Supplementary Information

Novel insights on causes of disproportionate trends between particulate NO_3^- and NO_x emissions in Canadian urban atmospheres

Qinchu Fan[1], Xiaohong Yao[1], and Leiming Zhang[2]

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

Text S1. Analysis of the relative importance of 15 major variables on f-NO_3^- formation using a Random Forest model

The Random Forest (RF) model embedded in the *"ranger"* R package was used here to simulate the nonlinear relationship between daily f-NO_3^- and a set of predictors comprising 15 variables (Wright and Ziegler, 2017; Schmid et al., 2016). The 15 variables include NO_2 and $PM_{2.5}$ concentrations, near-surface meteorological variables including wind speed and direction (WS, WD), relative humidity (RH), temperature (T), dew point, visibility, and surface pressure, boundary-layer height (BLH), surface shortwave radiation (SSR), total cloud cover (TCC), total precipitation (TP), and two seasonal timing variables. The two seasonal timing variables was encoded with two Fourier terms—sin and cos of day-of-year (DOY)—to capture the annual cycle without year-end discontinuity. Hourly WS, WD, RH, T, surface pressure, dew point, and visibility were obtained from the Edmonton International Airport station via *"worldMet"* R package and aggregated to daily averages. Daily BLH, SSR, TCC, and TP were sourced from ERA5 (Copernicus Climate Data Store) and harmonized to the time-zone-specified daily data. Models were trained using 70% of 2010–2019 data that are randomly selected. The remaining 30% data were used for testing the model performance (Fig. S8). We used 3,000 trees with a minimum node size of 6. For the 30% test data set, the ordinary-least-squares slope of predictions versus observations was 0.53, with RMSE, MAE, and R^2 being 1.2, 0.42, and 0.62, respectively. The relative importance of all the input variables is shown in the right panel in Fig.S8 with importance increasing from the top to the bottom. Of these variables temperature is ranked as having the highest importance (Importance ≈ 0.104), followed by $PM_{2.5}$

(0.063), NO_2 (0.057), and BLH (0.053). Moisture metrics (dew point 0.031; RH 0.012) and SSR (0.020) were informative, but secondary. DOY Fourier terms indicated seasonal asymmetry (cosine, 0.026; sine, 0.011). Transport/optics terms rank also have certain extent of importance (WD, 0.012; WS, 0.007; visibility, 0.009), but weaker than variables mentioned above. Cloud, precipitation, and pressure played minimum role on the daily scale (TCC, 0.002; TP, 0.002; pressure, 0.001).

Text S2. Description of the hypothesis: Was the observed f-NO_3^- entirely derived from the condensation of HNO_{3gas}* in Canadian urban atmospheres during the cold season?

To address the hypothesis if the observed f-NO_3^- was entirely derived from the condensation of HNO_{3gas}* (≈HNO_{3gas} + N_2O_{5gas}) in Canadian urban atmospheres during the cold season, we first analyzed the 20 daily samples having f-NO_3^- concentration of > 4 μg m^{-3} in 2010, noting that there are a total of 116 samples in this year. The mass ratios of f-NO_3^- to HNO_{3gas} varied from 21 to >100, with a median value of 63, in these 20 samples. We first hypothesized that the observed f-NO_3^- was entirely derived from the condensation of HNO_{3gas}* (≈HNO_{3gas} + N_2O_{5gas}), and this hypothesis was tested using the following theory. Considering mass conservation in the process of f-NO_3^- formation, the net generation of f-NO_3^- should equal to the net loss of HNO_{3gas}^*. Given that the concentrations of f-NO_3^- are one to two orders of magnitude higher than those of HNO_{3gas}^*, the quasi-static condition must be satisfied if the above hypothesis is true, i.e., $d[HNO_{3gas}^*]/dt \approx 0$, and this condition is driven by the assumption that the rapid formation of HNO_{3gas}^* is balanced by its removal through conversion into f-NO_3^-. Such

an approximation has been commonly used in literature to estimate extremely low concentrations of reactive species such as the OH radical. However, the low time-resolution of our data (with daily samples) does not allow us to directly estimate formation rates of f-NO_3^-. Indirectly, if the condition $d[HNO_{3gas}^*]/dt \approx 0$ accompanied with the balanced rapid formation and removal of HNO_{3gas}^* does not hold for the 20 samples, the above hypothesis fails. If the removal rate of HNO_{3gas}^* declined to near zero, the concentration of HNO_{3gas}^* would substantially increase due to its rapid formation. In such a case, a low concentration of f-NO_3^- would be associated with a low removal rate of HNO_{3gas}^* in the process of f-NO_3^- formation. Thus, the examination of $d[HNO_{3gas}^*]/dt \approx 0$ accompanied with the balanced rapid formation and removal of HNO_{3gas}^* is further reframed into testing a weaker hypothesis: whether HNO_{3gas}^* concentrations significantly increased under conditions with low f-NO_3^- concentrations, compared to cases with high f-NO_3^- concentrations. Such an increase is a necessity, but not sufficient condition to accept the quasi-static approximation. If no such increase was observed, the hypothesis must be rejected, i.e., the observed low concentrations of HNO_{3gas}^* could not have given rise to high f-NO_3^- concentrations prior to their detection at the sampling site, based on the principle of mass conservation.

Text S3. Simulation of secondary f-NO_3^- formation using the Flexible 0-D Atmospheric Model (F0AM)

The Flexible 0-D Atmospheric Model (F0AM) with the CB6r2 chemical mechanism was used to simulate f-NO_3^- formation on ten days with f-NO_3^- concentration higher than 9 μg m^{-3} in 2010 (00:00–23:00 local time), results from which can quantify the

maximum retainable secondary f-NO_3^- (Luecken et al., 2019; Wolfe et al., 2016). Simulations were run in MATLAB with a 1-h external step (adaptive internal sub-stepping) and full time-series output. The box model was constrained by the observed NO_2, O_3, and CO; all other species evolved freely. To avoid numerical instability, initial conditions were set to NO = 10^{-4} ppb and HNO_3 = 10^{-3} ppb. Meteorological inputs including surface pressure (P), temperature (T), relative humidity (RH), and solar zenith angle (SZA) were computed from site latitude/longitude and local time. Surface albedo was fixed at 0.1. Photolysis frequencies followed F0AM's SZA-driven parameterization using a TOMS-like total O_3 column from OMI/Aura Level-3 (OMTO3d, 1°×1°) retrieved via NASA GES DISC Giovanni and interpolated to the model time resolution. The dilution constant k_{dil} was inferred from CO as a quasi-conservative tracer. HNO_3 dry deposition was represented as $k_{dep}=(v_d/100)/H$, where v_d is the dry-deposition velocity (cm s^{-1}) and H is the boundary-layer height (m).

From model reaction rates we diagnosed total gas-phase HNO_3 production P_{gas} and the OH + NO_2 → HNO_3 channel P_{OH}. Nighttime heterogeneous N_2O_5 hydrolysis was included as a diagnostic upper bound and applied only when SZA > 90°:

$$P_{N_2O_5}^{het} = Yk_{het}[N_2O_5] \quad (1)$$

$$k_{het} = \frac{\gamma \bar{c} S_A}{4} \quad (2)$$

where Y is the stoichiometric HNO_3 yield per N_2O_5 uptake, γ is the uptake coefficient of N_2O_5 (removal probability of N_2O_5 per collision with the wet aerosol surface), $\bar{c}$ is the mean molecular speed of N_2O_5, and S_A is the aerosol surface area concentration.

Over each 24-h window we integrated total chemical production and diagnostic losses:

$$P_{chem} = \int (P_{gas} + P_{N_2O_5}^{het})dt \quad (3)$$

$$P_{loss} = \int (k_{dil}[HNO_3]^{+} + k_{dep}[HNO_3])dt \quad (4)$$

where P_{gas} is the total gas-phase HNO_3 production rate (ppb s^{-1}), and $[HNO_3]^+$ = max ($[HNO_3] - [HNO_3]_{bkg}$, 0) with $[HNO_3]_{bkg}$ = 0 to provide a local-increment upper bound. The retainable upper bound of secondary particulate nitrate was then calculated as:

$$NO_{3\,sec,max}^{-} = \alpha \max(P_{chem} - P_{loss}, 0) \quad (5)$$

where α converts ppb to μg m^{-3} for nitrate using the daily mean T and P. To avoid double counting, $P_{N_2O_5}^{het}$ (Eq. 1) was added only if the gas-phase mechanism lacked an explicit $N_2O_5 \rightarrow HNO_3$ pathway.

In a baseline configuration designed to represent a typical mid-latitude winter urban setting (γ = 0.01, S_A=200 , Y=1.5; Mielke et al., 2016; Zang et al., 2022), the retainable secondary particulate nitrate was 2.5 μg m^{-3} on January 19, 2010 (the highest observed f-NO_3^- in 2010) and 0.3 μg m^{-3} on December 7, 2010 (the second highest observed f-NO_3^- in 2010) , corresponding to 14% and 1.4% of observed f-NO_3^-, respectively (Table S2a). Using empirical estimates of aerosol surface area derived from $PM_{2.5}$ with RH growth and T (S_A=583 and 554 for 2010-01-29 and 2010-12-07; γ held at 0.01), the retainable secondary burden increased to 4.4 and 6.7 μg m^{-3}, i.e., 25% and 39%, respectively, of the observed values (Table S2b). The latter parameters, however, resulted in the predicted f-NO_3^- to as high as 86 μg m^{-3} on February 28, 2010, which

were six times of the observed value and is practically impossible in Canada. Thus, all predicted values larger than the observations were excluded for further analysis.

When examining the well-behaved simulation cases in Table S2ab, all predicated values were substantially smaller than the observations, except for 1 December 2010 (S_A=338 and γ at 0.015). Even in the empirical-estimate cases, where predictions are generally larger, the contribution of retainable secondary particulate nitrate never exceeds 45% of the observed f-NO_3^-. Overall, the test results suggested that the primary f-NO_3^- dominantly contributed to the higher values of f-NO_3^- in winter.

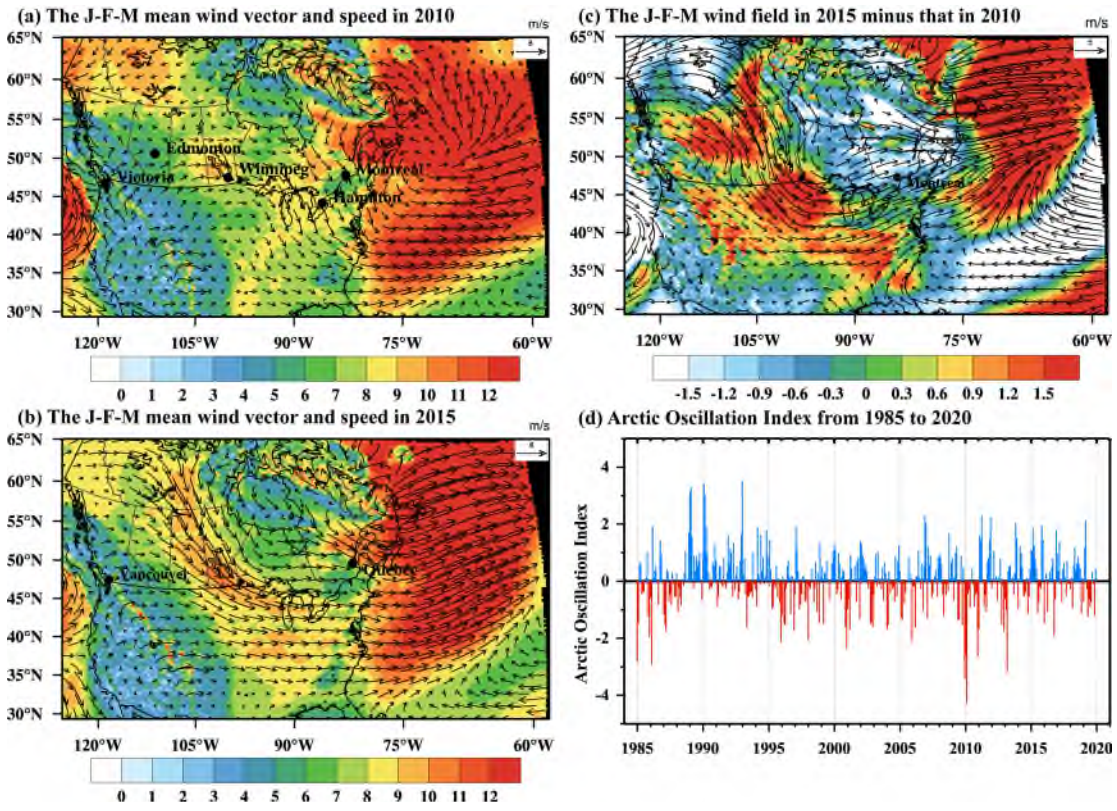

Figure S1. Mean wind vector in January-February-March (JFM) in 2010 (a) and 2015 (b), the difference in the JFM mean wind vector between 2010 and 2015 (2015 minus 2010) (c), and Arctic Oscillation Index from 1990 to 2020 (d)

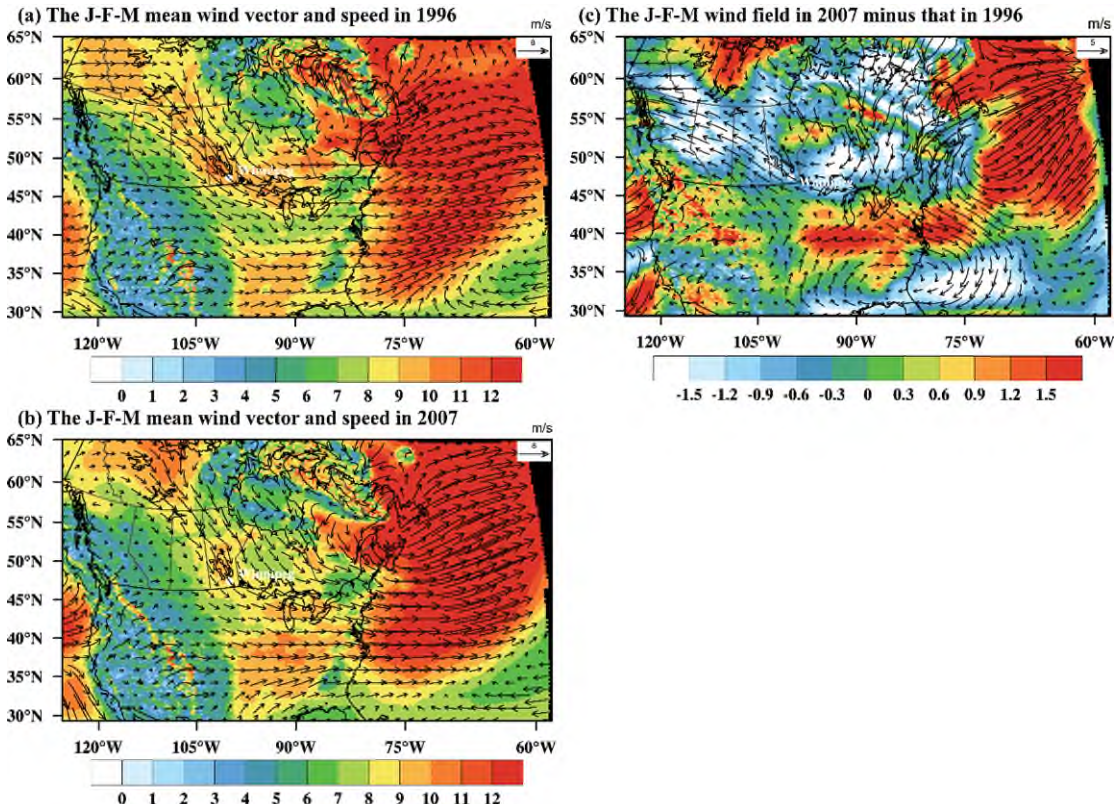

Figure S2. Same as in Figure S1a-c except for 1996 (a) and 2007 (b).

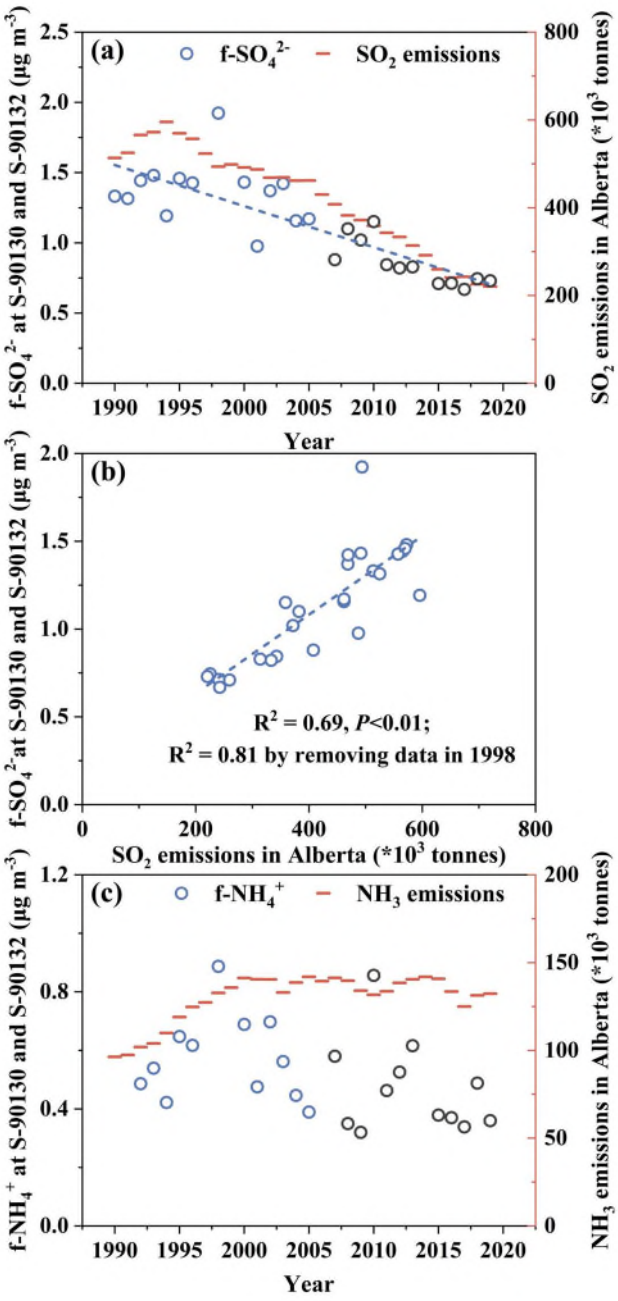

Figure S3. (a) Annual variations of mass concentrations of f-SO_4^{2-} at S-90130 and S-90132 in Edmonton and provincial total SO_2 emissions, (b) f-SO_4^{2-} vs. SO_2 emissions, and (c) same as in (a) except for f-NH_4^+ and NH_3 emissions. Blue and black markers represent the data points obtained at S-90130 and S-90132, respectively.

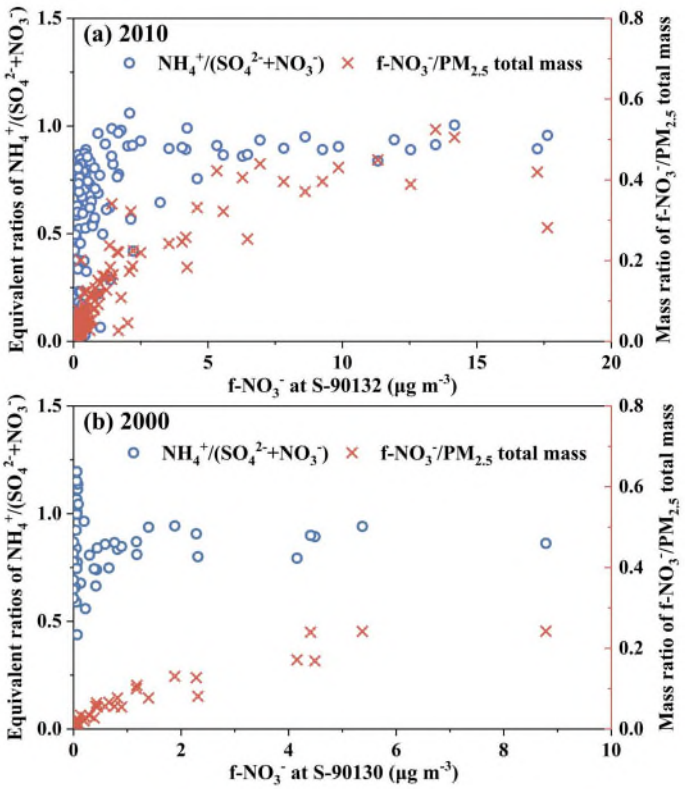

Figure S4. Variations of equivalent ratio of $NH_4^+/(SO_4^{2-}+NO_3^-)$ and mass ratio of f-$NO_3^-/PM_{2.5}$ with f-NO_3^- concentration at S-901932 in Edmonton in 2010 (a) and 2000 (b).

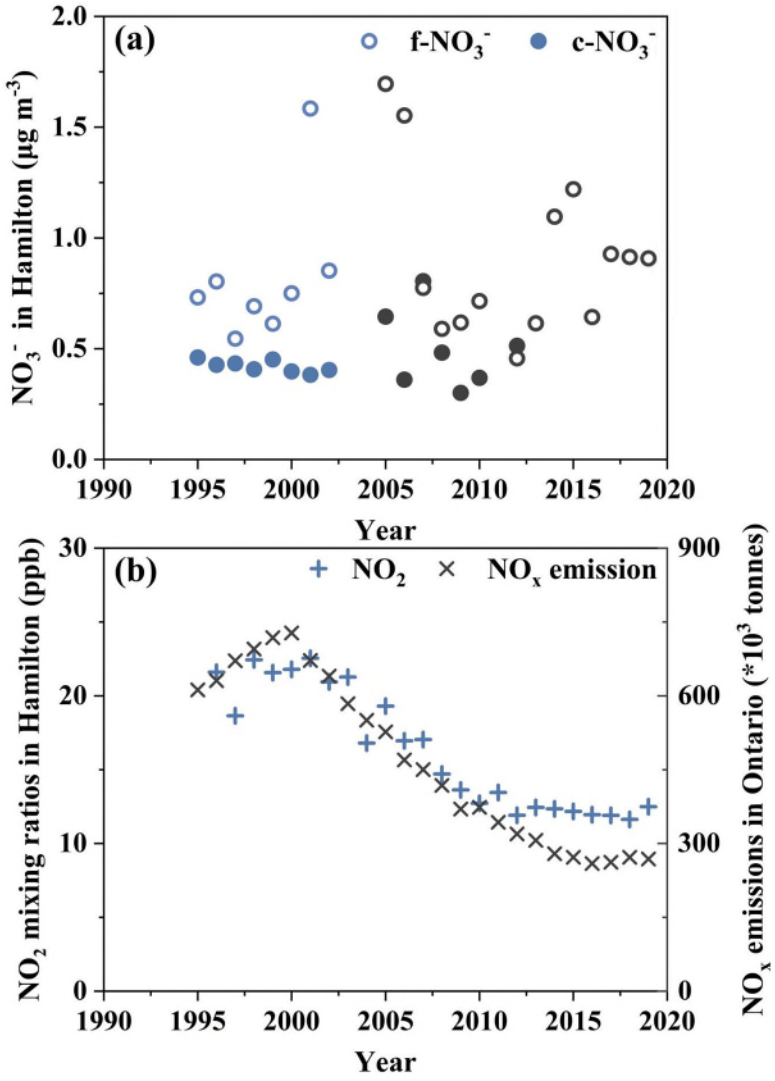

Figure S5. (a) Annual variations of mass concentrations of f-NO_3^- and c-NO_3^- in Hamilton, and (b) annual variations of mixing ratio of NO_2 in Hamilton and provincial total NO_x emissions. Blue and black markers in (a) represent data points before and after 2005, respectively.

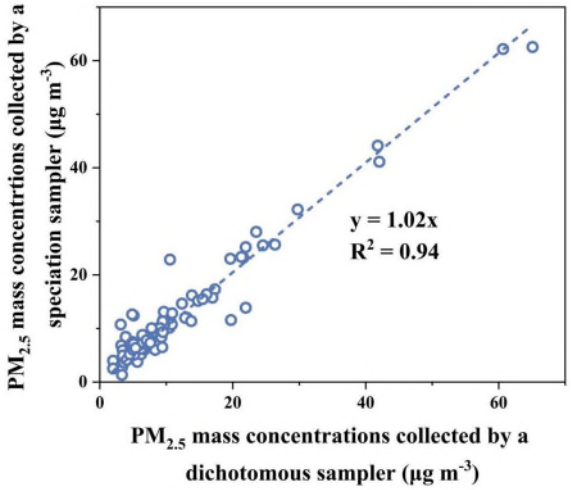

Figure S6. $PM_{2.5}$ mass concentrations measured by a speciation sampler against those measured by a dichotomous sampler at S-90132 in 2010.

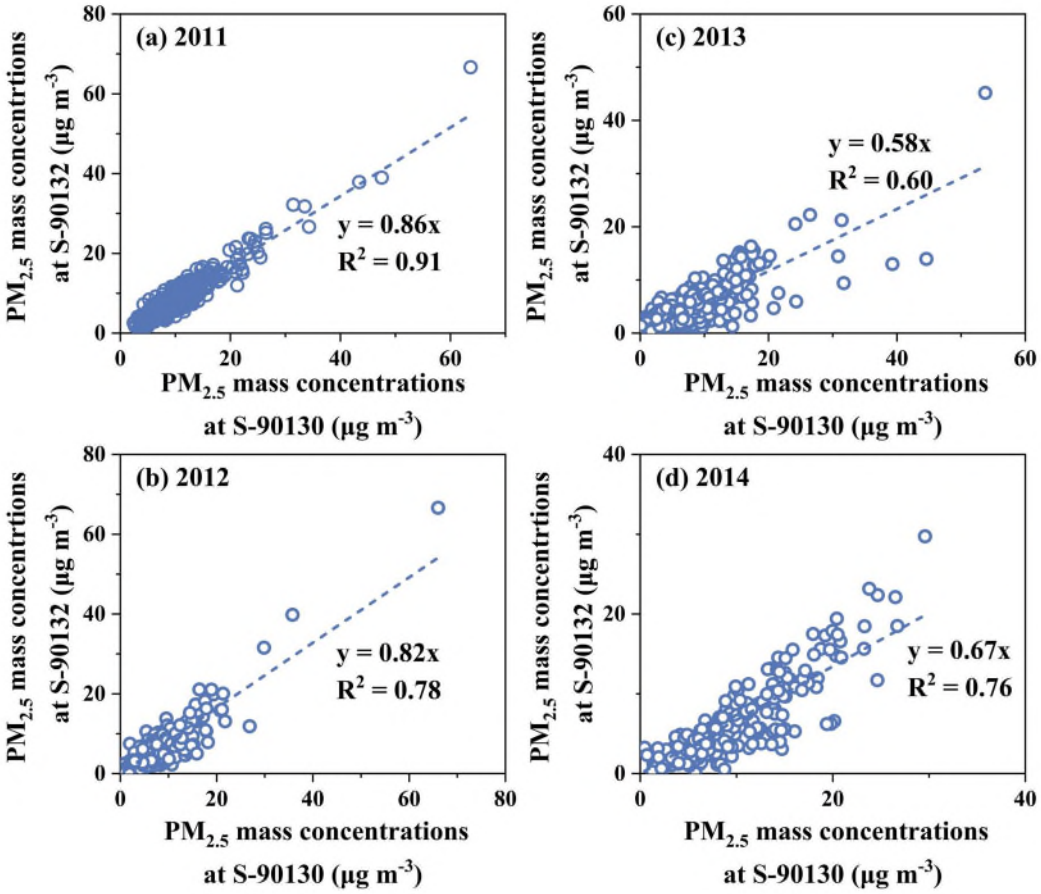

Figure S7. Inter-comparison between real-time measurements of $PM_{2.5}$ mass concentrations simultaneously made at S-90132 and S-90130 in (a) 2011, (b) 2012, (c) 2013, and (d) 2014

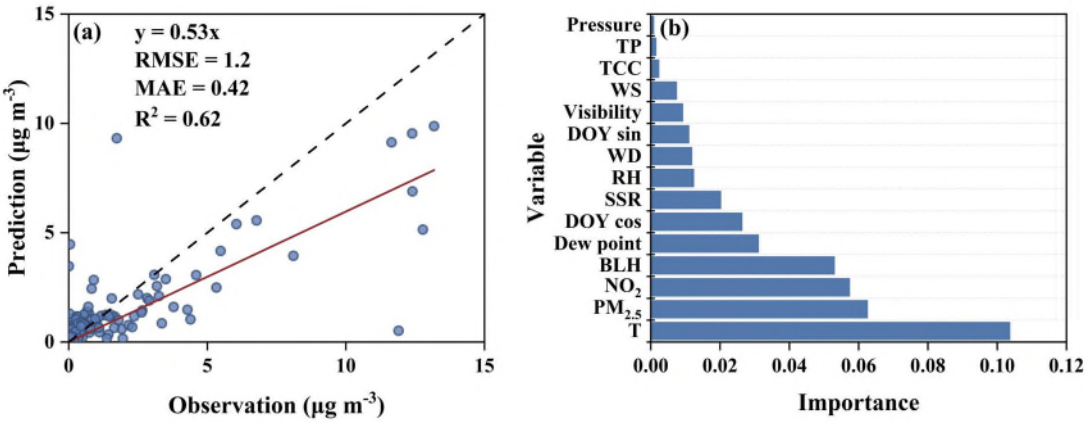

Figure S8. (a) Performance of the Random Forest model for daily f-NO_3^- during 2010–2019: predicted vs. observed on the 30% test data set (1:1 line shown; metrics inset), and (b) Permutation-based variable importance diagnosing the relative influence of meteorology (WS, WD, RH, T, Pressure, Visibility, Dew point, BLH, SSR, TCC, and TP), chemistry (NO_2 and $PM_{2.5}$), and seasonality (DOY sin, DOY cos) on f-NO_3^- formation.

Table S1. Information on pollutant concentrations, meteorological variables, and emissions datasets for each Canadian site.

City (Site)	Variable	Time span (Time resolution)
Edmonton (S-90132)	f-NO_3^- (μg m^{-3})	2007-2019 (with 2014 missing, 3-day)
	f-SO_4^{2-} (μg m^{-3})	2007-2019 (with 2014 missing, 3-day)
	f-NH_4^+ (μg m^{-3})	2007-2019 (with 2014 missing, 3-day)
	$PM_{2.5}$ (μg m^{-3})	2010-2014 (1-hr)
Edmonton (S-90130)	f-NO_3^- (μg m^{-3})	1990-2005 (with 1997 and 1999 missing, 6-day)
	c-NO_3^- (μg m^{-3})	1990-2005 (with 1997 and 1999 missing, 6-day)
	f-SO_4^{2-} (μg m^{-3})	1990-2005 (with 1997 and 1999 missing, 6-day)
	f-NH_4^+ (μg m^{-3})	1992-2005 (with 1997 and 1999 missing, 6-day)
	NO_2 mixing ratio (ppb)	1994-2019 (1-hr)
	$PM_{2.5}$ concentration (μg m^{-3})	2011-2014 (1-hr)
	HNO_{3gas} (μg m^{-3})	2010, 2015 (1-day)
	NO_x emissions (*10^3 tones)	1990-2019 (1-yr)
	SO_2 emissions (*10^3 tones)	1990-2019 (1-yr)
	NH_3 emissions (*10^3 tones)	1990-2019 (1-yr)
	T (°C)	2010, 2015 (1-day)
	WS (m s^{-1})	2010, 2015 (1-day)
	RH (%)	2010, 2015 (1-day)
Winnipeg	f-NO_3^- (μg m^{-3})	1990-2018 (with 1998, 2000, 2005, 2009, 2013, 2014 missing, 6-day)
	c-NO_3^- (μg m^{-3})	1990-2012 (with 1998, 2000, 2005, 2009 missing,6-day)
	NO_2 mixing ratio (ppb)	1990-2019 (1-hr)
	NO_x emissions (*10^3 tones)	1990-2018 (1-yr)
Quebec City	f-NO_3^- (μg m^{-3})	1995-2018 (with 1997, 2000, 2002, 2004, 2005, 2008 missing, 6-day)
	c-NO_3^- (μg m^{-3})	1995-2018 (with 1997, 2000, 2002, 2004, 2005, 2008 missing, 6-day)
	NO_2 mixing ratio (ppb)	1996-2019 (1-hr)
	NO_x emissions (*10^3 tones)	1995-2019 (1-yr)
Montreal	f-NO_3^- (μg m^{-3})	1997-2018 (with 2004, 2005, 2015, and 2016 missing, 6-day)
	c-NO_3^- (μg m^{-3})	1997-2014 (with 2004 and 2005 missing, 6-day)
	NO_2 mixing ratio (ppb)	1996-2019 (1-hr)
Hamilton	f-NO_3^- (μg m^{-3})	1995-2019 (with 2003, 2004, and 2011 missing, 3-day)
	c-NO_3^- (μg m^{-3})	1998-2012 (with 2003, 2004, and 2011 missing, 3-day)
	NO_2 mixing ratio (ppb)	1996-2019 (1-hr)
	NO_x emissions (*10^3 tones)	1995-2019 (1-yr)
Victoria	f-NO_3^- (μg m^{-3})	1990-2018 (with 1991, 1992, 1997, 1998, 2004, 2005, 2009, and 2016 missing, 6-day)
	c-NO_3^- (μg m^{-3})	1990-2012 (with 1991, 1992, 1997, 1998, 2004, 2005, 2009, and 2010 missing, 6-day)
	NO_2 mixing ratio (ppb)	1993-2019 (1-hr)
	NO_x emissions (*10^3 tones)	1990-2019 (1-yr)
Vancouver	f-NO_3^- (μg m^{-3})	1990-2018 (with 1992, 1994, 1999, 2003, 2005, and 2012-2014 missing, 6-day)
	c-NO_3^- (μg m^{-3})	1990-2012 (with 1992, 1994, 1999, 2003, 2005, and 2012-2014 missing, 6-day)
	NO_2 mixing ratio (ppb)	1990-2019 (1-hr)

Table S2. Sensitivity of the retainable upper bound secondary f-NO_3^- (f-$NO_{3\ sec,\ max}^-$) to the uptake coefficient of N_2O_5 (γ) and aerosol surface area concentration (S_A) on ten days in 2010 with f-NO_3^- concentration > 9 μg m^{-3}, evaluated with the F0AM box model.

Date	f-$NO_{3\ obs}^-$ (μg m^{-3})	γ	S_A (μm^2 cm^{-3})	P_{chem} (ppb)	P_{OH} (ppb)	$P_{N_2O_5}$ (ppb)	P_{loss} (ppb)	f-$NO_{3\ sec,\ max}^-$ (μg m^{-3})
2010/01/29	18	0.010	200	2	0.3	2	1	2
2010/12/07	17	0.010	200	1	0.2	0.8	1	0.3
2010/12/25	13	0.010	200	3	0.3	2	2	2
2010/02/25	11	0.010	200	3	1	2	8	0
2010/03/03	9	0.010	200	13	2	11	14	0
2010/01/08	9	0.010	200	1	0.2	0.9	1	0
2010/02/28	14	0.010	200	33	2	30	18	43
2010/02/01	13	0.010	200	13	0.6	13	8	14
2010/01/17	12	0.010	200	27	0.7	25	4	64
2010/12/01	10	0.010	200	7	0.3	7.0	2	15
2010/01/29	18	0.01	583	3	0.3	2	1	4
2010/12/07	17	0.01	554	3	0.2	3	1	7
2010/12/25	13	0.015	478	4	0.3	4	2	5
2010/01/17	12	0.01	95	6	0.7	4	4	5
2010/12/01	10	0.015	338	5	0.3	5	2	8
2010/01/08	9	0.01	182	1	0.2	0.8	1	0
2010/02/28	14	0.015	558	48	2	46	18	86
2010/02/01	13	0.01	360	16	0.6	15	8	22
2010/02/25	11	0.015	513	16	1	14	8	23
2010/03/03	9	0.015	513	13	2	11	10	10

- f-$NO_{3\ obs}^-$, the observed f- NO_3^- concentration; P_{chem}, total chemical production of f-NO_3^-; P_{OH}, production of f-NO_3^- via $OH+NO_2 \rightarrow HNO_3$; $P_{N_2O_5}$, production of f-NO_3^- by nighttime heterogeneous N_2O_5 hydrolysis; P_{loss}, losses of f-NO_3^- via dilution and dry deposition.
- The top block reports day-specific γ and S_A estimates and the bottom block shows a baseline representative of a mid-latitude winter urban setting (γ = 0.01, S_A=200).
- Cells highlighted in orange indicate cases where f-$NO_{3\ sec,\ max}^-$ exceeds f-$NO_{3\ obs}^-$.

References

Wright, M. N.; Ziegler, A. Ranger: a fast implementation of random forests for high dimensional data in C++ and R, *J. Stat. Softw.* **2017**, 77, 1-17.

Schmid, M.; Wright, M. N.; Ziegler, A. On the use of Harrell's C for clinical risk prediction via random survival forests, *Expert Syst. Appl.* **2016**, 63, 450-459.

Luecken, D. J.; Yarwood, G.; Hutzell, W. T. Multipollutant modeling of ozone, reactive nitrogen and HAPs across the continental US with CMAQ-CB6, *Atmos. Environ.* **2019**, 201, 62-72.

Wolfe, G. M.; Marvin, M. R.; Roberts, S. J.; Travis, K. R.; Liao, J. The framework for 0-d atmospheric modeling (F0AM) v3.1, *Geosci. Model Dev.* **2016**, 9, 3309-3319.

Mielke, L. H.; Furgeson, A.; Odame-Ankrah, C. A.; Osthoff, H. D. Ubiquity of $ClNO_2$ in the urban boundary layer of Calgary, Alberta, Canada, *Can. J. Chem.* **2016**, 94, 414-423.

Zang, H.; Zhao, Y.; Huo, J.; Zhao, Q.; Fu, Q.; Duan, Y.; Shao, J.; Huang, C.; An, J.; Xue, L.; Li, Z.; Li, C.; Xiao, H. High atmospheric oxidation capacity drives wintertime nitrate pollution in the eastern Yangtze River Delta of China, *Atmos. Chem. Phys.* **2022**, 22, 4355-4374.