	1	1
$\mathbf{Q/Q}_{exp}$	6.64	5.98	5.18	4.60	4.21	3.76

Table S4: The squared correlation coefficients between the model-predicted species and actual observations

Element	Ca	V	Mn	As	Ni	Ba	Pb
	0.92	0.91	0.86	0.70	0.62	0.76	0.61
Bulk species	NH_4^+	SO_4^{2-}	NO_3^-	OC	EC		
	0.98	0.86	0.88	0.79	0.82		

Table S5: Performance metrics of the imputation methods under Case 1

Species	% of missing	Method	R^2	IoA	MAPE(%)
NH_4^+	15	Proposed	0.96	0.99	10.63
		Linear	0.91	0.98	18.17
		KNN	0.94	0.98	14.96
		DBN	0.94	0.98	20.12
		BPCA	0.92	0.97	21.63
NO_3^-	15	Proposed	0.91	0.97	26.64
		Linear	0.35	0.82	28.69
		KNN	0.84	0.95	33.62
		DBN	0.86	0.98	55.44
		BPCA	0.89	0.97	29.77
SO_4^{2-}	15	Proposed	0.79	0.92	15.09
		Linear	0.49	0.79	28.33
		KNN	0.79	0.94	17.04
		DBN	0.83	0.96	19.81
		BPCA	0.61	0.87	31.68
Ca	15	Proposed	0.93	0.98	13.97
		Linear	0.73	0.88	22.05
		KNN	0.84	0.94	18.33
		DBN	0.90	0.97	21.84
		BPCA	0.91	0.97	24.00
Si	15	Proposed	0.82	0.95	17.17
		Linear	0.74	0.87	33.42
		KNN	0.79	0.94	16.43
		DBN	0.82	0.94	21.30
		BPCA	0.76	0.93	20.90
Fe	15	Proposed	0.83	0.95	13.21
		Linear	0.62	0.86	34.23
		KNN	0.84	0.96	13.54
		DBN	0.86	0.96	14.79
		BPCA	0.84	0.94	19.35
OC	15	Proposed	0.81	0.94	17.42
		Linear	0.80	0.94	14.29

Table S5 (continued)

Species	% of missing	Method	R^2	IoA	MAPE(%)
		KNN	0.86	0.93	17.68
		DBN	0.83	0.90	45.14
		BPCA	0.81	0.93	19.43
EC	15	Proposed	0.82	0.95	15.53
		Linear	0.64	0.89	23.32
		KNN	0.87	0.93	13.89
		DBN	0.86	0.96	12.53
		BPCA	0.80	0.94	16.53

Table S6: Performance metrics of the imputation methods under Cases 2 and 3.

Species	% of Missing	Case	Method	R^2	IoA	MAPE(%)
NH_4^+	10	2	Proposed	0.85	0.95	30.66
			Linear	0.53	0.86	70.93
			KNN	0.56	0.62	66.02
			DBN	0.67	0.87	30.05
			BPCA	0.52	0.84	34.59
SO_4^{2-}			Proposed	0.74	0.83	40.98
			Linear	0.64	0.89	59.35
			KNN	0.59	0.85	61.63
			DBN	0.63	0.78	44.43
			BPCA	0.43	0.75	49.08
NO_3^-			Proposed	0.81	0.94	27.09
			Linear	0.37	0.79	95.53
			KNN	0.60	0.87	49.42
			DBN	0.66	0.88	48.75
			BPCA	0.53	0.85	41.45
NH_4^+	20		Proposed	0.61	0.84	38.05
			Linear	0.50	0.75	29.61
			KNN	0.40	0.69	67.67
			DBN	0.52	0.75	37.92
			BPCA	0.26	0.66	52.40
SO_4^{2-}			Proposed	0.36	0.59	45.01
			Linear	0.57	0.71	41.67
			KNN	0.26	0.68	69.34
			DBN	0.34	0.52	52.09
			BPCA	0.13	0.47	61.83
NO_3^-			Proposed	0.67	0.86	38.90
			Linear	0.51	0.80	33.70
			KNN	0.53	0.84	57.11
			DBN	0.53	0.77	44.71
			BPCA	0.31	0.72	52.86
OC	10	3	Proposed	0.73	0.92	19.25
			Linear	0.34	0.74	31.32
			KNN	0.68	0.89	25.89
			DBN	0.74	0.92	22.51
			BPCA	0.66	0.86	29.95
EC			Proposed	0.84	0.96	13.14
			Linear	0.24	0.71	46.02
			KNN	0.79	0.94	13.89
			DBN	0.80	0.91	18.66
			BPCA	0.85	0.93	19.03
OC	20		Proposed	0.74	0.92	21.24
			Linear	0.61	0.87	20.99
			KNN	0.71	0.87	28.07
			DBN	0.76	0.92	22.54
			BPCA	0.64	0.86	27.88
EC			Proposed	0.83	0.95	13.96
			Linear	0.72	0.92	18.34
			KNN	0.82	0.94	13.84
			DBN	0.85	0.94	15.46
			BPCA	0.79	0.93	18.89
OC	30		Proposed	0.68	0.88	19.47
			Linear	0.54	0.85	29.54
			KNN	0.67	0.90	23.40
			DBN	0.71	0.91	19.68

Table S6 (continued)

Species	% of Missing	Case	Method	R^2	IoA	MAPE(%)
			BPCA	0.57	0.86	24.71
EC			Proposed	0.80	0.94	16.51
			Linear	0.48	0.81	32.10
			KNN	0.75	0.93	15.89
			DBN	0.84	0.95	13.00
			BPCA	0.78	0.94	18.73

Table S7: Performance metrics of the imputation methods under Case 4.

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
NH_4^+	10	MCMS	Proposed	0.92	0.95	20.67
			Linear	0.74	0.87	53.23
			KNN	0.88	0.96	24.00
			DBN	0.91	0.95	24.67
			BPCA	0.93	0.97	22.13
NO_3^-			Proposed	0.85	0.95	23.92
			Linear	0.70	0.86	43.88
			KNN	0.83	0.94	23.84
			DBN	0.88	0.88	39.81
			BPCA	0.86	0.96	36.14
NH_4^+	20	MCMS	Proposed	0.83	0.95	25.87
			Linear	0.48	0.80	87.64
			KNN	0.79	0.93	45.33
			DBN	0.78	0.90	42.45
			BPCA	0.85	0.96	36.85
NO_3^-			Proposed	0.81	0.95	28.46
			Linear	0.41	0.77	91.45
			KNN	0.79	0.94	29.27
			DBN	0.53	0.78	67.21
			BPCA	0.62	0.88	56.23
NH_4^+	10	MCMI	Proposed	0.93	0.98	9.57
			Linear	0.31	0.69	63.88
			KNN	0.85	0.95	16.43
			DBN	0.72	0.92	22.14
			BPCA	0.92	0.98	12.86
NO_3^-			Proposed	0.84	0.95	14.82
			Linear	0.15	-0.01	162.09
			KNN	0.69	0.91	21.98
			DBN	0.57	0.84	34.34
			BPCA	0.76	0.93	27.68
NH_4^+	20	MCMI	Proposed	0.95	0.98	13.63
			Linear	0.00	0.23	54.07
			KNN	0.72	0.91	35.28
			DBN	0.84	0.96	22.07
			BPCA	0.95	0.98	15.42
NO_3^-			Proposed	0.74	0.90	22.81
			Linear	0.12	0.57	80.46
			KNN	0.67	0.88	25.61
			DBN	0.54	0.82	47.62
			BPCA	0.73	0.92	32.61

Table S8: Performance metrics of the imputation methods under Case 5.

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
Fe	10	MCMS	Proposed	0.90	0.93	25.83
			Linear	0.16	-0.24	57.18
			KNN	0.69	0.78	53.78
			DBN	0.77	0.93	31.88
			BPCA	0.90	0.96	26.58
Ca			Proposed	0.84	0.95	53.84
			Linear	0.06	-0.02	63.88
			KNN	0.72	0.85	79.88
			DBN	0.83	0.95	46.35
			BPCA	0.87	0.96	56.30
Si			Proposed	0.83	0.92	31.88
			Linear	0.01	-0.24	43.74
			KNN	0.77	0.91	30.26
			DBN	0.86	0.95	24.42
			BPCA	0.84	0.89	38.23
Ti			Proposed	0.62	0.88	35.93
			Linear	0.02	-0.39	52.24
			KNN	0.53	0.78	41.22
			DBN	0.62	0.87	50.67
			BPCA	0.63	0.87	48.09
Fe	20	MCMS	Proposed	0.81	0.92	30.14
			Linear	0.33	0.66	111.94
			KNN	0.74	0.83	57.10
			DBN	0.89	0.97	22.96
			BPCA	0.84	0.95	34.96
Ca			Proposed	0.83	0.95	73.39
			Linear	0.70	0.90	55.72
			KNN	0.79	0.88	92.22
			DBN	0.87	0.95	61.60
			BPCA	0.84	0.94	81.37
Si			Proposed	0.86	0.93	28.74
			Linear	0.76	0.93	25.78
			KNN	0.86	0.93	29.56
			DBN	0.76	0.93	28.74
			BPCA	0.79	0.92	38.40
Ti			Proposed	0.52	0.82	40.11
			Linear	0.43	0.78	39.53
			KNN	0.48	0.73	47.21
			DBN	0.51	0.77	77.15
			BPCA	0.56	0.80	78.77
Fe	10	MCMI	Proposed	0.65	0.71	20.22
			Linear	0.04	0.46	56.84
			KNN	0.67	0.80	19.17
			DBN	0.72	0.90	14.43
			BPCA	0.73	0.82	17.06
Ca			Proposed	0.87	0.91	18.66
			Linear	0.25	0.68	49.85
			KNN	0.85	0.89	21.26
			DBN	0.67	0.89	32.19
			BPCA	0.81	0.91	19.80
Si			Proposed	0.91	0.97	12.65
			Linear	0.83	0.87	29.43
			KNN	0.74	0.91	21.37
			DBN	0.56	0.83	38.28
			BPCA	0.78	0.93	26.58
Ti			Proposed	0.82	0.95	9.69
			Linear	0.11	-0.34	56.65
			KNN	0.68	0.89	15.85
			DBN	0.39	0.46	23.20
			BPCA	0.42	0.69	31.26
Fe	20	MCMI	Proposed	0.79	0.81	20.98
			Linear	0.00	0.19	72.71
			KNN	0.66	0.72	32.84
			DBN	0.82	0.94	19.52
			BPCA	0.86	0.94	17.65
Ca			Proposed	0.59	0.83	22.67
			Linear	0.05	-0.16	39.87
			KNN	0.33	0.68	32.53
			DBN	0.69	0.88	35.79
			BPCA	0.58	0.87	28.40

Table S8 (contiued)

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
Si			Proposed	0.84	0.95	16.37
			Linear	0.28	0.31	101.43
			KNN	0.70	0.91	24.20
			DBN	0.63	0.87	29.35
			BPCA	0.44	0.77	40.08
Ti			Proposed	0.69	0.84	22.97
			Linear	0.00	0.24	47.41
			KNN	0.56	0.81	31.61
			DBN	0.59	0.68	54.23
			BPCA	0.46	0.78	37.12

Table S9: Performance metrics of the imputation methods under Case 6.

Species	% of Missing	MCMI or MCMS	Method	R2	IoA	MAPE(%)
K	10	MCMS	Proposed	0.76	0.91	16.99
			Linear	0.10	0.54	37.61
			KNN	0.81	0.92	22.72
			DBN	0.83	0.93	13.24
			BPCA	0.79	0.93	18.87
NH_4^+			Proposed	0.85	0.94	42.55
			Linear	0.40	0.35	158.94
			KNN	0.84	0.82	62.48
			DBN	0.83	0.90	60.45
			BPCA	0.75	0.90	54.55
NO_3^-			Proposed	0.74	0.90	57.80
			Linear	0.53	0.54	158.74
			KNN	0.73	0.90	48.46
			DBN	0.55	0.52	157.23
			BPCA	0.59	0.84	70.03
K	20	MCMS	Proposed	0.90	0.96	15.47
			Linear	0.12	0.48	69.39
			KNN	0.86	0.96	20.51
			DBN	0.93	0.98	11.50
			BPCA	0.90	0.96	13.89
NH_4^+			Proposed	0.78	0.93	39.41
			Linear	0.03	0.10	234.85
			KNN	0.67	0.84	66.19
			DBN	0.74	0.86	39.67
			BPCA	0.49	0.79	49.58
NO_3^-			Proposed	0.68	0.89	52.58
			Linear	0.01	0.13	234.58
			KNN	0.65	0.87	57.17
			DBN	0.42	0.50	76.89
			BPCA	0.52	0.82	52.73
K	10	MCMI	Proposed	0.71	0.90	12.04
			Linear	0.00	0.09	43.73
			KNN	0.67	0.86	15.09
			DBN	0.77	0.93	10.26
			BPCA	0.74	0.92	12.27
NH_4^+			Proposed	0.94	0.98	12.81
			Linear	0.41	0.76	62.08
			KNN	0.84	0.92	22.87
			DBN	0.84	0.95	19.83
			BPCA	0.91	0.97	13.85
NO_3^-			Proposed	0.85	0.95	17.58
			Linear	0.24	0.58	73.33
			KNN	0.78	0.93	20.12
			DBN	0.61	0.86	34.08
			BPCA	0.74	0.92	31.49
K	20	MCMI	Proposed	0.79	0.93	13.96
			Linear	0.02	0.40	33.43
			KNN	0.75	0.90	15.81
			DBN	0.75	0.92	13.54
			BPCA	0.80	0.94	12.77
NH_4^+			Proposed	0.90	0.95	21.24
			Linear	0.00	0.21	52.32
			KNN	0.63	0.86	38.68
			DBN	0.86	0.96	21.19
			BPCA	0.94	0.98	15.97
NO_3^-			Proposed	0.87	0.96	26.96
			Linear	0.12	0.57	85.94
			KNN	0.65	0.86	24.32
			DBN	0.53	0.84	37.23
			BPCA	0.72	0.91	34.03

Table S10: Performance metrics of the imputation methods under Case 7.

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
K	10	MCMS	Proposed	0.86	0.96	14.41
			Linear	0.03	0.10	67.61
			KNN	0.82	0.83	35.49

Table S10 (continued)

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
			DBN	0.89	0.97	12.40
			BPCA	0.86	0.96	13.12
OC			Proposed	0.81	0.95	12.40
			Linear	0.00	0.00	61.98
			KNN	0.70	0.85	24.09
			DBN	0.61	0.87	20.10
			BPCA	0.65	0.89	17.53
EC			Proposed	0.80	0.94	15.37
			Linear	0.01	-0.07	45.14
			KNN	0.66	0.87	17.83
			DBN	0.66	0.89	16.55
			BPCA	0.71	0.91	18.08
K	20	MCMS	Proposed	0.90	0.97	13.40
			Linear	0.13	0.61	51.78
			KNN	0.77	0.92	22.49
			DBN	0.89	0.97	13.37
			BPCA	0.92	0.98	12.82
OC			Proposed	0.73	0.90	19.65
			Linear	0.18	0.47	41.49
			KNN	0.58	0.83	33.84
			DBN	0.63	0.88	25.93
			BPCA	0.63	0.86	20.21
EC			Proposed	0.81	0.90	13.66
			Linear	0.12	0.35	46.47
			KNN	0.69	0.90	20.81
			DBN	0.72	0.91	16.67
			BPCA	0.75	0.90	14.36
K	10	MCMI	Proposed	0.92	0.97	16.10
			Linear	0.35	0.70	67.05
			KNN	0.93	0.98	15.41
			DBN	0.92	0.97	13.18
			BPCA	0.91	0.97	15.23
OC			Proposed	0.79	0.91	22.36
			Linear	0.24	0.05	48.25
			KNN	0.74	0.88	19.66
			DBN	0.76	0.88	27.42
			BPCA	0.72	0.87	26.35
EC			Proposed	0.90	0.91	15.43
			Linear	0.71	0.87	57.79
			KNN	0.91	0.96	11.44
			DBN	0.86	0.87	17.56
			BPCA	0.88	0.92	17.42
K	20	MCMI	Proposed	0.95	0.97	15.55
			Linear	0.57	0.84	49.68
			KNN	0.89	0.86	18.95
			DBN	0.95	0.98	12.98
			BPCA	0.95	0.98	17.73
OC			Proposed	0.86	0.88	19.36
			Linear	0.00	-0.26	40.51
			KNN	0.63	0.77	21.40
			DBN	0.77	0.84	24.18
			BPCA	0.71	0.85	20.16
EC			Proposed	0.73	0.88	15.00
			Linear	0.00	0.20	42.39
			KNN	0.70	0.91	17.90
			DBN	0.69	0.91	20.31
			BPCA	0.62	0.86	17.27

Table S11: Performance metrics of the imputation methods under Case 8.

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
NH_4^+	10	MCMI	Proposed	0.87	0.96	20.04
			Linear	0.00	0.15	133.22
			KNN	0.66	0.87	43.66
			DBN	0.72	0.82	47.55
			BPCA	0.80	0.75	70.29
NO_3^-			Proposed	0.84	0.95	19.78

Table S11 (continued)

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
			Linear	0.10	0.62	75.76
			KNN	0.74	0.92	23.25
			DBN	0.58	0.90	30.93
			BPCA	0.68	0.90	30.93
OC			Proposed	0.88	0.91	19.81
			Linear	0.23	0.60	76.21
			KNN	0.80	0.91	20.87
			DBN	0.79	0.90	24.06
			BPCA	0.78	0.84	23.86
EC			Proposed	0.88	0.96	12.01
			Linear	0.44	0.75	33.96
			KNN	0.84	0.95	12.75
			DBN	0.85	0.93	15.65
			BPCA	0.87	0.94	23.02
NH_4^+	20	MCMI	Proposed	0.90	0.97	23.66
			Linear	0.08	0.44	140.73
			KNN	0.67	0.84	44.03
			DBN	0.84	0.94	46.61
			BPCA	0.87	0.95	49.73
NO_3^-			Proposed	0.83	0.94	25.66
			Linear	0.42	0.78	58.26
			KNN	0.84	0.93	24.45
			DBN	0.64	0.82	48.05
			BPCA	0.77	0.92	39.99
OC			Proposed	0.86	0.89	19.89
			Linear	0.49	0.34	41.37
			KNN	0.78	0.89	20.38
			DBN	0.79	0.89	24.39
			BPCA	0.77	0.86	21.74
EC			Proposed	0.87	0.96	13.46
			Linear	0.51	0.77	38.88
			KNN	0.84	0.95	11.57
			DBN	0.79	0.94	14.54
			BPCA	0.83	0.95	18.28
NH_4^+	10	MCMS	Proposed	0.81	0.92	30.32
			Linear	0.00	0.36	159.90
			KNN	0.76	0.92	51.24
			DBN	0.87	0.91	25.72
			BPCA	0.86	0.85	29.97
NO_3^-			Proposed	0.75	0.88	44.35
			Linear	0.01	0.41	169.02
			KNN	0.71	0.88	48.13
			DBN	0.71	0.60	100.14
			BPCA	0.79	0.78	51.22
OC			Proposed	0.86	0.96	12.69
			Linear	0.05	-0.01	104.49
			KNN	0.69	0.90	20.87
			DBN	0.61	0.87	22.32
			BPCA	0.60	0.86	21.36
EC			Proposed	0.89	0.97	12.07
			Linear	0.34	0.70	32.59
			KNN	0.79	0.93	14.71
			DBN	0.78	0.93	12.49
			BPCA	0.66	0.90	17.80
NH_4^+	20	MCMS	Proposed	0.82	0.94	32.44
			Linear	0.29	0.67	52.16
			KNN	0.75	0.90	40.32
			DBN	0.88	0.97	23.75
			BPCA	0.55	0.81	34.39
NO_3^-			Proposed	0.66	0.88	47.50
			Linear	0.37	0.74	50.80
			KNN	0.53	0.84	38.43
			DBN	0.68	0.77	43.09
			BPCA	0.54	0.83	38.60
OC			Proposed	0.63	0.88	20.19
			Linear	0.17	0.59	31.67
			KNN	0.58	0.82	27.06
			DBN	0.57	0.85	21.11
			BPCA	0.55	0.85	21.74

Table S11 (continued)

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
EC			Proposed	0.82	0.94	14.53
			Linear	0.17	0.61	31.56
			KNN	0.76	0.91	15.14
			DBN	0.83	0.94	14.00
			BPCA	0.76	0.92	17.16

Table S12: Performance of imputation methods under Case 9.

Species	% of Missing	MCMI or MCMS	Method	R^2	IoA	MAPE(%)
K	10	MCMI	Proposed	0.96	0.96	17.61
			Linear	0.02	-1.13	142.37
			KNN	0.89	0.91	24.24
			DBN	0.96	0.98	11.29
			BPCA	0.96	0.98	11.45
NH_4^+			Proposed	0.86	0.96	20.30
			Linear	0.00	0.15	133.22
			KNN	0.64	0.86	43.61
			DBN	0.76	0.84	42.19
			BPCA	0.79	0.74	72.12
NO_3^-			Proposed	0.84	0.96	19.87
			Linear	0.10	0.62	75.76
			KNN	0.76	0.93	23.13
			DBN	0.07	0.15	84.06
			BPCA	0.68	0.89	31.22
OC			Proposed	0.89	0.89	20.63
			Linear	0.23	0.60	76.21
			KNN	0.81	0.91	20.67
			DBN	0.79	0.91	24.73
			BPCA	0.78	0.84	23.96
EC			Proposed	0.88	0.96	11.78
			Linear	0.44	0.75	33.96
			KNN	0.82	0.94	13.22
			DBN	0.78	0.89	19.52
			BPCA	0.87	0.94	22.90
K	20	MCMI	Proposed	0.81	0.93	14.56
			Linear	0.37	0.73	27.62
			KNN	0.75	0.91	18.56
			DBN	0.81	0.93	15.41
			BPCA	0.84	0.95	13.87
NH_4^+			Proposed	0.89	0.97	23.51
			Linear	0.08	0.44	140.73
			KNN	0.67	0.84	44.16
			DBN	0.86	0.95	40.71
			BPCA	0.87	0.95	51.33
NO_3^-			Proposed	0.83	0.94	25.89
			Linear	0.42	0.78	58.26
			KNN	0.84	0.93	24.83
			DBN	0.66	0.87	52.42
			BPCA	0.77	0.93	39.75
OC			Proposed	0.86	0.90	19.69
			Linear	0.49	0.34	41.37
			KNN	0.78	0.89	20.15
			DBN	0.79	0.90	22.81
			BPCA	0.76	0.85	21.54
EC			Proposed	0.86	0.95	14.09
			Linear	0.51	0.77	38.88
			KNN	0.81	0.93	13.02
			DBN	0.68	0.89	18.64
			BPCA	0.81	0.84	18.85

References

(1) Moritz, S.; Sardá, A.; Bartz-Beielstein, T.; Zaefferer, M.; Stork, J. *arXiv preprint arXiv:1510.03924* **2015**.

(2) Ibrahim, I. A.; Khatib, T. *Energy Conversion and Management* **2017**, *138*, 413–425.

(3) Liu, B.; Wu, J.; Zhang, J.; Wang, L.; Yang, J.; Liang, D.; Dai, Q.; Bi, X.; Feng, Y.; Zhang, Y., et al. *Environmental Pollution* **2017**, *222*, 10–22.

(4) Kim, E.; Hopke, P. K.; Qin, Y. *Journal of the Air & Waste Management Association* **2005**, *55*, 1190–1199.

(5) Kim, E.; Hopke, P. K. *Atmospheric Environment* **2007**, *41*, 567–575.

(6) Tian, S.; Pan, Y.; Wang, Y. *Atmospheric Chemistry and Physics* **2016**, *16*, 1–19.

(7) Polissar, A. V.; Hopke, P. K.; Paatero, P.; Malm, W. C.; Sisler, J. F. *Journal of Geophysical Research: Atmospheres* **1998**, *103*, 19045–19057.