

High-Resolution Air Quality Modeling over Santiago, Chile: A Comparative Assessment of EDGAR and INEMA Anthropogenic Emission Inventories

Mohsen Pourkiaei^{1*}, Jing Wang^{1,2}

¹ Dept. of Civil, Environmental and Geomatic Engineering (D-BAUG), Institute of Environmental Engineering (IfU), Air Quality and Particle Technology, ETH Zürich, Zürich, Switzerland

² Empa, Laboratory for Building Energy Materials and Components, Überlandstrasse 129, 8600 Dübendorf, Switzerland

*Corresponding author, email: spourkiaei@ethz.ch

Supplementary Notes

Inventory

Figure S1 illustrates the long-term evolution (1975–2020) and current sectoral composition of key emissions in Chile according to the EDGAR v8.1 Emissions in Urban Areas¹. Time-series panels show a sustained increase in CO₂ and total greenhouse gas (GHG) emissions, driven primarily by the industry and transport sectors, with waste and agriculture also contributing substantially to the GHG levels. NO_x emissions rose steadily through the early 2000s before plateauing, overwhelmingly dominated by the transport sector throughout the period. In contrast, PM_{2.5} emissions peaked around 2010, largely attributable to the residential/buildings sector, before declining, while waste and transport contributions continued to grow. Complementary stacked bar charts reveal the sectoral breakdown for multiple air pollutants in both percentage and absolute terms (×10⁴ Tonnes). Transport accounts for the largest share of NO_x (approximately 55–60 %), whereas PM_{2.5} shows a more distributed profile involving residential buildings, agriculture, transport, and waste. Agriculture is the dominant source of NH₃, and CO emissions are heavily associated with buildings' energy and transportation. The patterns highlight the persistent importance of the transport and residential sectors in driving Chile's air quality and climate-relevant emissions via EDGAR v8.1.

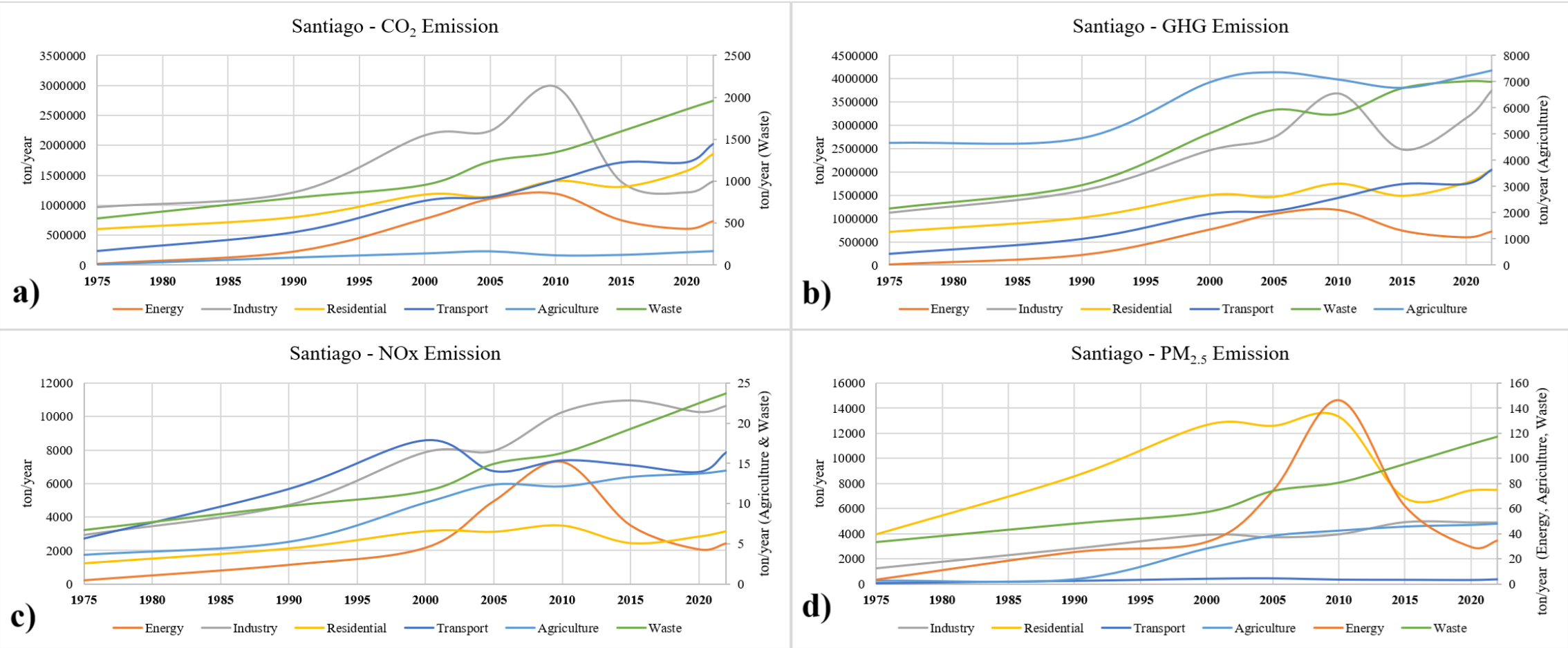


Figure S1. Long-term sectoral evolution of CO₂, GHG, NO_x, and PM_{2.5} emissions in Santiago according to the EDGAR v8.1 Emissions in Urban Areas dataset. (a) CO₂ emissions by energy, industry, residential, transport, agriculture, and waste sectors. (b) GHG emissions by the same six sectors. (c) NO_x emissions by sector, with agriculture and waste shown using the secondary y-axis. (d) PM_{2.5} emissions by sector, with energy, agriculture, and waste shown using the secondary y-axis. The time series illustrate changes in the magnitude and relative importance of individual emission sectors over the period shown.

Supplementary Results
Santiago Las Condes Station (Base Model)

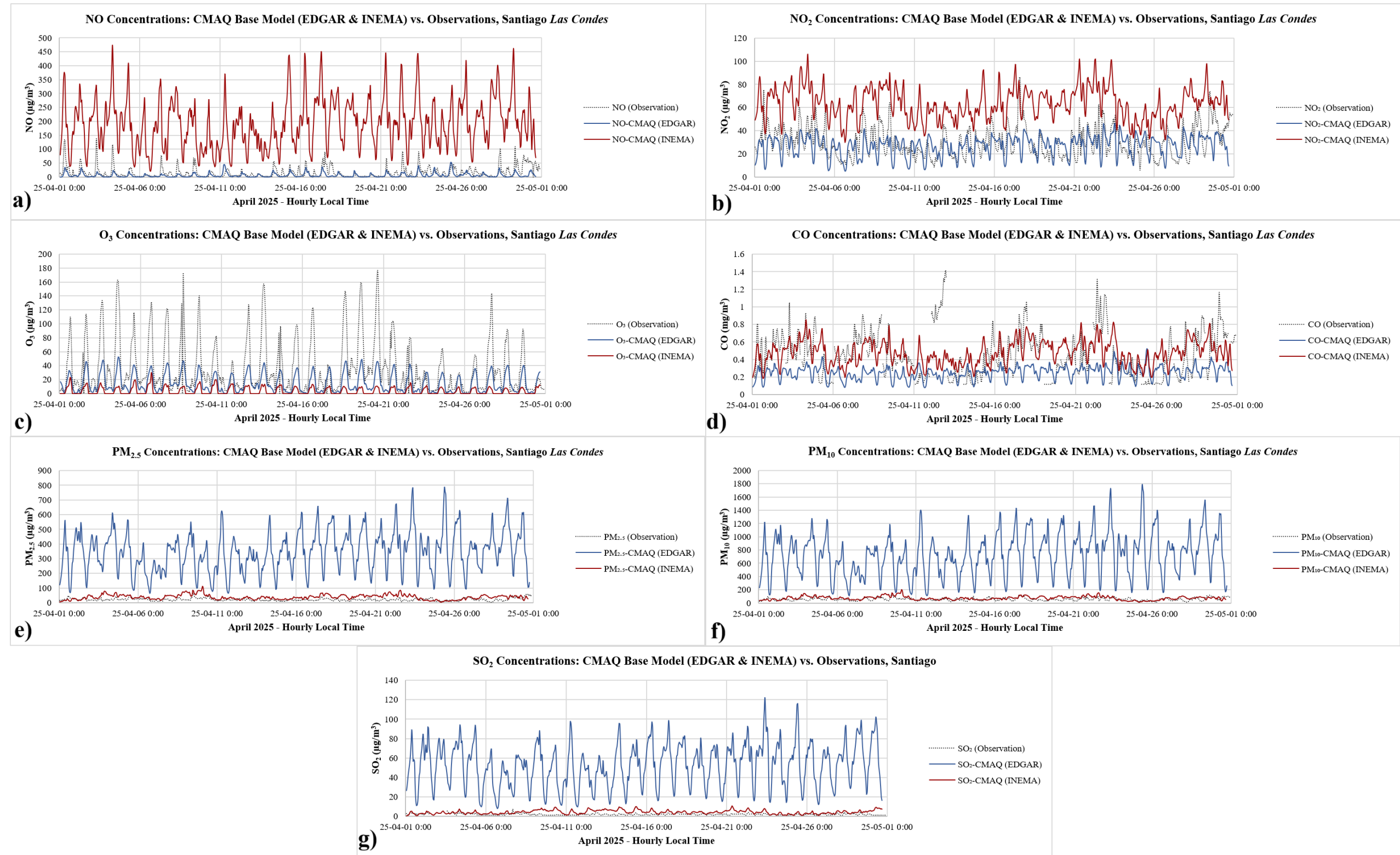


Figure S2. Hourly observed and CMAQ base model concentrations at Las Condes during April 2025. CMAQ simulations driven by base EDGAR v8.1 (2022) and INEMA v2 (2020) emissions are compared with SINCA observations for (a) NO, (b) NO₂, (c) O₃, (d) CO, (e) SO₂, (f) PM₁₀, and (g) PM_{2.5}. Model values correspond to the 1-km grid cell containing the monitoring station. SO₂ observations are provided from Santiago's representative dataset (La Florida station). Each panel shows hourly concentrations, with solid blue and red lines representing CMAQ outputs driven by EDGAR and INEMA emissions, respectively, and a dashed black line representing observations. All timestamps are reported in Santiago local time. Concentrations are reported in $\mu\text{g}\cdot\text{m}^{-3}$, except CO, which is reported in $\text{mg}\cdot\text{m}^{-3}$. Abbreviations: CMAQ = Community Multiscale Air Quality model; EDGAR = Emissions Database for Global Atmospheric Research; INEMA = Inventario Nacional de Emisiones Antropogénicas; NO = nitric oxide; NO₂ = nitrogen dioxide; O₃ = ozone; CO = carbon monoxide; PM_{2.5} = particulate matter with aerodynamic diameter $\leq 2.5 \mu\text{m}$; PM₁₀ = particulate matter with aerodynamic diameter $\leq 10 \mu\text{m}$; SO₂ = sulfur dioxide.

Santiago *O'Higgins* Station (Base Model)

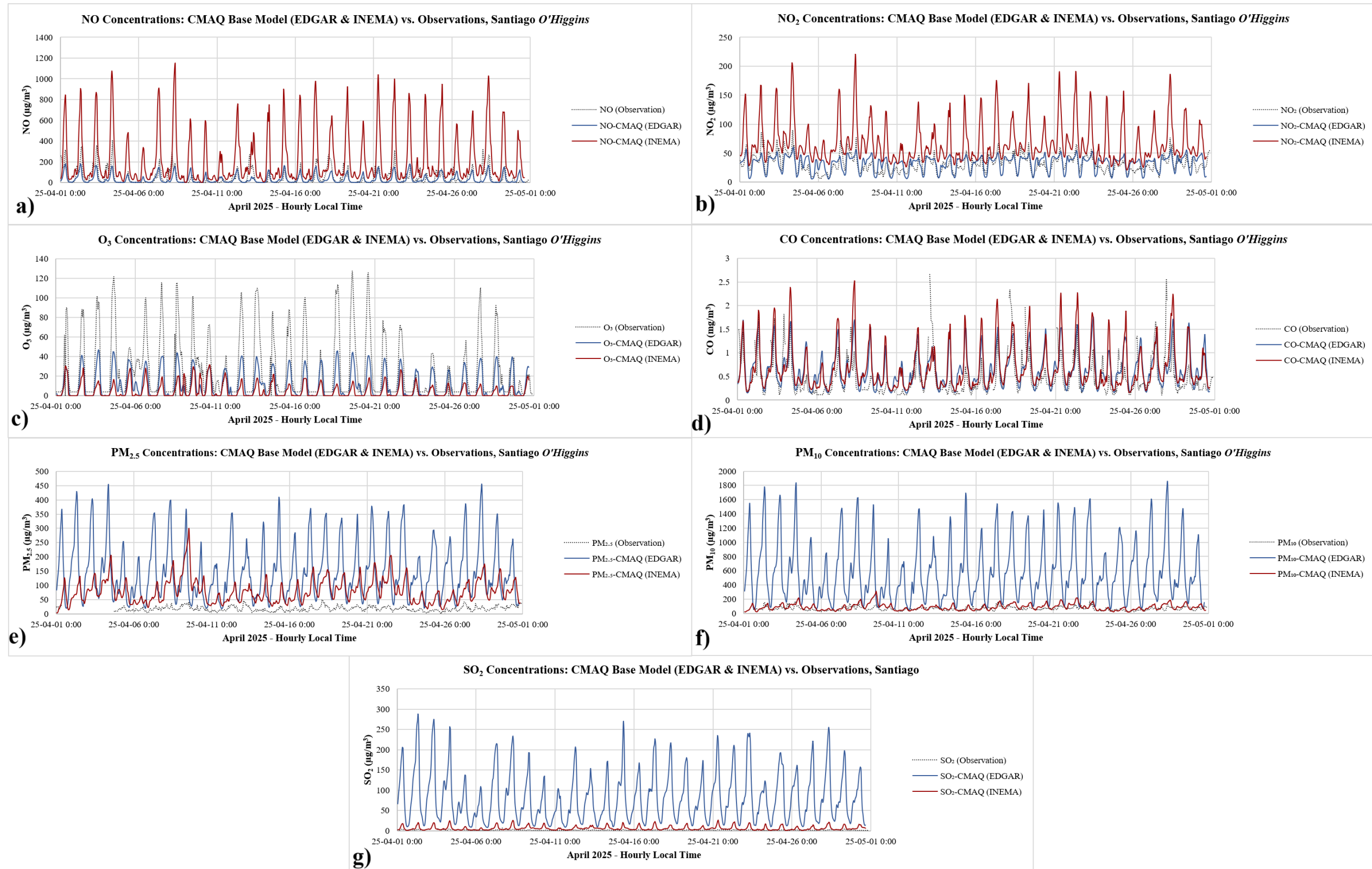


Figure S3. Hourly observed and CMAQ base model concentrations at O'Higgins during April 2025. CMAQ simulations driven by base EDGAR v8.1 (2022) and INEMA v2 (2020) emissions are compared with SINCA observations for (a) NO, (b) NO₂, (c) O₃, (d) CO, (e) SO₂, (f) PM₁₀, and (g) PM_{2.5}. Model values correspond to the 1-km grid cell containing the monitoring station. SO₂ observations are provided from Santiago's representative dataset (La Florida station). Each panel shows hourly concentrations, with solid blue and red lines representing CMAQ outputs driven by EDGAR and INEMA emissions, respectively, and a dashed black line representing observations. All timestamps are reported in Santiago local time. Concentrations are reported in µg·m⁻³, except CO, which is reported in mg·m⁻³. Abbreviations: CMAQ = Community Multiscale Air Quality model; EDGAR = Emissions Database for Global Atmospheric Research; INEMA = Inventario Nacional de Emisiones Antropogénicas; NO = nitric oxide; NO₂ = nitrogen dioxide; O₃ = ozone; CO = carbon monoxide; PM_{2.5} = particulate matter with aerodynamic diameter ≤ 2.5 µm; PM₁₀ = particulate matter with aerodynamic diameter ≤ 10 µm; SO₂ = sulfur dioxide.

Santiago Pudahuel Station (Base Model)

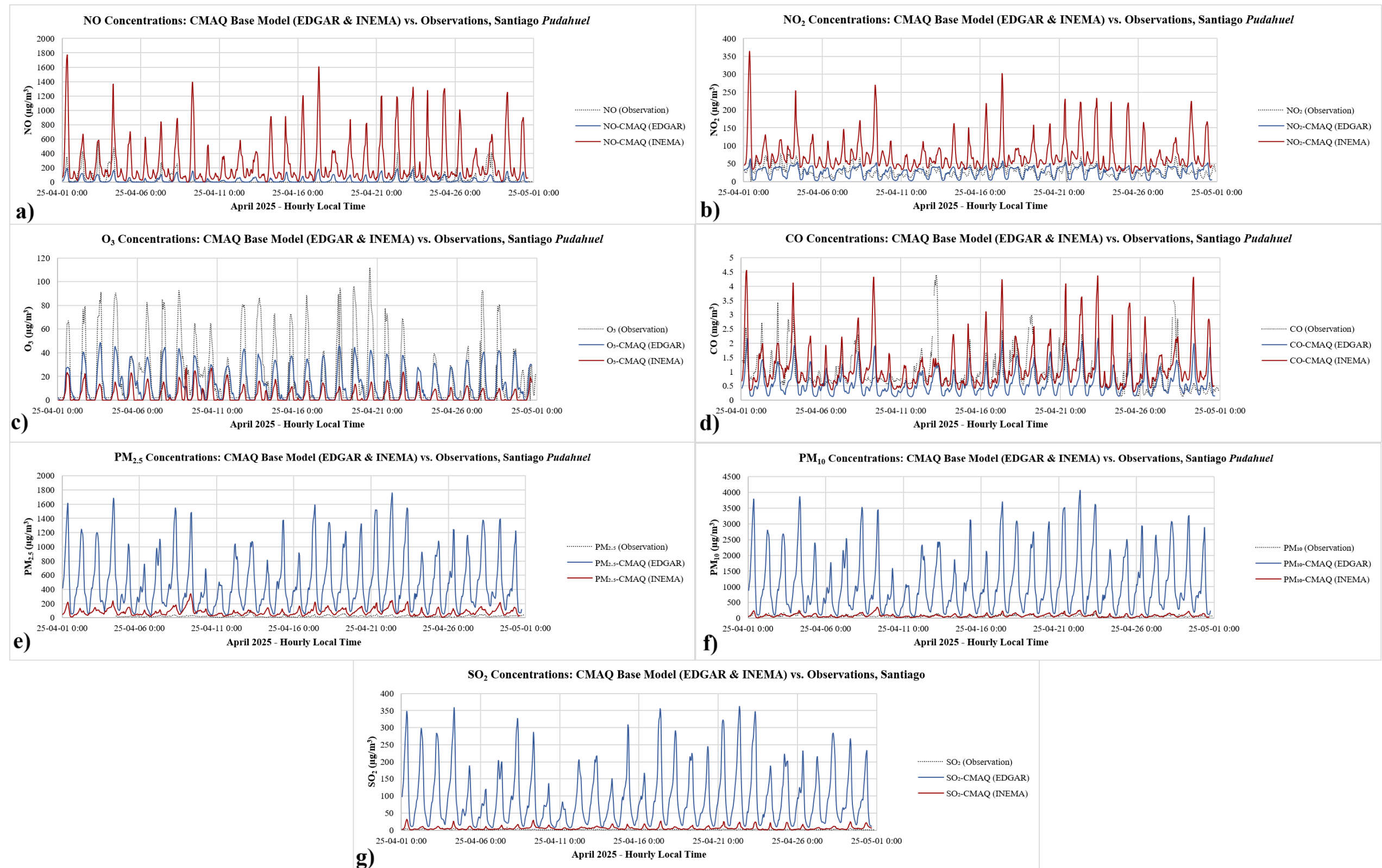


Figure S4. Hourly observed and CMAQ base model concentrations at Pudahuel during April 2025. CMAQ simulations driven by base EDGAR v8.1 (2022) and INEMA v2 (2020) emissions are compared with SINCA observations for (a) NO, (b) NO₂, (c) O₃, (d) CO, (e) SO₂, (f) PM₁₀, and (g) PM_{2.5}. Model values correspond to the 1-km grid cell containing the monitoring station. SO₂ observations are provided from Santiago's representative dataset (La Florida station). Each panel shows hourly concentrations, with solid blue and red lines representing CMAQ outputs driven by EDGAR and INEMA emissions, respectively, and a dashed black line representing observations. All timestamps are reported in Santiago local time. Concentrations are reported in $\mu\text{g}\cdot\text{m}^{-3}$, except CO, which is reported in $\text{mg}\cdot\text{m}^{-3}$. Abbreviations: CMAQ = Community Multiscale Air Quality model; EDGAR = Emissions Database for Global Atmospheric Research; INEMA = Inventario Nacional de Emisiones Antropogénicas; NO = nitric oxide; NO₂ = nitrogen dioxide; O₃ = ozone; CO = carbon monoxide; PM_{2.5} = particulate matter with aerodynamic diameter $\leq 2.5 \mu\text{m}$; PM₁₀ = particulate matter with aerodynamic diameter $\leq 10 \mu\text{m}$; SO₂ = sulfur dioxide.

Size-Resolved PM Mass & Number Diagnostics

Figures S5 and S6 provide supplementary diagnostics of the size-dependent aerosol mass and particle-number concentrations produced by the EDGAR- and INEMA-driven base simulations at Parque O'Higgins during April 2025. Figure S5 compares total and modal aerosol mass with PM_{10} , $PM_{2.5}$, and PM_{10} to examine differences in both overall loading and the relative magnitudes of the modeled size-dependent quantities. Figure S6 provides the corresponding total, fine-mode, and coarse-mode particle-number diagnostics. These model-to-model comparisons complement the observational $PM_{2.5}$ and PM_{10} evaluation in Figures S2–S4; they do not independently validate the simulated size distributions. Table S1 defines the aerosol-mass quantities shown in Figure S5.

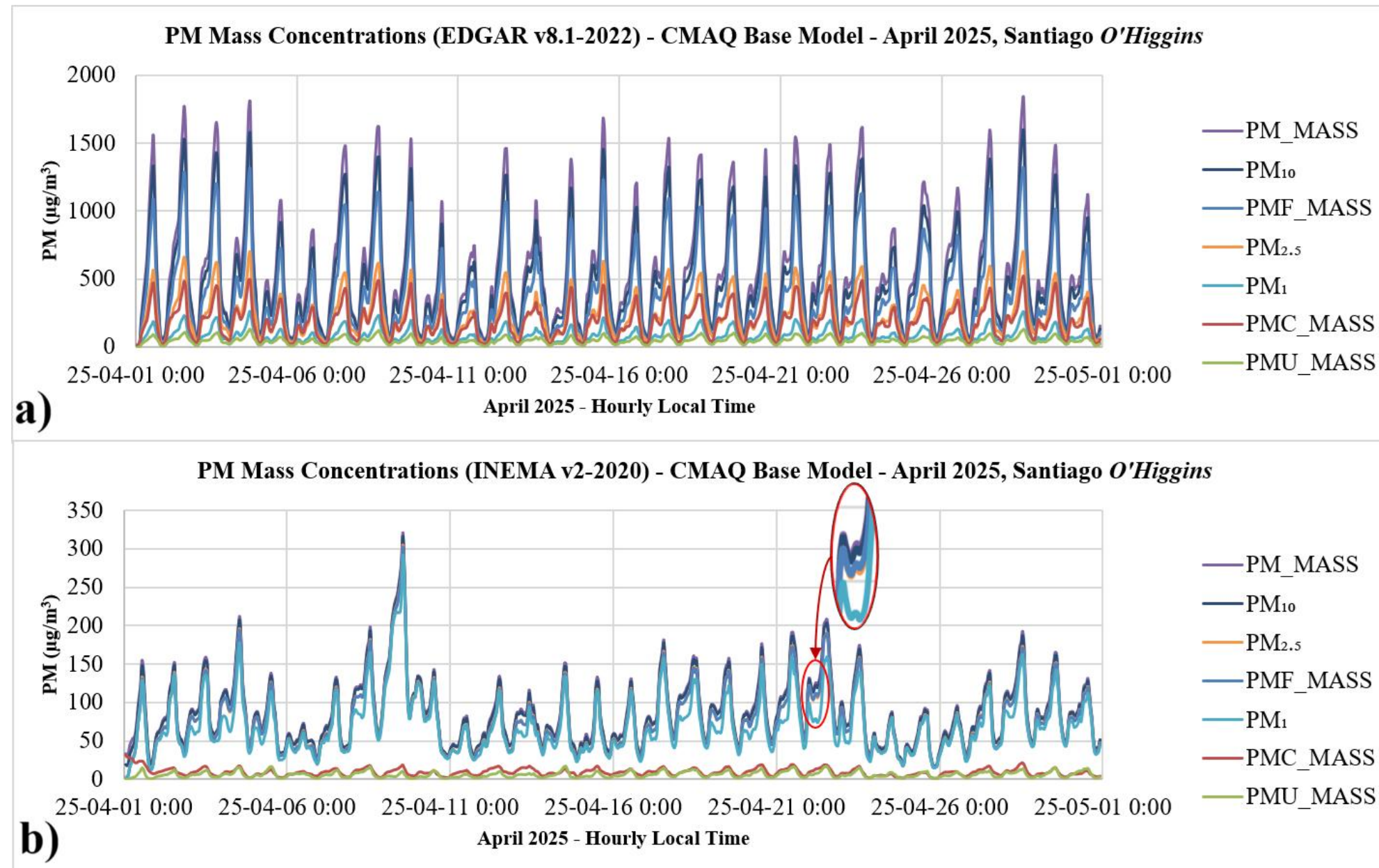


Figure S5. Hourly size-resolved aerosol mass concentrations from the 1-km CMAQ base simulations at Parque O'Higgins during April 2025. (a) CMAQ simulation driven by EDGAR v8.1 (2022) emissions. (b) CMAQ simulation driven by INEMA v2 (2020) emissions. The hourly model-derived aerosol diagnostics are distinguished by color: PM_MASS is shown in purple, PM_{10} in dark blue, PMF_MASS in blue, $PM_{2.5}$ in orange, PM_1 in light blue/cyan, PMC_MASS in red, and PMU_MASS in green.

PM_MASS denotes total modeled aerosol mass across the Aitken (I), accumulation (J), and coarse (K) modes; PMF_MASS represents the combined Aitken and accumulation modes; PMC_MASS represents coarse-mode aerosol mass, approximately corresponding to $PM_{10} - PM_{2.5}$; and PMU_MASS represents the Aitken-mode/ultrafine aerosol mass diagnostic. PM_{10} includes fine- and coarse-mode aerosol subject to the 10- μm inlet cutoff. Concentrations are expressed in $\mu g \cdot m^{-3}$ and timestamps are reported in local time. The vertical-axis ranges differ between panels because of the substantially different concentration magnitudes produced by the two base configurations. All curves represent CMAQ-derived quantities. The figure compares the modeled size-dependent aerosol mass distributions between the two base configurations and is not an observational evaluation; $PM_{2.5}$ and PM_{10} comparisons with measurements are provided in Figures S2–S4.

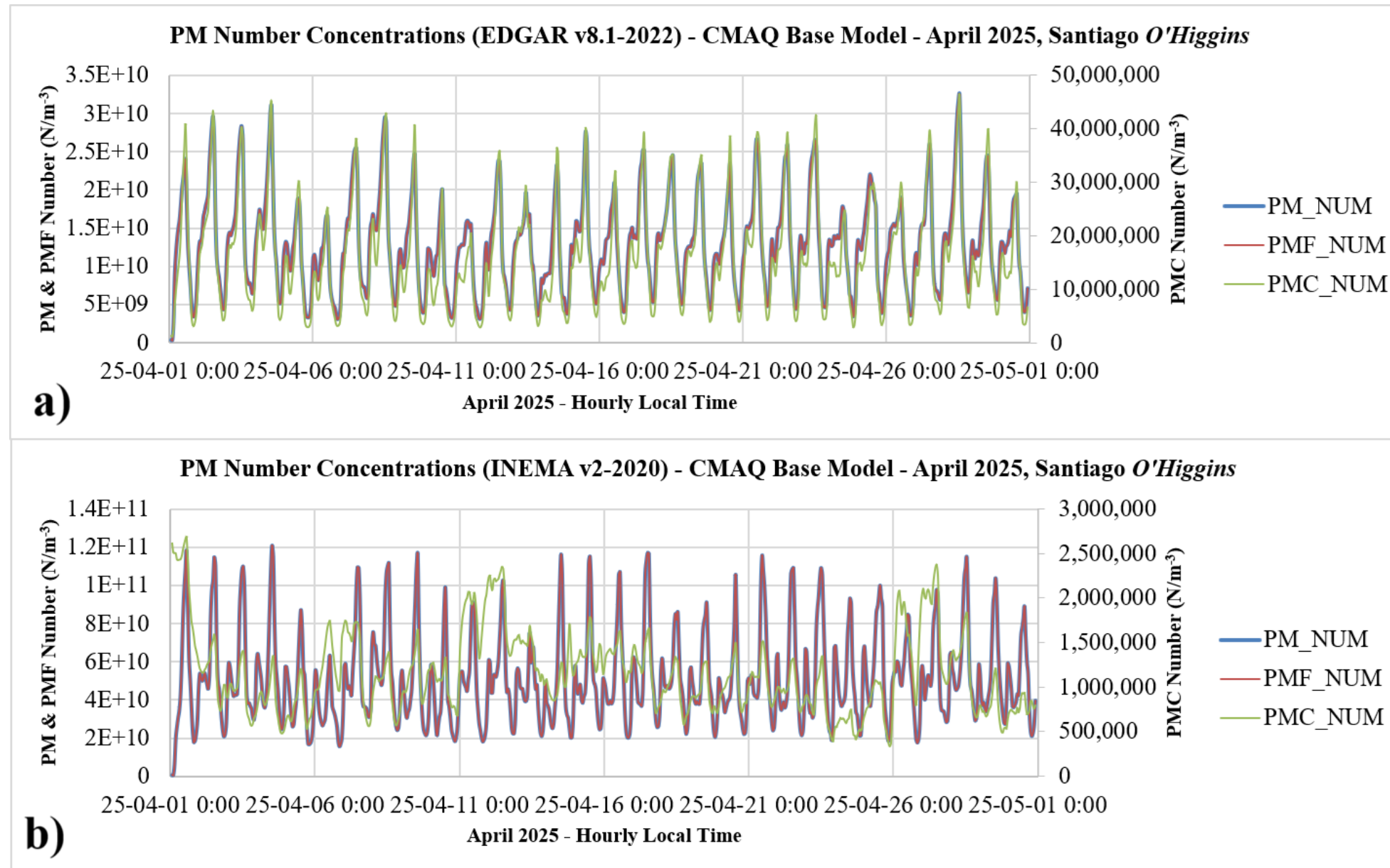


Figure S6. Hourly particulate-matter number concentrations from the 1-km CMAQ base simulations at Parque O'Higgins during April 2025. (a) CMAQ simulation driven by EDGAR v8.1 (2022) emissions. (b) CMAQ simulation driven by INEMA v2 (2020) emissions. The hourly model-derived particle-number diagnostics are distinguished by color: PM_NUM is shown in blue, PMF_NUM in red, and PMC_NUM in green. PM_NUM denotes total particle number concentration, PMF_NUM represents fine-mode particle number associated with the Aitken and accumulation modes, and PMC_NUM represents coarse-mode particle number concentration. PM_NUM and PMF_NUM are plotted against the left y-axis, whereas PMC_NUM is plotted against the right y-axis because of its substantially smaller magnitude. Number concentrations are expressed in $\text{N}\cdot\text{m}^{-3}$ and timestamps are reported in local time. The vertical-axis ranges differ between the EDGAR- and INEMA-driven panels to accommodate the different simulated concentration magnitudes. All curves represent CMAQ-derived quantities; no observational data are shown.

The EDGAR-driven run produced extremely high total aerosol mass (PM_MASS), with hourly peaks exceeding $1,800 \mu\text{g}\cdot\text{m}^{-3}$. The near-perfect overlap between PM_MASS and PM₁₀ indicates that virtually all modeled aerosol mass resides below $10 \mu\text{m}$ aerodynamic diameter, consistent with a small coarse-mode contribution in the present model configuration. PM_{2.5} concentrations were only marginally lower than PM₁₀/PM_MASS, while PM₁ remained substantially lower, revealing a strong dominance of the accumulation mode. The close tracking between PMF_MASS and PM_{2.5} further confirms that the majority of fine-mode mass falls below the regulatory $2.5 \mu\text{m}$ cutoff, as expected for an accumulation-mode geometric mean diameter typically centered between $0.3\text{--}0.8 \mu\text{m}$. Coarse-mode mass (PMC_MASS) was minor, and ultrafine mass (PMU_MASS) was negligible, as anticipated since mass scales with the cube of particle diameter ($m \propto D_p^3$). However, the absolute magnitudes are physically implausible for Santiago and point to substantial overestimation of primary particulate emissions (owed to industrial and residential combustion sectors) in the EDGAR v8.1 dataset.

In contrast, the INEMA v2 simulation yielded closer PM concentrations to observations, with PM_MASS ranging between $20\text{--}320 \mu\text{g}\cdot\text{m}^{-3}$, values consistent with observed urban winter/autumn haze episodes in central Chile. The same modal hierarchy was observed (PM_MASS $\approx$ PM₁₀ > PM_{2.5} $\approx$ PMF_MASS > PM₁ >> PMC_MASS >> PMU_MASS) but scaled to credible levels. The close agreement between PMF_MASS and PM_{2.5}, together with the relatively small PMC_MASS contribution, indicates that the INEMA v2 inventory produces substantially lower and more plausible modeled aerosol loadings that enable the CMAQ model to reproduce

the dominance of secondary inorganic and organic aerosols in the accumulation mode, while producing only limited coarse-particle emissions. The modest but visible separation between PM_{2.5} and PM₁, and the low PMU_MASS, further align with expected urban aerosol dynamics in which accumulation-mode particles dominate mass while ultrafine particles dominate number concentration.

The contrasting base simulations demonstrate substantial differences in modeled aerosol mass and modal distribution between the two inventory-based configurations. The INEMA-based simulation produces lower aerosol mass concentrations than the EDGAR-based simulation. However, these size-resolved diagnostics do not include corresponding observational comparisons and therefore do not independently establish the accuracy of the simulated size distribution or the overall superiority of either inventory.

The CMAQ base model simulations of particulate matter number concentration further reinforce the contrasting performance of the EDGAR v8.1 and INEMA v2 inventories while providing important insights into the aerosol size distribution over Santiago. The EDGAR-driven run produced extremely high total particle number concentrations (PM_NUM), frequently exceeding 3.5×10^{10} particles·m⁻³. The near-perfect overlap between PM_NUM and PMF_NUM demonstrates that particle number is overwhelmingly dominated by the fine mode (Aitken + accumulation modes), consistent with the mass distribution observed in previous analyses. The coarse-mode number concentration (PMC_NUM) remained orders of magnitude lower (secondary vertical axis), confirming that coarse particles contribute negligibly to total number despite their larger individual mass. This pattern aligns with the earlier finding that PM_MASS ≈ PM₁₀ and PMF_MASS ≈ PM_{2.5}, indicating a strongly fine-mode-dominated aerosol population in which accumulation-mode particles drive both mass and number concentrations. On the other hand, the INEMA v2 simulation exhibited substantially lower particle number concentrations (PM_NUM ranging from $\sim 1.5 \times 10^{10}$ to 1.2×10^{11} particles·m⁻³), representing a more realistic urban background²⁻⁵. Once again, PM_NUM and PMF_NUM were nearly identical, confirming the dominance of the fine mode in number concentration. The coarse-mode number (PMC_NUM) was even smaller in absolute terms than in the EDGAR run and displayed greater temporal variability. This result is fully consistent with the mass-based analysis, where INEMA v2 produced credible PM_{2.5} and PM₁₀ levels with limited coarse-mode mass (PMC_MASS). The combined mass and number distributions indicate that INEMA v2 produces an accumulation-mode-dominated modeled aerosol distribution of urban environments influenced by secondary inorganic and organic aerosol formation.

Both inventories exhibit the same fundamental modal structure: fine-mode particles dominate both mass and number, while the coarse mode contributes minimally to number and only modestly to mass. This behavior is physically expected for Santiago’s autumn aerosol, which is characterized by high anthropogenic emissions, secondary aerosol formation. The much higher absolute values in the EDGAR simulation (both mass and number) further support the conclusion that the global inventory overestimates primary particulate emissions in this region.

Table S1. Definition and typical size ranges of particulate matter variables used in the CMAQ simulations.

PM Variable	Definition	Typical Size Range
PM ₁	Total particulate matter ≤ 1 μm aerodynamic diameter	Dp ≤ 1 μm, “sharp-cut” PM ₁ using modal fractions
PM _{2.5}	Total particulate matter ≤ 2.5 μm	Dp ≤ 2.5 μm, regulatory fine PM
PM ₁₀	Total particulate matter ≤ 10 μm	Dp ≤ 10 μm, includes fine + coarse
PM_MASS	Total modeled aerosol mass	All aerosol mass, usually includes all CMAQ aerosol modes/species
PMF_MASS	Fine-mode aerosol mass	Mostly Aitken + accumulation modes, “F” = Fine
PMC_MASS	Coarse-mode aerosol mass	Coarse particles, often roughly PM ₁₀ – PM _{2.5}
PMU_MASS	Ultrafine aerosol mass	Aitken-mode aerosol mass, “U” = ultrafine

PM_MASS = Aitken (I) + Accumulation (J) + Coarse mode mass (K), PMF_MASS = Aitken + Accumulation, PMC_MASS = Coarse mode mass ≈ PM₁₀-PM_{2.5}, PMU_MASS = Aitken, PM₁₀ = Fine mode + Coarse mode (with inlet cutoff). PMU_MASS is the Aitken-mode/ultrafine aerosol mass diagnostic, while PMF_MASS represents the Aitken plus accumulation modes.

NVPOA & BCO Diagnostic (ISAM)

Inspection of the hourly ISAM NVPOA output revealed that the source apportionment was disproportionately influenced by a small number of extreme values in APOCJ_BCO (boundary condition aerosol primary organic carbon in the accumulation mode) and APNCOMJ_BCO (boundary condition aerosol primary non-carbon organic matter in the accumulation mode). In CMAQ-ISAM, the NVPOA tag class specifically represents the nonvolatile primary organic aerosol species APOC and APNCOM. However, in the standard AERO7/DESID configuration used with cb6r5_ae7_aq, emitted POC (primary organic carbon) and PNCOM (primary non-carbon organic matter) are not mapped predominantly to APOC and APNCOM. Instead, a substantial fraction of emitted primary organic aerosol (POA) mass is distributed among semi- and intermediate-volatility POA surrogates. Consequently, a small local contribution to the ISAM NVPOA class does not imply that local primary organic aerosol emissions are negligible, irrespective of the total emitted POA mass. After excluding the short BCO-dominated episode, locally emitted sources become substantially more prominent: transport is the leading contributor in the EDGAR simulation, whereas energy use in buildings is the leading contributor in the INEMA simulation.

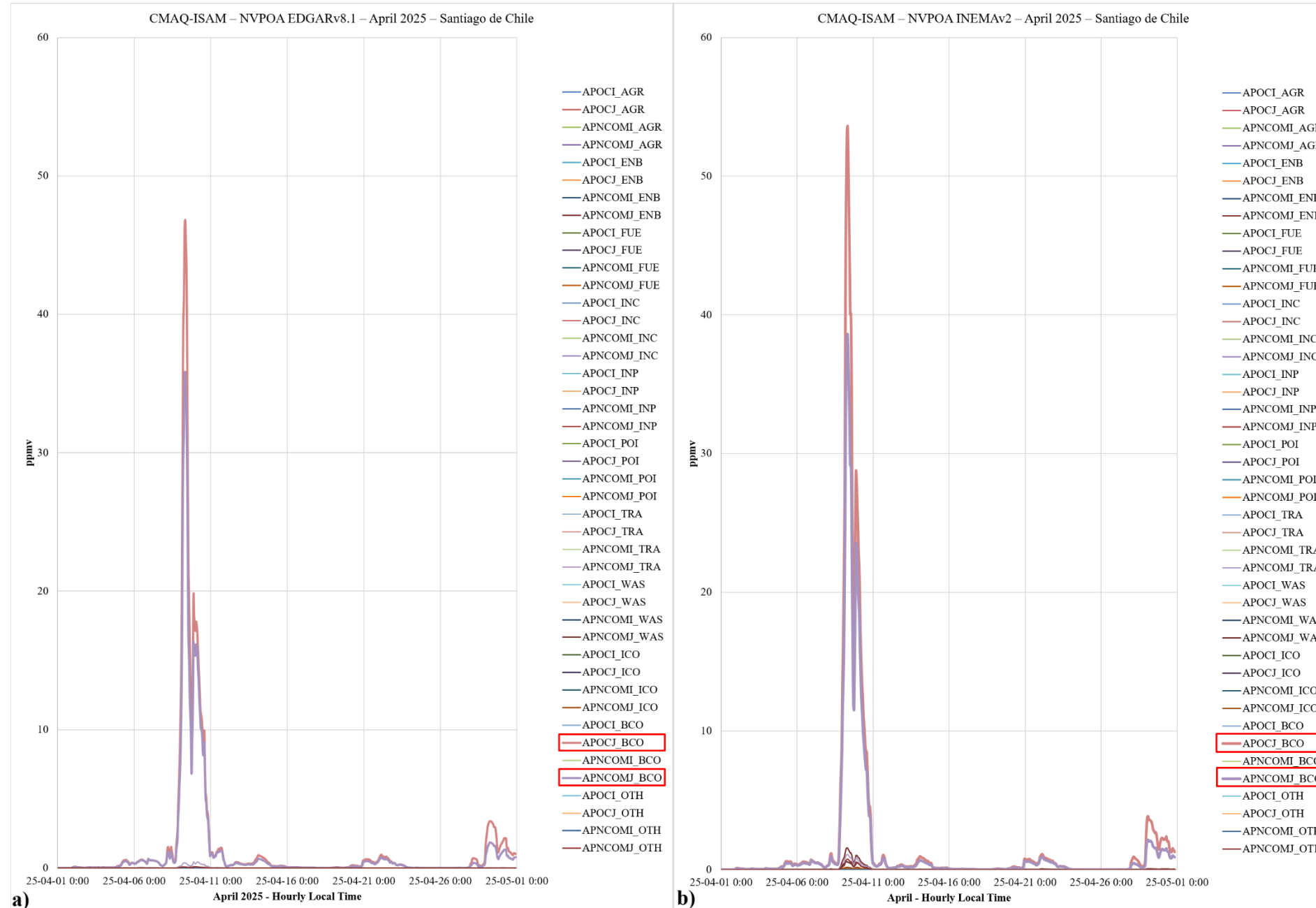


Figure S7. Full 720-h CMAQ-ISAM NVPOA source contributions over the Santiago 1-km domain during April 2025. Panel (a) shows the calibrated EDGAR v8.1 simulation, and panel (b) shows the calibrated INEMA v2 simulation. The time series display the ISAM-tagged I- and J-mode APOC and APNCOM components for each of the eight harmonized anthropogenic sectors, agriculture (AGR), energy for buildings (ENB), fuel exploitation (FUE), industrial combustion (INC), industrial processes (INP), power industry (POI), transport (TRA), and waste (WAS), together with the corresponding initial condition (ICO), boundary condition (BCO), and residual/other (OTH) contributions. Within the boundary condition components, APOCJ_BCO denotes accumulation-mode aerosol primary organic carbon and APNCOMJ_BCO denotes accumulation-mode aerosol primary non-carbon organic matter. The dominant boundary condition components, APOCJ_BCO (salmon/red) and APNCOMJ_BCO (purple), account for the pronounced short-duration NVPOA enhancement around 8–10 April in both inventory configurations (fire episodes). NVPOA represents the modeled nonvolatile APOC/APNCOM subset and should not be interpreted as total primary organic aerosol.

Inspection of the hourly ISAM output showed that the large monthly BCO contribution was concentrated within a short episode around 8–10 April rather than being sustained throughout the simulation. This episode coincided with independently documented smoke and haze over the Santiago region. A major fire in Paine on 7 April produced a dense smoke plume visible across the Metropolitan Region and remained active into 8 April, while meteorological observations at Arturo Merino Benítez Airport reported smoke on 7–8 April and haze on 9–10 April^{6,7}. These observations support the interpretation that the elevated boundary organic-aerosol contribution represents an episodic transported-aerosol event rather than a persistent feature of the April simulation. We therefore no longer interpret the full monthly BCO percentage as representative of the typical source distribution throughout April. Instead, the excluded episodic emission boundary contribution is discussed separately from the source-apportionment pattern during the remainder of the simulation. Outside this short BCO-dominated episode, transport becomes the largest EDGAR NVPOA contributor (~49%), whereas energy use in buildings becomes the largest INEMA contributor (~31%).

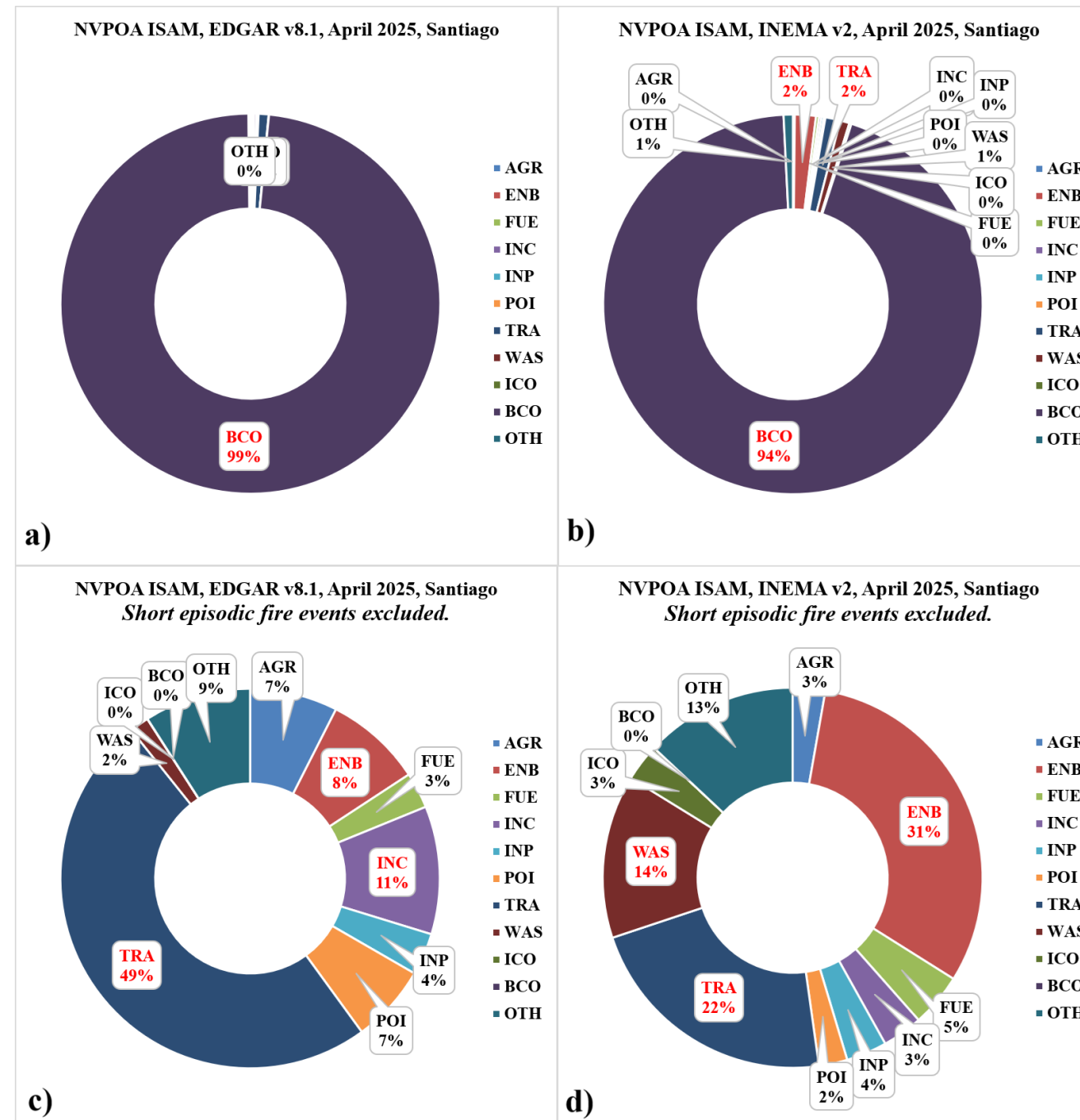


Figure S8. Effect of the short-duration April BCO-dominated episode on CMAQ-ISAM source apportionment of NVPOA over the Santiago 1-km domain. April-mean source contributions are shown for the calibrated (a) EDGAR v8.1 and (b) INEMA v2 simulations using the complete April 2025 period, and for the corresponding (c) EDGAR and (d) INEMA simulations after excluding the identified short-duration BCO-dominated episode (coincident with regional smoke/fire activity). The source categories are agriculture (AGR; blue), energy for buildings (ENB; red), fuel exploitation (FUE; light green), industrial combustion (INC; purple), industrial processes (INP; cyan), power industry (POI; orange), transport (TRA; dark blue), waste (WAS; dark red/brown), initial conditions (ICO; olive green), boundary conditions (BCO; dark purple), and residual/other contributions (OTH; teal). When the complete month is considered, boundary conditions (BCO) account for 99% and 94% of modeled NVPOA in the EDGAR and INEMA configurations, respectively. After removal of the episodic period, the BCO contribution becomes negligible, and the underlying source-apportionment structure is dominated by local anthropogenic contributions. Transport (TRA) is the largest contributor under EDGAR (49%), whereas energy use in buildings (ENB) is the largest contributor under INEMA (31%), followed by transport (22%) and waste (14%).

CMAQ-DDM/HDDM (Base Model)

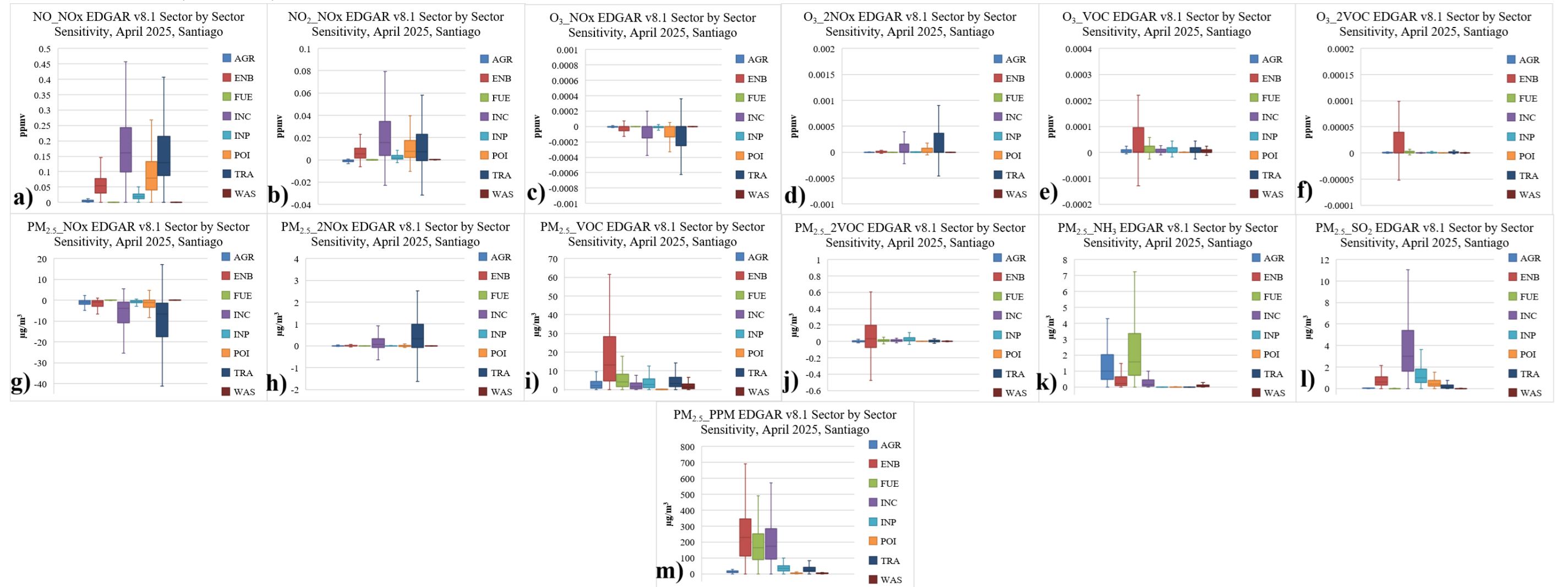


Figure S9. Sector-resolved first- and second-order sensitivities from the base EDGAR v8.1 CMAQ-DDM/HDDM simulation during April 2025. The figure displays a grid of box plots, each showing the distribution of first-order (and selected second-order) sensitivity values for a specific pollutant–precursor pair across emission sectors. Colored boxes with whiskers represent the sensitivity distributions for each sector. Sectors are agriculture (AGR), energy for buildings (ENB), fuel exploitation (FUE), industrial combustion (INC), industrial processes (INP), power industry (POI), transport (TRA), and waste (WAS). Panels show sensitivities for a) NO_NO_x, b) NO₂_NO_x, c) O₃_NO_x, d) O₃_2NO_x, e) O₃_VOC, f) O₃_2VOC, g) PM_{2.5}_NO_x, h) PM_{2.5}_2NO_x, i) PM_{2.5}_VOC, j) PM_{2.5}_2VOC, k) PM_{2.5}_NH₃, l) PM_{2.5}_SO₂, and m) PM_{2.5}_PPM. Large positive PM_{2.5}_PPM sensitivities appear for ENB, INC, and FUE, while negative PM_{2.5}_NO_x sensitivities occur for several sectors. Abbreviations: CMAQ = Community Multiscale Air Quality model; DDM-3D = Decoupled Direct Method in Three Dimensions; HDDM = Higher-order Decoupled Direct Method; EDGAR = Emissions Database for Global Atmospheric Research; AGR = Agriculture; ENB = Energy for Buildings; FUE = Fuel Exploitation; INC = Industrial Combustion; INP = Industrial Processes; POI = Power Industry; TRA = Transport; WAS = Waste; NO = nitric oxide; NO₂ = nitrogen dioxide; O₃ = ozone; PM_{2.5} = particulate matter ≤ 2.5 µm; NO_x = nitrogen oxides; VOC = volatile organic compounds; NH₃ = ammonia; SO₂ = sulfur dioxide; PPM = primary particulate matter.

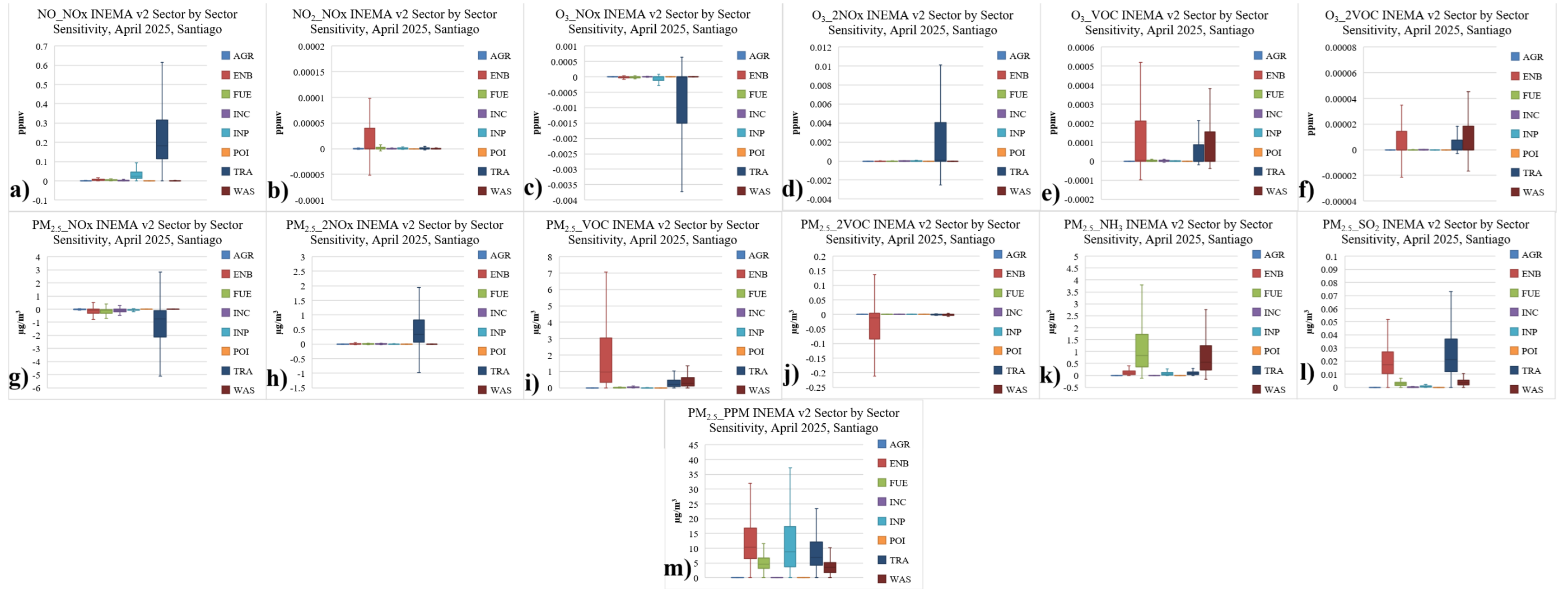


Figure S10. Sector-resolved first- and second-order sensitivities from the base INEMA v2 CMAQ-DDM/HDDM simulation during April 2025. The figure displays a grid of box plots, each showing the distribution of first-order (and selected second-order) sensitivity values for a specific pollutant–precursor pair across emission sectors. Colored boxes with whiskers represent the sensitivity distributions for each sector. Sectors are agriculture (AGR), energy for buildings (ENB), fuel exploitation (FUE), industrial combustion (INC), industrial processes (INP), power industry (POI), transport (TRA), and waste (WAS). Panels show sensitivities for a) NO_NO_x, b) NO₂_NO_x, c) O₃_NO_x, d) O₃_2NO_x, e) O₃_VOC, f) O₃_2VOC, g) PM_{2.5}_NO_x, h) PM_{2.5}_2NO_x, i) PM_{2.5}_VOC, j) PM_{2.5}_2VOC, k) PM_{2.5}_NH₃, l) PM_{2.5}_SO₂, and m) PM_{2.5}_PPM. Large positive PM_{2.5}_PPM sensitivities appear for ENB, INP, and TRA, while negative PM_{2.5}_NO_x sensitivities occur for several sectors. Abbreviations: CMAQ = Community Multiscale Air Quality model; DDM-3D = Decoupled Direct Method in Three Dimensions; HDDM = Higher-order Decoupled Direct Method; INEMA = Inventario Nacional de Emisiones Antropogénicas; AGR = Agriculture; ENB = Energy for Buildings; FUE = Fuel Exploitation; INC = Industrial Combustion; INP = Industrial Processes; POI = Power Industry; TRA = Transport; WAS = Waste; NO = nitric oxide; NO₂ = nitrogen dioxide; O₃ = ozone; PM_{2.5} = particulate matter $\leq 2.5 \mu\text{m}$; NO_x = nitrogen oxides; VOC = volatile organic compounds; NH₃ = ammonia; SO₂ = sulfur dioxide; PPM = primary particulate matter.

CMAQ-DDM/HDDM (Calibrated Emissions)

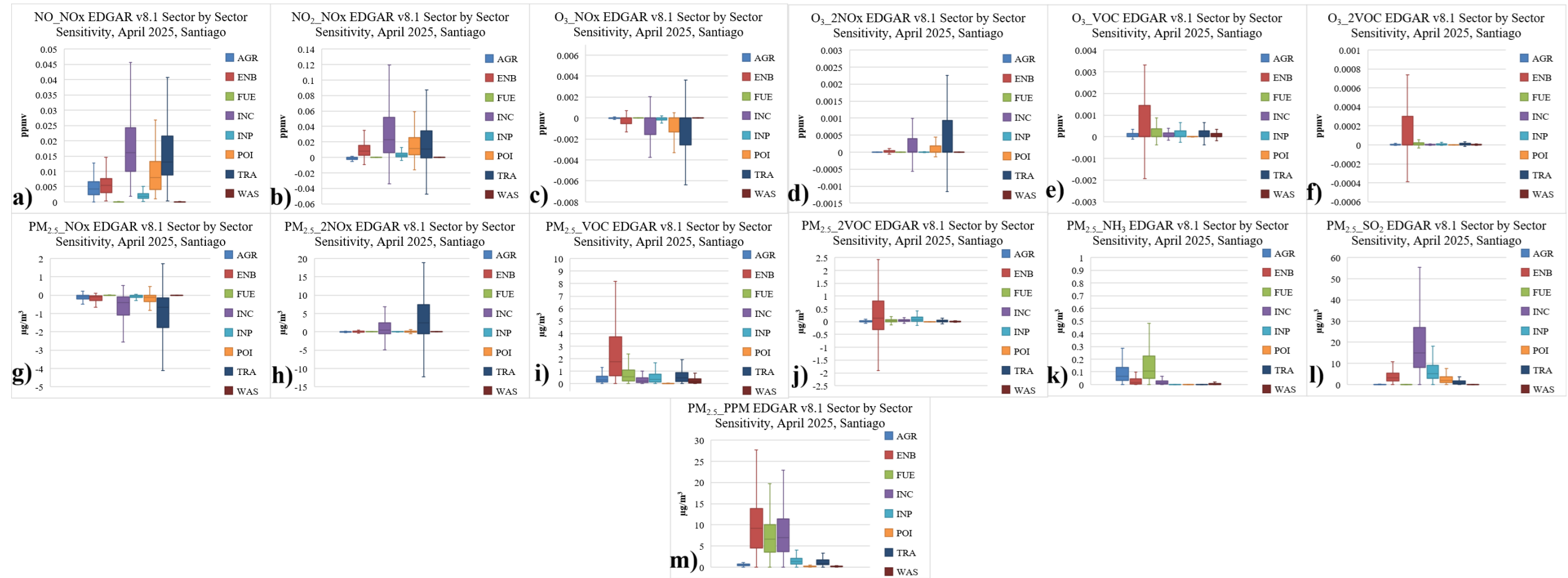


Figure S11. Sector-resolved first- and second-order sensitivities from the calibrated EDGAR v8.1 CMAQ-DDM/HDDM simulation during April 2025. The figure displays a grid of box plots, each showing the distribution of first-order (and selected second-order) sensitivity values for a specific pollutant–precursor pair across emission sectors. Colored boxes with whiskers represent the sensitivity distributions for each sector. Sectors are agriculture (AGR), energy for buildings (ENB), fuel exploitation (FUE), industrial combustion (INC), industrial processes (INP), power industry (POI), transport (TRA), and waste (WAS). Panels show sensitivities for a) NO_NO_x , b) $NO_2_NO_x$, c) $O_3_NO_x$, d) $O_3_2NO_x$, e) O_3_VOC , f) O_3_2VOC , g) $PM_{2.5_NO_x}$, h) $PM_{2.5_2NO_x}$, i) $PM_{2.5_VOC}$, j) $PM_{2.5_2VOC}$, k) $PM_{2.5_NH_3}$, l) $PM_{2.5_SO_2}$, and m) $PM_{2.5_PPM}$. Large positive $PM_{2.5_PPM}$ sensitivities appear for ENB, INC, and FUE, while negative $PM_{2.5_NO_x}$ sensitivities occur for several sectors. Abbreviations: CMAQ = Community Multiscale Air Quality model; DDM-3D = Decoupled Direct Method in Three Dimensions; HDDM = Higher-order Decoupled Direct Method; EDGAR = Emissions Database for Global Atmospheric Research; AGR = Agriculture; ENB = Energy for Buildings; FUE = Fuel Exploitation; INC = Industrial Combustion; INP = Industrial Processes; POI = Power Industry; TRA = Transport; WAS = Waste; NO = nitric oxide; NO_2 = nitrogen dioxide; O_3 = ozone; $PM_{2.5}$ = particulate matter $\leq 2.5 \mu m$; NO_x = nitrogen oxides; VOC = volatile organic compounds; NH_3 = ammonia; SO_2 = sulfur dioxide; PPM = primary particulate matter.

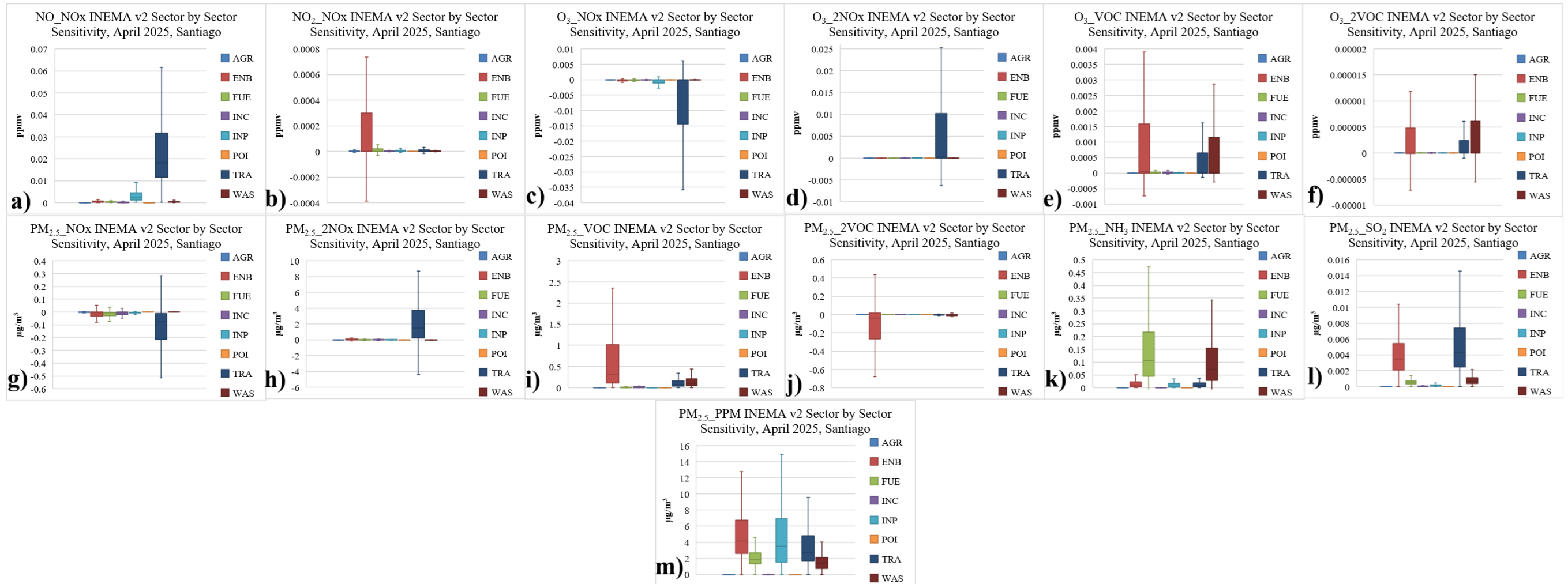


Figure S12. Sector-resolved first- and second-order sensitivities from the calibrated INEMA v2 CMAQ-DDM/HDDM simulation during April 2025. The figure displays a grid of box plots, each showing the distribution of first-order (and selected second-order) sensitivity values for a specific pollutant–precursor pair across emission sectors. Colored boxes with whiskers represent the sensitivity distributions for each sector. Sectors are agriculture (AGR), energy for buildings (ENB), fuel exploitation (FUE), industrial combustion (INC), industrial processes (INP), power industry (POI), transport (TRA), and waste (WAS). Panels show sensitivities for a) NO_NO_x, b) NO₂_NO_x, c) O₃_NO_x, d) O₃_2NO_x, e) O₃_VOC, f) O₃_2VOC, g) PM_{2.5}_NO_x, h) PM_{2.5}_2NO_x, i) PM_{2.5}_VOC, j) PM_{2.5}_2VOC, k) PM_{2.5}_NH₃, l) PM_{2.5}_SO₂, and m) PM_{2.5}_PPM. Large positive PM_{2.5}_PPM sensitivities appear for ENB, INP, and TRA, while negative PM_{2.5}_NO_x sensitivities occur for several sectors. Abbreviations: CMAQ = Community Multiscale Air Quality model; DDM-3D = Decoupled Direct Method in Three Dimensions; HDDM = Higher-order Decoupled Direct Method; INEMA = Inventario Nacional de Emisiones Antropogénicas; AGR = Agriculture; ENB = Energy for Buildings; FUE = Fuel Exploitation; INC = Industrial Combustion; INP = Industrial Processes; POI = Power Industry; TRA = Transport; WAS = Waste; NO = nitric oxide; NO₂ = nitrogen dioxide; O₃ = ozone; PM_{2.5} = particulate matter $\leq 2.5 \mu\text{m}$; NO_x = nitrogen oxides; VOC = volatile organic compounds; NH₃ = ammonia; SO₂ = sulfur dioxide; PPM = primary particulate matter.

Supplementary Materials

WRF Model Configuration

WRF-ARW v4.1.1 was run in nonhydrostatic mode (`non_hydrostatic = .true.`) using the hybrid vertical coordinate (`hybrid_opt = 2`)⁸. The hybrid coordinate is terrain-following near the surface and transitions toward isobaric coordinates aloft; `etac = 0.1` specifies the critical eta level controlling this transition, with levels satisfying $znw(k) < etac$ being isobaric. The real-data base-state temperature was set to 290 K (`base_temp = 290`). Horizontal diffusion used `diff_opt = 1`, for which diffusion gradients are evaluated along model coordinate surfaces, together with `km_opt = 4`, for which the horizontal eddy coefficient is diagnosed from horizontal deformation. Because the YSU PBL scheme was active, vertical turbulent mixing was represented by the PBL parameterization. Sixth-order numerical diffusion was disabled (`diff_6th_opt = 0`). The constant diffusion coefficients `khdif` and `kvdif` were both zero. Vertical-velocity damping was enabled (`w_damping = 1`). Upper-level damping used `damp_opt = 3`, with a damping depth of 5000 m measured downward from the model top (`zdamp = 5000`) and `dampcoef = 0.2`. Positive-definite advection was enabled for moisture and scalar fields (`moist_adv_opt = 1`, `scalar_adv_opt = 1`). Subgrid orographic effects were represented using the gravity-wave-drag option (`gwd_opt = 1`), which includes gravity-wave drag and low-level flow blocking.

Microphysics used the Thompson graupel scheme (`mp_physics = 8`) in all domains. Longwave and shortwave radiation were represented with RRTMG (`ra_lw_physics = 4`, `ra_sw_physics = 4`). The parent 9-km domain used the Grell–Freitas cumulus scheme (`cu_physics = 3`), while cumulus parameterization was disabled (`cu_physics = 0`) in the 3- and 1-km inner domains. Boundary-layer mixing used the YSU scheme (`bl_pbl_physics = 1`) with the revised MM5 Monin–Obukhov surface-layer formulation (`sf_sfclay_physics = 1`), and land-surface processes used the unified Noah land-surface model (`sf_surface_physics = 2`). Radiation was called every 9 min (`radt = 9`); `cutd = 0` and `bldt = 0` indicate that cumulus and PBL physics, respectively, were called every model time step. Cloud effects on radiation were enabled using the Xu–Randall cloud-fraction option (`icloud = 1`). The MODIS-with-lakes dataset used 21 land-use categories (`num_land_cat = 21`), urban canopy physics was disabled (`sf_urban_physics = 0`), and land-use and soil-category fields were obtained from WPS/geogrid with dominant categories recomputed by `real.exe` (`surface_input_source = 1`).

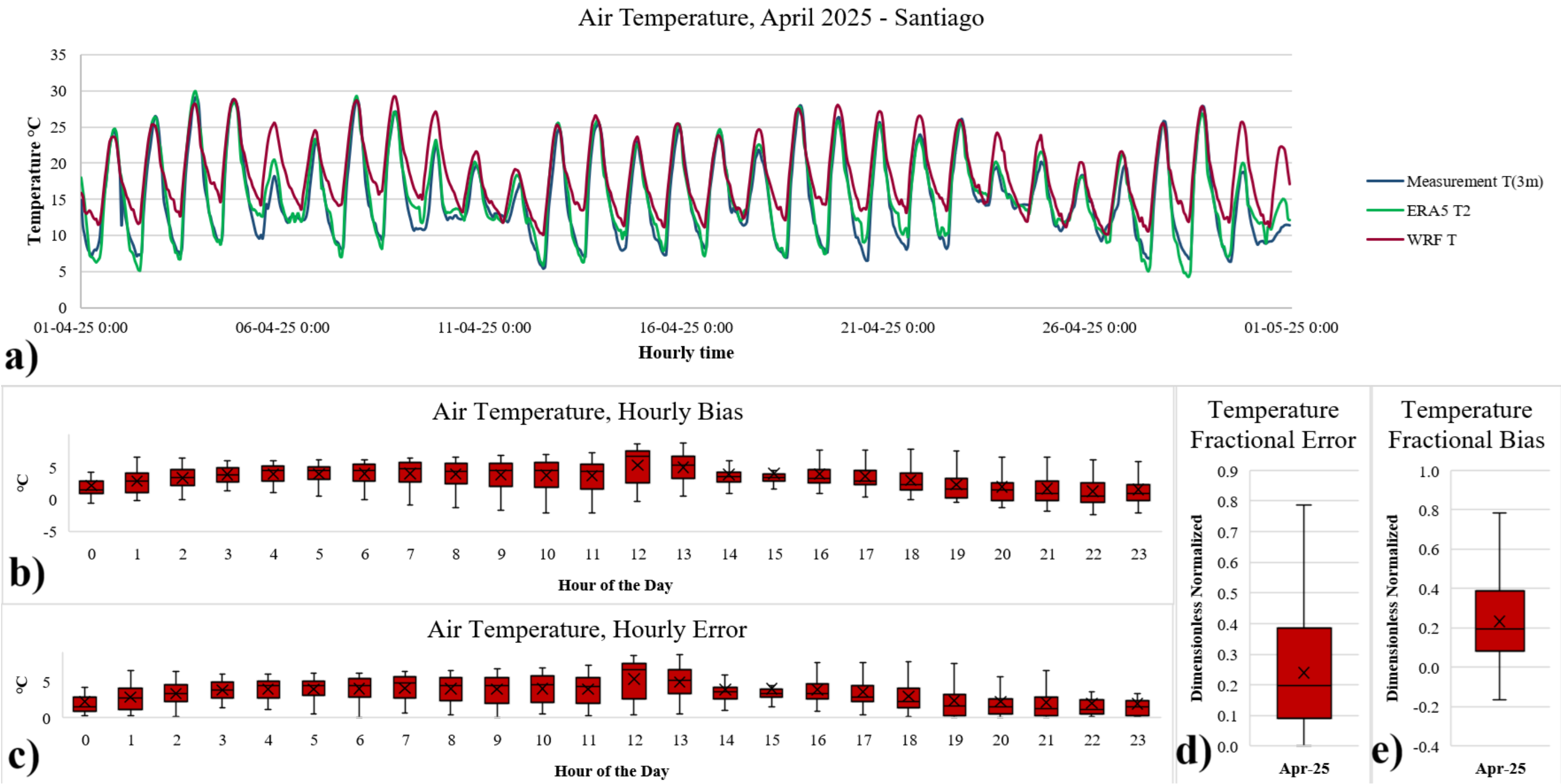


Figure S13. Performance evaluation of the WRF-simulated 2-m air temperature in central Santiago during April 2025. The figure contains five panels. Panel a shows the hourly air-temperature time-series comparison for April 2025, with the solid red line representing WRF output, the solid green line representing ERA5 reanalysis, and the solid blue line representing ground-based observations at 3 m height from O'Higgins Park station. Panel b shows box plots of hourly temperature bias for each hour of the day, and panel c shows box plots of hourly temperature error for each hour of the day; red boxes indicate the interquartile ranges and whiskers show the full ranges. Panel d shows the April 2025 temperature fractional error, and panel e shows the corresponding temperature fractional bias. WRF reproduces the observed diurnal temperature cycle reasonably well but exhibits a generally positive bias, particularly during daytime hours. Abbreviations: WRF = Weather Research and Forecasting model; ERA5 = European Centre for Medium-Range Weather Forecasts Reanalysis v5.

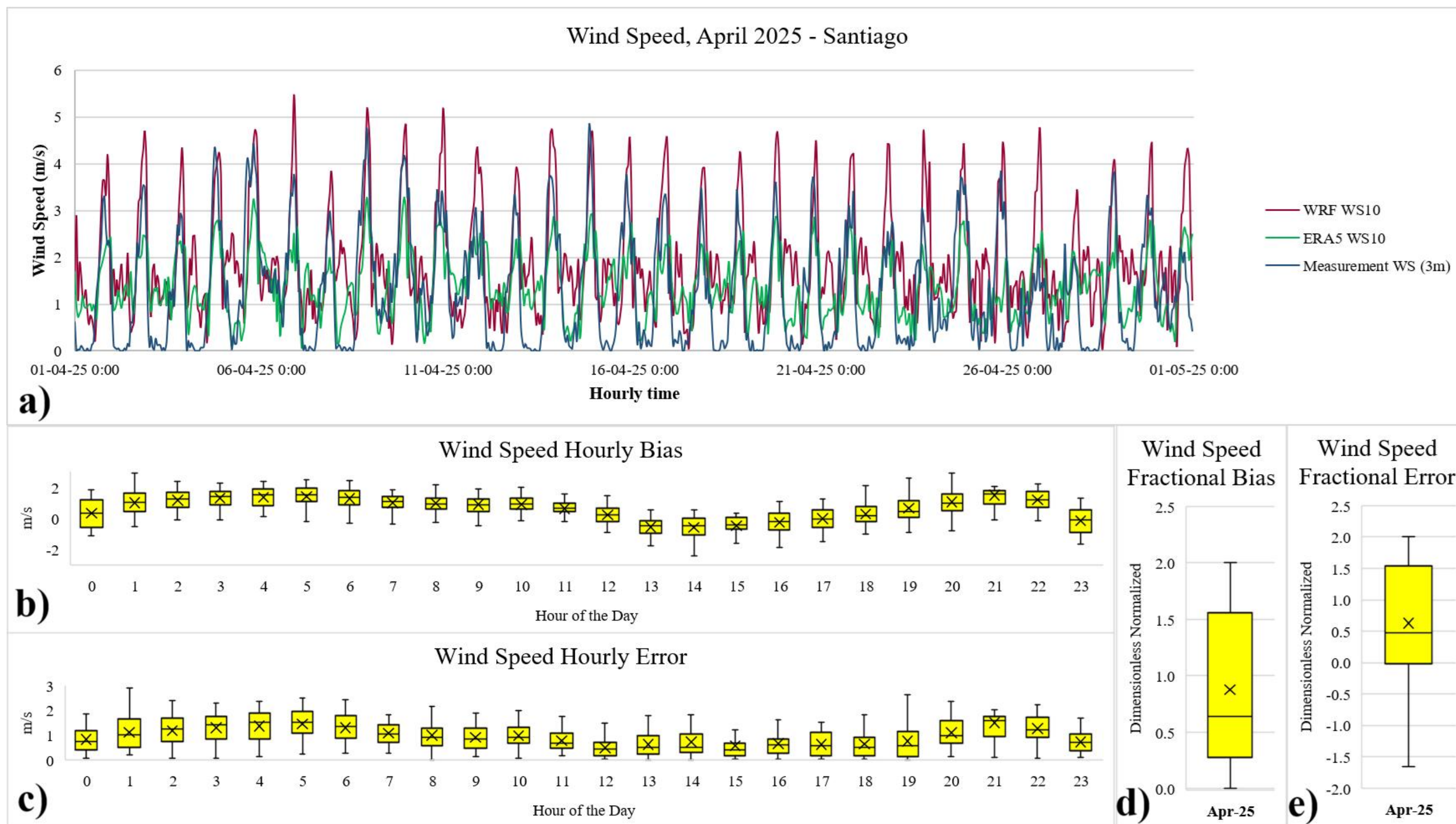


Figure S14. Performance evaluation of the WRF-simulated wind speed in central Santiago during April 2025. The figure contains five panels. Panel a shows the hourly wind-speed time-series comparison for April 2025, with the solid red line representing WRF 10-m wind speed, the solid green line representing ERA5 10-m wind speed, and the solid blue line representing ground-based wind-speed observations at 3 m height from O'Higgins Park station. Panel b shows box plots of hourly wind-speed bias for each hour of the day, and panel c shows box plots of hourly wind-speed error for each hour of the day; yellow boxes indicate the interquartile ranges and whiskers show the full ranges. Panel d shows the April 2025 wind-speed fractional bias, and panel e shows the corresponding wind-speed fractional error. Abbreviations: WRF = Weather Research and Forecasting model; ERA5 = European Centre for Medium-Range Weather Forecasts Reanalysis v5.

The prevailing wind patterns during the simulation period are illustrated in Figure S15. Observations show dominant winds from the southwest (approximately 210°–240°), with frequent speeds between 2–5 m/s. The WRF simulation successfully reproduced the dominant southwest direction and the overall wind speed distribution, although it slightly overestimated the frequency of stronger winds (>4 m/s). In contrast, ERA5 reanalysis exhibited a more northerly component and weaker overall speeds compared to both observations and the WRF model. These differences highlight the added value of dynamical downscaling with WRF for capturing local topographic influences on wind fields in the Santiago basin. Figure S16 of the Supplementary Information shows time series comparison of WRF results with ERA5 reanalysis and ground-based observations for air pressure and relative humidity.

Wind Roses - April 2025 - Santiago - 30° & 1m/s Bins

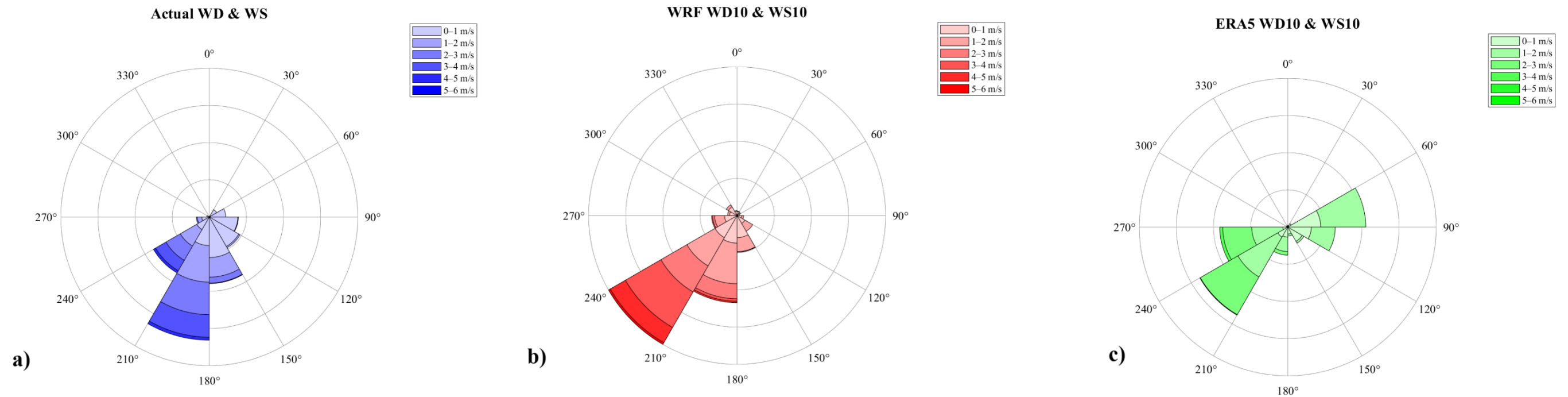


Figure S15. Wind roses for April 2025 in central Santiago, Chile, showing wind direction (30° bins) and wind speed (1 m/s bins). The figure presents three circular wind rose plots side by side. Panel a) shows observed wind conditions from the ground-based station at O'Higgins Park; panel b) shows WRF-simulated 10-m winds; and panel c) shows ERA5 reanalysis 10-m winds. Each rose displays the frequency of wind directions (calculated as $\text{mod}(180 + 180/\pi \times \text{atan2}(u, v), 360)$) and speeds (calculated as $\sqrt{u^2 + v^2}$), with color shading indicating wind speed bins from calm (light) to strong (dark). Observations and WRF both show dominant southwesterly winds (210°–240°), while ERA5 exhibits a more northerly component and weaker speeds. WRF captures the dominant southwesterly flow and diurnal variability better than ERA5. Abbreviations: WRF = Weather Research and Forecasting model; ERA5 = European Centre for Medium-Range Weather Forecasts Reanalysis v5; u = east-west wind component; v = north-south wind component.

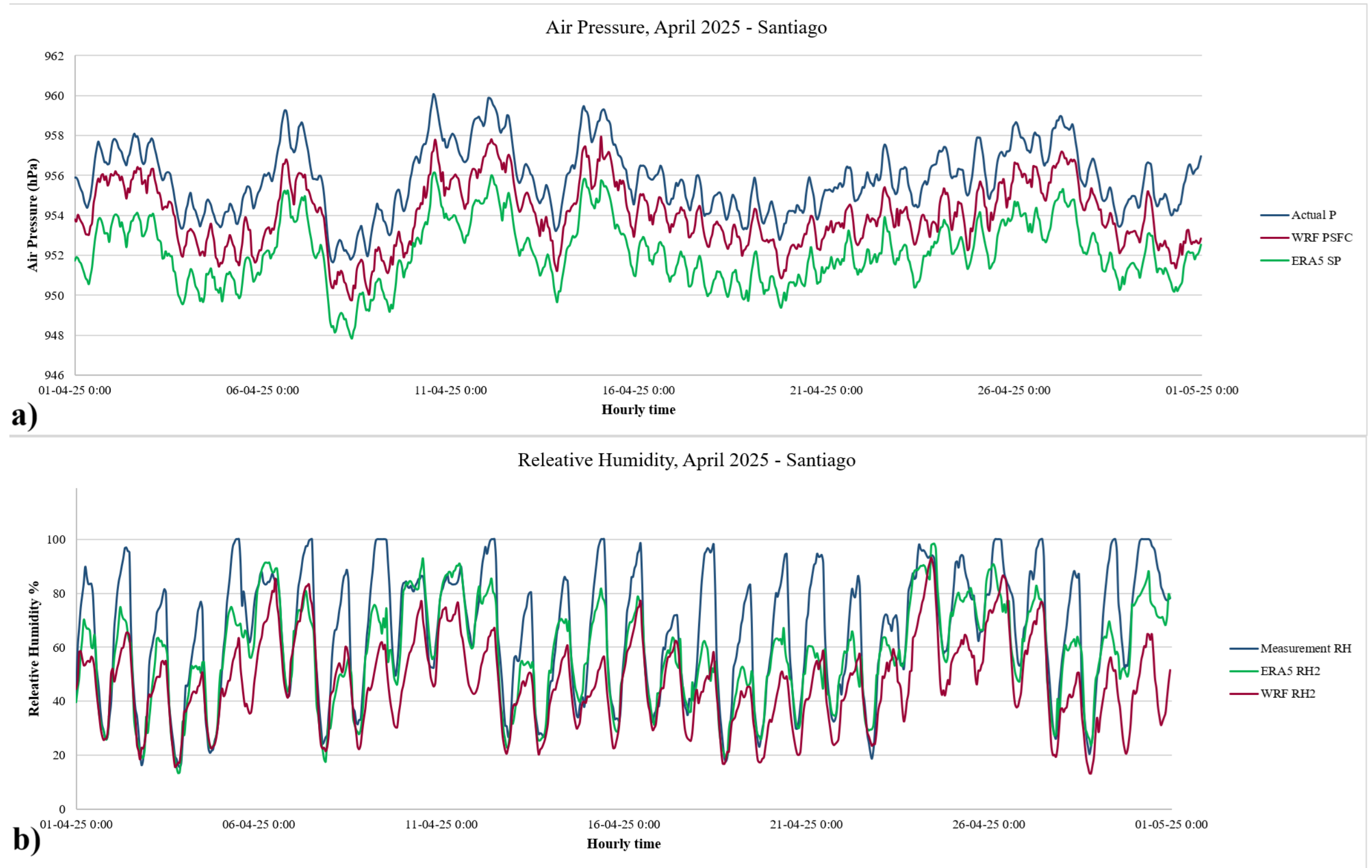


Figure S16. Performance evaluation of the WRF-simulated air pressure and relative humidity in central Santiago during April 2025. The figure shows time series comparisons of WRF output (red), ERA5 reanalysis (green), and ground-based observations (blue) at the O'Higgins Park monitoring station (SINCA Station D14) for (a) air pressure and (b) relative humidity. Relative humidity (RH, %) is calculated from specific humidity (q , $\text{kg}\cdot\text{kg}^{-1}$) and air temperature (T , $^{\circ}\text{C}$) using the following approximation: $RH = 100 (e / e_s) = 100 \times [qp / (0.622 + 0.378q)] / [611.2 \times \exp(17.67T / (T + 243.5))]$, where e is the actual vapor pressure and e_s is the saturation vapor pressure and p is pressure.

SMOKE Model CMAQ-Ready Emissions

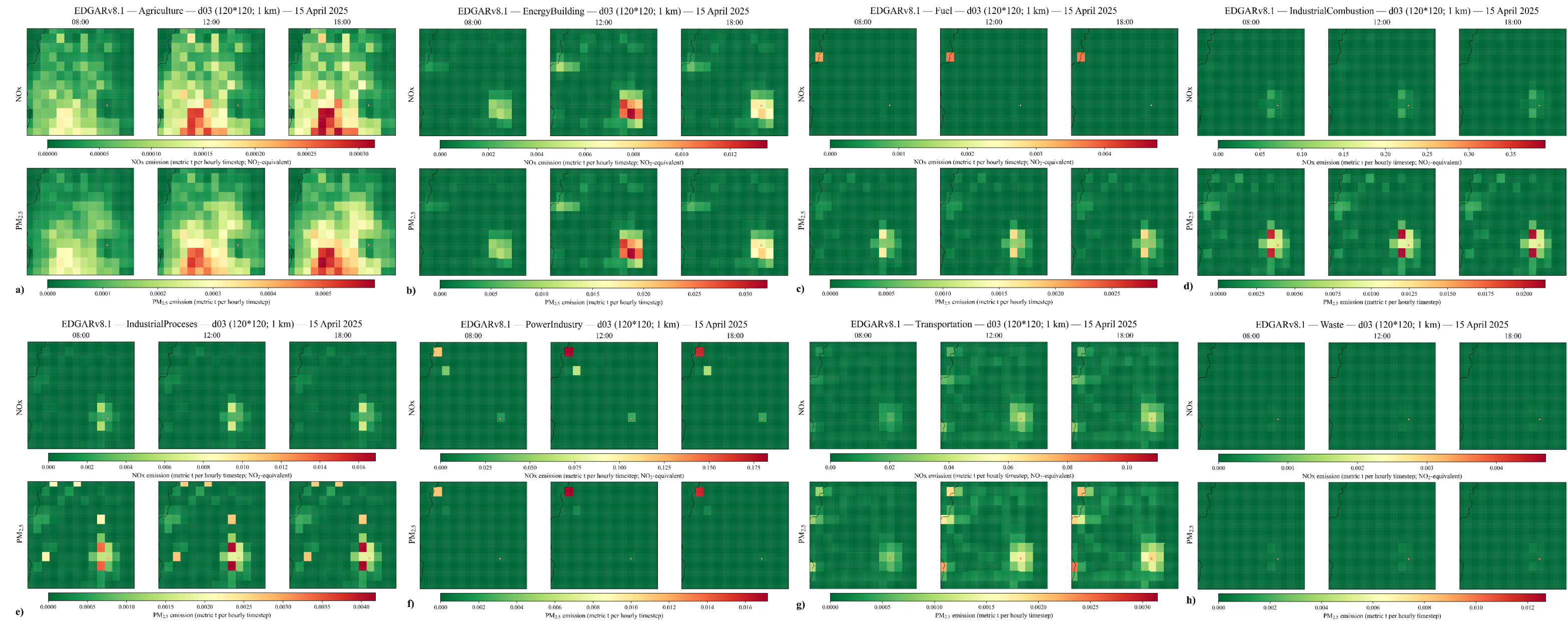


Figure S17. Spatial distribution of CMAQ-ready EDGAR v8.1 (2022) NO_x and $\text{PM}_{2.5}$ emissions over the Santiago 1-km resolution ($120 \text{ km} \times 120 \text{ km}$) domain on 15 April 2025. Panels show the eight harmonized anthropogenic source sectors: (a) agriculture (AGR), (b) energy for buildings (ENB), (c) fuel exploitation (FUE), (d) industrial combustion (INC), (e) industrial processes (INP), (f) power industry (POI), (g) transport (TRA), and (h) waste (WAS). Within each panel, the upper row shows NO_x emissions and the lower row shows $\text{PM}_{2.5}$ emissions at 08:00, 12:00, and 18:00 local time, illustrating sector-specific spatial patterns and their diurnal variation across the 120×120 -cell domain. Color scales are sector- and pollutant-specific and therefore should not be compared directly between panels. The white circle with a red outline marks the location of the Parque O'Higgins monitoring station. Abbreviations: CMAQ = Community Multiscale Air Quality model; EDGAR = Emissions Database for Global Atmospheric Research; NO_x = nitrogen oxides; $\text{PM}_{2.5}$ = particulate matter with aerodynamic diameter $\leq 2.5 \text{ }\mu\text{m}$; AGR = agriculture; ENB = energy for buildings; FUE = fuel exploitation; INC = industrial combustion; INP = industrial processes; POI = power industry; TRA = transport; WAS = waste.

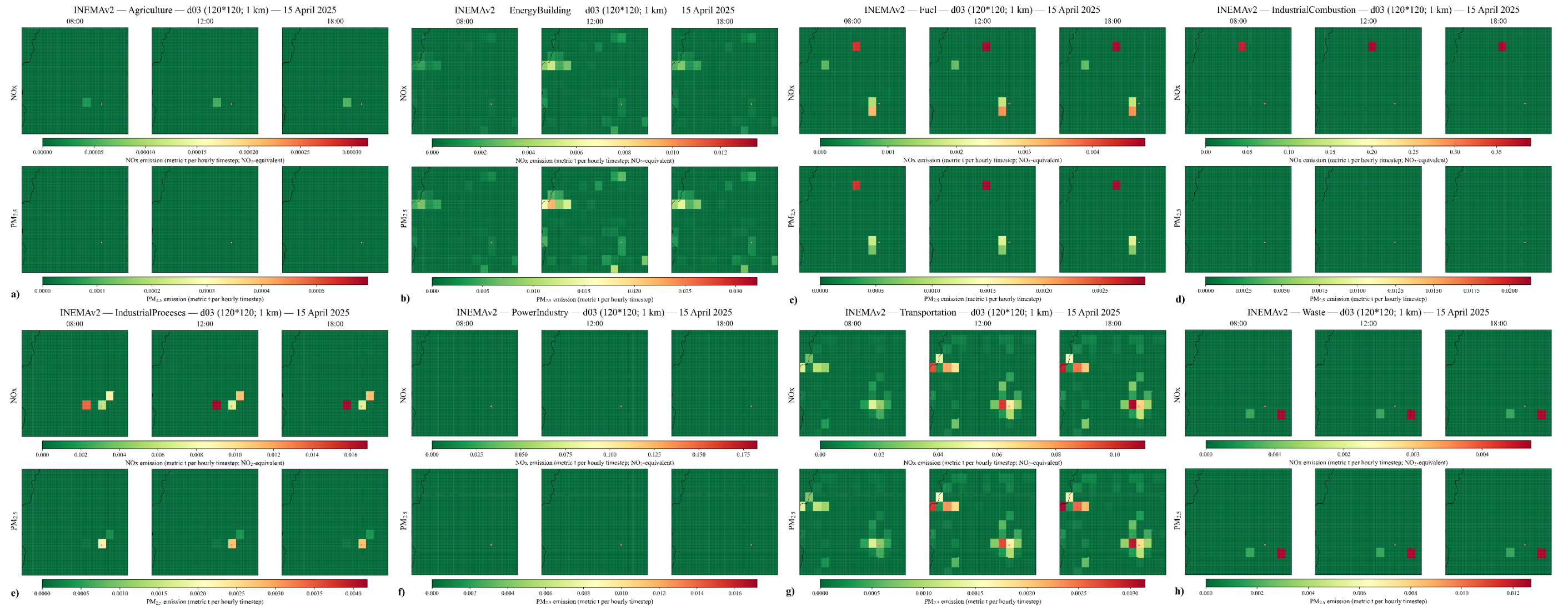


Figure S18. Spatial distribution of CMAQ-ready INEMA v2 (2020) NO_x and PM_{2.5} emissions over the Santiago 1-km resolution (120 km × 120 km) domain on 15 April 2025. Panels show the eight harmonized anthropogenic source sectors: (a) agriculture (AGR), (b) energy for buildings (ENB), (c) fuel exploitation (FUE), (d) industrial combustion (INC), (e) industrial processes (INP), (f) power industry (POI), (g) transport (TRA), and (h) waste (WAS). Within each panel, the upper row shows NO_x emissions and the lower row shows PM_{2.5} emissions at 08:00, 12:00, and 18:00 local time, illustrating sector-specific spatial patterns and their diurnal variation across the 120 × 120-cell domain. Color scales are sector- and pollutant-specific and therefore should not be compared directly between panels. The white circle with a red outline marks the location of the Parque O'Higgins monitoring station. Abbreviations: CMAQ = Community Multiscale Air Quality model; INEMA = Inventario Nacional de Emisiones Antropogénicas; NO_x = nitrogen oxides; PM_{2.5} = particulate matter with aerodynamic diameter ≤ 2.5 μm; AGR = agriculture; ENB = energy for buildings; FUE = fuel exploitation; INC = industrial combustion; INP = industrial processes; POI = power industry; TRA = transport; WAS = waste.

SMOKE Model NMVOC Speciation

Table S2. List of the 25 GEIA NMVOC species groups and their representative molecular formulas used for chemical speciation of the EDGAR v8.1 inventory.

GEIA ID	GEIA Group	Molecular Formula
VOC1	Alkanols (alcohols)	C _n H _{2n+1} OH
VOC2	Ethane	C ₂ H ₆
VOC3	Propane	C ₃ H ₈
VOC4	Butanes	C ₄ H ₁₀
VOC5	Pentanes	C ₅ H ₁₂
VOC6	Hexanes and higher alkanes	C _n H _{2n+2} (n ≥ 6)
VOC7	Ethene (ethylene)	C ₂ H ₄
VOC8	Propene	C ₃ H ₆
VOC9	Ethyne (acetylene)	C ₂ H ₂
VOC10	Isoprenes	C ₅ H ₈
VOC11	Monoterpenes	C ₁₀ H ₁₆
VOC12	Other alk(adi)enes/alkynes (olefines)	C _n H _{2n-2}
VOC13	Benzene (benzol)	C ₆ H ₆
VOC14	Methylbenzene (toluene)	C ₇ H ₈
VOC15	Dimethylbenzenes (xylenes)	C ₆ H ₄ (CH ₃) ₂
VOC16	Trimethylbenzenes	C ₆ H ₃ (CH ₃) ₃
VOC17	Other aromatics	C _n H _{2n-6}
VOC18	Esters	R-C(=O)O-R'
VOC19	Ethers (alkoxy alkanes)	R-O-R'
VOC20	Chlorinated hydrocarbons	CH ₃ Cl
VOC21	Methanal (formaldehyde)	CH ₂ O
VOC22	Other alkanals (aldehydes)	R-CHO
VOC23	Alkanones (ketones)	R-C(=O)-R'
VOC24	Acids (alkanoic)	R-C _n H _n COOH
VOC25	Other NMVOC (HCFCs, nitriles, etc.)	NA
NOTES: R & R' denote functional groups. Where general formulae are not appropriate, the simplest molecular formula representing the group is provided. NA = Not Available.		

To ensure compatibility with the CB06R5 chemical mechanism (*cb6r5_ae7_aq*) employed in our CMAQ modeling (see the main manuscript, section Methods; CMAQ Processing), the original EDGAR NMVOC groups were mapped into CMAQ-ready mechanism species. Also, species-dependent speciation factors were derived through the emission-weighted allocation methodology based on the total sector-specific contribution of each NMVOC group (within the parent modeling domain) to preserve the relative chemical composition of the original EDGAR speciation profiles. The final speciation profiles were subsequently integrated into the *spcmat* processing stage within SMOKE.

SMOKE Model Flow Chart

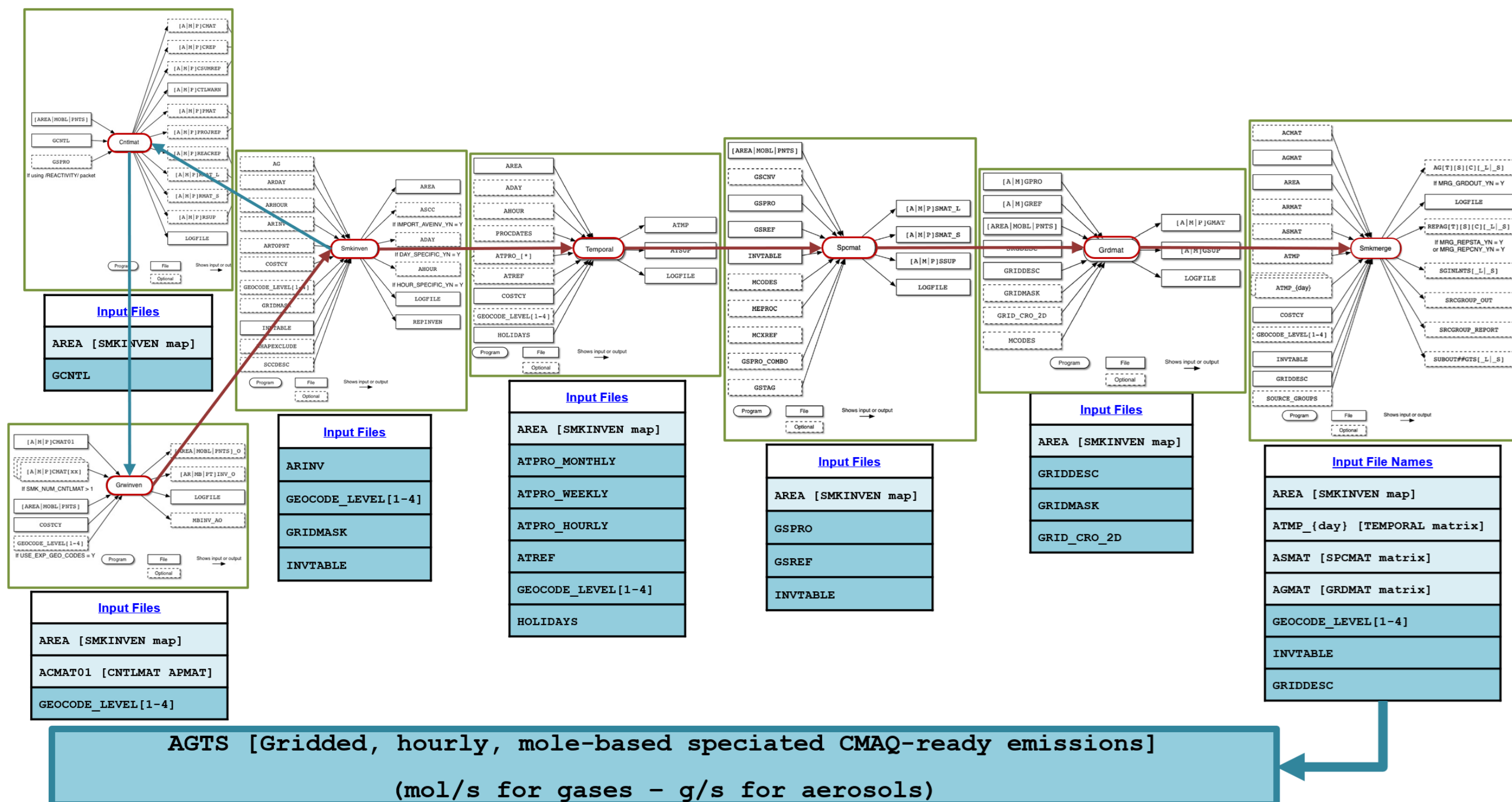


Figure S19. The full workflow of the SMOKE program followed in this study⁹. The diagram shows the Sparse Matrix Operator Kernel Emissions (SMOKE) workflow used to convert annual/monthly, county- or source-level inventories into gridded, hourly, chemically speciated emissions for CMAQ. Green boxes are SMOKE programs; teal tables list the controlling input files at each step. Optional control and growth matrices are built first (Cntlmat, Grwinven) from the area inventory and control packets (GCNTL). Smkinven imports the raw AREA inventory, geocodes and source cross-references (INVTABLE, GEOCODE_LEVEL [1-4], GRIDMASK) and writes the internal SMOKE inventory map. Temporal applies month-, week- and hour-specific profiles (ATPRO_MONTHLY, ATPRO_WEEKLY, ATPRO_HOURLY, ATREF, HOLIDAYS) to produce the temporal matrix (ATMP). Spcmat maps inventory pollutants onto the CMAQ chemical mechanism using speciation profiles and cross-references (GSPRO, GSREF) and writes the speciation matrix (ASMAT). Grdmat allocates county- or source-level mass onto the model grid with the grid description, mask and 2-D grid-cross file (GRIDDESC, GRIDMASK, GRID_CRO_2D) and writes the gridding matrix (AGMAT). Smkmerge multiplies the inventory by the temporal, speciation and gridding matrices and merges source categories to produce AGTS: gridded, hourly, mole-based speciated emissions ready for CMAQ ($\text{mol}\cdot\text{s}^{-1}$ for gases; $\text{g}\cdot\text{s}^{-1}$ for aerosols). Red arrows indicate mandatory processing pathways required in the workflow, whereas blue arrows indicate optional or non-mandatory processing pathways used only when the corresponding processing step or input is required.

CMAQ Chemistry-Transport Model (CCTM)

Air quality simulations were performed using the CMAQ model version 5.5.0.3. The model was compiled with the Carbon Bond 6 revision 5 (CB6r5) chemical mechanism coupled with the AERO7 aerosol module and aqueous-phase chemistry and employed the M3DRY dry deposition scheme. The computational domain features 40 vertical layers extending from the surface to the upper troposphere. Horizontal domain decomposition used a 4×11 processor grid (NPCOL = 4, NPROW = 11), resulting in 44 MPI processes for efficient parallel execution. The base-case simulation covered the full month of April 2025 and was integrated for 720 hours (NSTEPS = 7200000) with an hourly output timestep (TSTEP = 010000). Gridded anthropogenic emissions were supplied as eight sector-specific NetCDF files (N_EMIS_GR = 8) prepared at the native resolutions of the modeling domains (9 km, 3 km, and 1 km). No inline biogenic emissions (BEIS or MEGAN), lightning NO_x, or windblown dust emissions were activated; all emissions were pre-processed externally and provided as gridded inputs.

Key science options selected for the Santiago case study included M3DRY dry deposition, with gravitational sedimentation of aerosols enabled (CTM_GRAV_SETL = Y) and surface HONO interactions activated (CTM_SFC_HONO = Y). Ocean halogen chemistry was disabled (CTM_OCEAN_CHEM = N), as were ammonia bidirectional flux (CTM_ABFLUX = N) and potential vorticity-based stratospheric O₃ transport (CTM_PVO3 = N). Photolysis was calculated inline using OMI-derived total-column ozone input together with the standard PHOT_OPTICS.dat file, while aerosol optics were treated with the case-approximation Mie theory method (AEROSOL_OPTICS = 3). These settings were selected based on the availability of input datasets. The executable (CCTM_v55.exe) was launched using mpirun -np 44 on a Linux cluster with Intel oneAPI compilers and the Intel MPI library.

Meteorological inputs generated by WRF v4.1.1 were processed through the MCIP to produce CMAQ-compatible meteorological fields, while hourly anthropogenic emissions processed through SMOKE v5.2.1 were supplied as gridded chemically speciated emissions inputs. The CMAQ configuration incorporated online transport and chemistry calculations together with dry deposition, cloud processes, aerosol thermodynamics, and gas-particle partitioning mechanisms required for urban and regional air quality simulations.

The primary CMAQ simulations were subsequently performed using the CCTM module together with the *M3DRY* dry deposition scheme for surface-atmosphere exchange processes. Beyond the base air quality simulations, advanced CMAQ extensions including ISAM and sector-by-sector DDM-3D/HDDM-3D were activated to investigate pollutant source contributions and first- and second-order precursor sensitivities. CMAQ-ISAM was employed to quantify the relative contributions of emission sectors to NO_x, NO_z, NH₃, NO₃, O₃, PM, SO₄, VOC, SO₂, OA, EC, and NVPOA concentrations, while CMAQ-DDM-3D was used to evaluate pollutant sensitivity responses to precursor emission perturbations under different atmospheric chemical regimes.

References

1. Crippa, M., et al. EDGAR_emissions_urban_areas. European Commission, Joint Research Centre (JRC) [Dataset] (2021).
2. Sánchez, P. L. F., Manzano, C. A., Leiva-Guzmán, M. A., Canales, M. & Araya, R. T. Urban atmospheric particle size distribution in Santiago, Chile. *Atmos. Pollut. Res.* 12, 101201 (2021).
3. Préndez, M., Nova, P., Romero, H., Mendes, F. & Fuentealba, R. Representativeness of the particulate matter pollution assessed by an official monitoring station of air quality in Santiago, Chile: projection to human health. *Environ. Geochem. Health* 45, 2985–3001 (2023).
4. Gramsch, E. V. et al. Relationship between UV energy and formation of secondary particles in Santiago de Chile. *Aerosol Air Qual. Res.* 20, 2117–2127 (2020).
5. Gramsch, E., Gidhagen, L., Wahlin, P., Oyola, P. & Moreno, F. Predominance of soot-mode ultrafine particles in Santiago de Chile: possible sources. *Atmos. Environ.* 43, 2260–2267 (2009).
6. Chicureo Hoy. Major fire reported in Paine. <https://www.chicureohoy.cl/actualidad/incendio-de-gran-magnitud-se-registra-en-planta-fruticola-en-paine/> (2025; accessed August 2026).
7. SoyChile. Forest fire on Paidahuén Hill. <https://www.soychile.cl/Quillota/Sociedad/2025/04/08/901484/san-esteban-incendio-forestal-cerropaidahuen.html> (2025; accessed August 2026).
8. Skamarock, W. C. et al. A Description of the Advanced Research WRF Model Version 4. NCAR Tech. Note NCAR/TN-556+STR (2019).
9. Baek, B. H. et al. CEMPD/SMOKE: SMOKE v5.0 public release (June 2023). Zenodo <https://doi.org/10.5281/zenodo.8034267> (2023).