Science of the Total Environment 906 (2024) 167350

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Air pollution analysis in Northwestern South America: A new Lagrangian framework

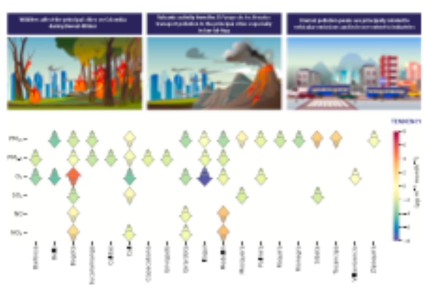
Alejandro Casallas [a,b,c,*], Ailin Cabrera [c], Marco-Andrés Guevara-Luna [d,e], Adrian Tompkins [a], Yuri González [f], Juan Aranda [g], Luis Carlos Belalcazar [h], Caroline Mogollon-Sotelo [i], Nathalia Celis [j], Ellie Lopez-Barrera [c], Carlos A. Peña-Rincon [c], Camilo Ferro [k]

[a] Earth System Physics, Abdus Salam International Centre for Theoretical Physics – ICTP, 34151 Trieste, Italy
[b] Department of Mathematics and Geoscience, University of Trieste, 34128 Trieste, Italy
[c] Escuela de Ciencias Exactas e Ingeniería, Universidad Sergio Arboleda, 11011 Bogotá, Colombia
[d] LIVE-Laboratoire Image Ville Environnement, Université de Strasbourg, 3 rue de l'Argonne, Strasbourg, France
[e] Conservación, Bioprospección y Desarrollo Sostenible (COBIDES), Universidad Nacional Abierta y a Distancia, Escuela de Ciencias Agrícolas, Pecuarias y del Medio Ambiente (ECAPMA), Bogotá, Colombia
[f] Facultad de Ingeniería y Ciencias Básicas, Fundación Universitaria Los Libertadores, 111221 Bogotá, Colombia
[g] Facultad de Ingeniería, Universidad de La Sabana, Campus del Puente del Común, Km 7 Autopista Norte de Bogotá, 250001 Chía, Cundinamarca, Colombia
[h] Departamento de Ingeniería Química y Ambiental, Universidad Nacional de Colombia, Bogotá, Colombia
[i] Department of Environmental Sciences and Policy, Central European University, Vienna, Austria
[j] Department of Civil, Environmental, and Architectural Engineering, University of Padova, Padova, Italy
[k] Departamento de Ingeniería, Aqualogs SAS, 11011 Bogotá, Colombia

HIGHLIGHTS

- PM is mostly decreasing. NO-NO_2 are increasing, their ratio controls the O_3 tendency.
- A novel Lagrangian Framework designed to analyze Long-Range Transport of pollution.
- Fires, wind speed, Boundary Layer Height, and traffic explain PM and NOx seasonality.
- O_3 seasonality is explained by NOx reactions, fires, and meteorology.
- SO_2 high pollution episodes are produced by plumes from the Nevado del Ruiz volcano.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Philip K. Hopke

Keywords:
Long-range transport
Forest fires
Volcanic emissions

ABSTRACT

This study examines the spatiotemporal variations of $PM_{2.5}$, PM_{10}, SO_2, O_3, NO, and NO_2 concentrations in Northwestern South America (NWSA). We assess the efficacy of existing policies, identify underlying phenomena, and highlight areas for further research. Significant findings have emerged by analyzing reanalysis and in-situ data, employing the WRF-Chem model, and utilizing a new Lagrangian framework designed to overcome some drawbacks common to analysis of pollution Long-Range Transport. Wildfires in the first half of the year and

* Corresponding author at: Earth System Physics, Abdus Salam International Centre for Theoretical Physics – ICTP, 34151 Trieste, Italy.
E-mail address: acasallas@ictp.it (A. Casallas).

https://doi.org/10.1016/j.scitotenv.2023.167350
Received 25 August 2023; Received in revised form 19 September 2023; Accepted 23 September 2023
Available online 27 September 2023

WRF-Chem model
CAMS/ERA5 reanalysis

volcanic activity (for SO_2) in July–August contribute to over 90 % of the pollutant's advection, leading to high pollution levels in urban areas. SO_2 volcanic emissions contribute to secondary PM, explaining the peak in PM concentrations in *Cali* in July. In the second half of the year, pollutant behavior varies based on factors such as city characteristics, vehicular-volume, air temperature, wind speed, and boundary layer height, and O_3 is influenced by solar radiation and the NO/NO_2 ratio. Diurnal variations of PM and NO_x correlate with vehicular density, SO_2 with industrial activity, and O_3 depends on solar radiation. Trend analysis reveals decreasing PM_{10} levels except in three *Cundinamarca* cities and *Cali* suggesting the need to implement/evaluate control plans in those locations. Although data is limited, NO and NO_2 levels show an increasing trend due to the rising number of vehicles. SO_2 levels are decreasing, except in *Cali*, potentially influenced by the nearby industrial and polluted city of *Yumbo*. O_3 displays a downward trend in most cities, except *Bogotá*, due to the NO/NO_2 ratio favoring O_3 increase. These findings provide a starting point for further research to deepen our understanding of NWSA air pollution. Such investigations are essential before modifying existing policies or enacting new ones. Collaborative efforts at the international, regional, and inter-city levels are crucial for effective air quality management.

1. Introduction

According to the World Health Organization – WHO (WHO, 2021), air pollution causes approximately 7 million deaths globally every year. The emissions from both human activities and natural sources, along with the complex dynamics of the atmosphere, impact human health worldwide (Lelieveld et al., 2015). Additionally, the intricate relationship between emissions and atmospheric dynamics, including wind patterns, temperature inversions, and chemical reactions, shapes the spatial and temporal distribution of pollutants, leading to regional and global challenges in air pollution (Seinfeld and Pandis, 2016). It is, therefore, crucial to understand the sources, transport, transformation, and fate of pollutants at different spatial dimensions (Zheng et al., 2022). Here we focus on North Western South America (NWSA), one of the most mountainous areas regarding climate, geography, society, and air quality of the world. This region lacks sufficient surface measurement systems for comprehensive air quality evaluation. NWSA is a tropical region with complex topography, a feature that drives a significant part of the atmospheric dynamics (Hernandez-Deckers, 2021) because of three mountain branches named: *cordilleras*. These *cordilleras* have a southwest-northeast orientation that passes through *Ecuador*, *Colombia* and ends at the Colombian-Venezuelan Border (CVB). Such complex orography makes it challenging to implement a well-suited in-situ observation network to measure air quality and meteorology [see Supplementary Material (hereafter SM) Fig. SM1].

A well-suited observation network of air quality monitoring is essential to design, enact and evaluate policies that aim to reduce the risk associated with air pollution, as recommended by the WHO (2021), which highlight maximum concentration levels and times of exposition for the different atmospheric pollutants to prevent adverse impacts on human health. This means that if the monitoring network is not well-suited as in NWSA, it is imperative to use other data sources as, for example, the Copernicus Atmosphere Monitoring Service (CAMS) reanalysis dataset (Inness et al., 2019), satellite products (e.g., MODIS – NASA, 2023), and the GEOS-Chem reanalysis (Bey et al., 2001). CAMS has already been validated for $PM_{2.5}$ in the region by Casallas et al. (2022 - hereafter CCG22) and IDEAM (2022) for other pollutants. Many studies have evaluated it in other regions (e.g., Guevara et al., 2021; Kuenan et al., 2022), showing acceptable results. Aerosol Optical Depth (AOD) from MODIS has been analyzed by the region by Guevara-Luna et al. (2018), and GEOS-Chem was first evaluated in NWSA by Rodriguez-Villamizar et al. (2022) with reasonable results, meaning that all these datasets could be used as tools to study pollution hotspots in NWSA.

CCG22 evaluated $PM_{2.5}$ trends in the NWSA, showing that CAMS can be used to determine the spatiotemporal distribution of the pollutant in NWSA. It also reported the $PM_{2.5}$ evolution in several cities, accounting for the role of policies (e.g., Franco et al., 2019; Gómez et al., 2020; Agudelo et al., 2021). The trends show that the pollutant is decreasing in many cities except in *Cali*, as reported by Mateus-Fontecha et al. (2022) and consistent with Peralta et al. (2023) results. CCG22 also shows that the diurnal variations of the pollutant depend on the number of vehicles (e.g., Rojas, 2004; Castillo-Camacho et al., 2020; Rodríguez-Gómez et al., 2022; Rojas et al., 2023), the wind speed and the Boundary Layer Height (BLH). The seasonal behavior depends on the Long-Range Transport (LRT) from wildfires (Mendez-Espinosa et al., 2020a) in the first half of the year, as identified using the Hybrid Single-Particle Lagrangian Integrated Trajectory model (HYSPLIT), which serves to approximate a relationship between wildfires and high-pollution events. During the second half of the year, $PM_{2.5}$ presented an increase in October–November, but their work did not describe the reasons for it. Here the analysis is extended to a larger set of pollutants (i.e., including gaseous pollutants) governed by complex dynamics and interactions, e.g., atmospheric chemistry, intermittency, and seasonality of the sources. We also include a new LRT analysis by means of a Lagrangian framework which is implemented using data from CAMS and from the Weather Forecast Research (WRF) model coupled with a chemistry module (WRF-Chem). The proposed framework allows the tracking and analysis of the evolution of plumes (Van Eaton et al., 2016). It also considers crucial aspects, including the spatial extent, concentration, and trends of pollutants within cities. As a novelty, the framework is able to determine and quantify the percentage of the pollution advected to the cities from regional phenomena.

2. Dataset and methods

2.1. Datasets

2.1.1. In-situ measurements and data quality control

This study employs hourly surface measurements of O_3, PM_{10}, $PM_{2.5}$, NO, NO_2, and SO2 concentrations from the Colombian Air Quality Information Subsystem Database (SISAIRE, 2023) spanning 2012-01-31 to 2022-12-31 across thirty-two Colombian cities. The data is utilized for two main objectives: (i) conducting a temporal trend analysis of pollutants based on their spatial distribution, and (ii) validating the CAMS reanalysis for NWSA. Before data usage, Chauvenet's criterion, following CCG22 and influenced by Lishu et al. (2015), is employed to mitigate data uncertainty. The criterion, typically applied to parametric distributions, is adapted here to identify data three standard deviations above the mean. Subsequently, these marked values are cross-referenced with data from the three nearest surface measuring stations to ascertain if values result from regional patterns or possible sensor outliers.

With validated data, cities' hourly mean concentrations are calculated using available background, urban, and rural monitoring stations (not traffic or industrial), resulting in one time series per city. Furthermore, only cities with a minimum of 75 % of annual data representation in at least the last four years are chosen for analysis. Nine cities yield suitable data for O_3, fourteen for PM_{10}, nine for $PM_{2.5}$, three for NO, and four for NO_2 and SO_2. The limited number of cities for NO, NO_2, and SO_2 underscores the necessity for expanded air quality monitoring, especially as increasing NO and NO_2 trends emerge in many regional cities. To enhance data accuracy, precipitation, wind speed, and radiation data from airport gauges (IDEAM, 2022) in each city are integrated. A quality control procedure employing Chauvenet's criterion and elimination of implausible data (e.g., negative radiation) reveals that 10 % of the data exhibits inconsistencies possibly linked to sensor/human errors. Furthermore, this analysis indicates that all stations can be employed for trend calculations and comparison against reanalysis results. It is worth noting that the inclusion of both traffic and industrial stations does not alter any of the conclusions presented in this research.

2.1.2. Reanalysis dataset and validation

The CAMS reanalysis dataset, spanning 8 complete years from 2014-01-01 to 2021-12-31, was acquired with a 3-hour temporal resolution and 0.75° spatial resolution. Validation involves comparing CAMS daily means (daily maximum for ozone) with pollutant averages (8-hour max daily mean for ozone) from available stations for each pollutant and city during 2021. Model validation scores, including Index Of Agreement (IOA), Spearman correlation (Rho), Factor of Two (FAC2), Normalized Mean Bias (NMB), Root Mean Square Error (RMSE), and Mean Bias (MB), along with categorical statistics such as Hit Rate (HIT), False Alarm Rate (FAR), and Proportion of Correctness (POC), are computed following Celis et al. (2022). Benchmarks and thresholds are selected from previous studies (Chai et al., 2013; Gupta and Mohan, 2015; Emery et al., 2017; CCG22) and from the WHO (2021) guidelines (see Tables SM1 and SM2).

Meteorological data, spanning 2012-01-31 to 2022-12-31, is sourced from ERA5 single levels (Hersbach et al., 2020) at a 0.25° spatial resolution and hourly temporal resolution. This data is supplemented with airport weather gauge data to assess trends (Singh et al., 2021). ERA5 pressure levels (Hersbach et al., 2023) from 1000 hPa to 600 hPa, with 25 hPa vertical resolution and 0.25° spatial resolution, are downloaded for the same period and hourly time resolution as single levels. This is used to compute Estimated Inversion Strength (EIS) —essential in explaining pollution variations (e.g., Clifton et al., 2013; Wu et al., 2019)—, following Wang et al. (2023). Calculation employs Romps' (2017) expression for Lifting Condensation Level temperature and height, accounting for city-specific heights.

2.2. Analysis of air pollution trends

For examining air pollution trends in NWSA, we harnessed monitoring station data already subjected to quality control processes outlined in Section 2.1.1. We initiated by conducting a multi-annual analysis for each pollutant, scrutinizing seasonal, monthly, and diurnal patterns, following CCG22's methodology, which draws from Singh et al. (2021). Next, we computed monthly means to execute two analyses. Firstly, we employed Seasonal Trend decomposition using LOESS smoothing (STL) based on Bigi and Ghermandi (2016). This method crafts a smoothed trend line to disentangle data into three components: seasonal, trend, and residual. Subsequently, using the monthly mean, a Generalized Least Squares (GLS) regression from Brockwell and Davis (2002) was fitted to quantify pollutants' linear trends, offering a comprehensive view of the region's air quality evolution.

2.3. Lagrangian framework

The framework arises as a proposal for conducting robust LRT evaluations of pollutants. It is an alternative face to other models such as HYSPLIT which have different drawbacks that we are looking to overcome in this research. It adopts a Lagrangian approach, allowing for tracking and analyzing the evolution of plumes. Our framework also considers crucial aspects, including the spatial extent, concentration, and distance travel by the plumes. Besides, it offers versatility to be applied to multiple datasets such as the CAMS reanalysis dataset, as well as outputs resulting from different models (e.g., WRF-Chem).

2.3.1. Past efforts to understand air pollutants' monthly variation in NWSA

Numerous studies (e.g., Mendez-Espinosa et al., 2020a; Ballesteros-González et al., 2020; Casallas et al., 2023a) have identified increased pollution levels in NWSA due to LRT from wildfires. They employed (i) HYSPLIT to associate wildfires with high $PM_{2.5}$ levels and (ii) mechanism denial experiments in WRF-Chem to account for ozone increments (Ballesteros-González et al., 2020). HYSPLIT, widely employed for LRT studies (e.g., McGowan and Clark, 2008; Stein et al., 2015; Chai et al., 2017), even though it has limitations: it assumes hotspot-to-city transport without factoring in turbulence or dispersion (Pouyaei et al., 2020); it can relate but not quantify pollution reaching regions; and it can't gauge pollution percentage or magnitude upon arrival.

To address these constraints, some authors adopt Lagrangian methodologies (e.g., Wernli and Bourqui, 2002; Putero et al., 2016) that mitigate these limitations to some extent. Brioude et al. (2013) introduced a Lagrangian model integrated with WRF's wind and turbulence schemes (e.g., Madala et al., 2016; Guo et al., 2018). Another approach is the Concentration Trajectory Route of Air pollution model (Pouyaei et al., 2020; Pouyaei et al., 2022), integrated with air quality models, implementing spawning and pruning to balance air packet density, reducing computational costs but forfeiting mass conservation. Our framework stands out as it utilizes gridded data (e.g., NetCDF), rendering it user-friendly and straightforward, unlike some other approaches that necessitate intricate model coupling, such as WRF.

2.3.2. Tracker theory and design

Our method employs a watershed algorithm as a Lagrangian tracker to discern and track pollution plumes to unveil their traits. We implement the watershed algorithm using Scikit-image version 0.2 (Virtonen et al., 2020) in Python 3.9. Initially, the algorithm segments the image, where each segment corresponds to a unique object or part based on distinct characteristics. By applying a threshold, the algorithm identifies and labels objects within image segments exceeding/lacking certain values. For instance, it can label zones where a pollutant exceeds a predefined threshold, which can be fixed or dynamically adjusted based on percentiles/magnitudes over time and space. The boundaries will be defined depending on the plume concentration akin to Pouyaei et al. (2022).

While the algorithm identifies objects, tracking them over time is essential (see Müller et al., 2022 for examples). Our tracker calculates object trajectories by assigning labels in the first time step (3-hourly for CAMS, hourly for WRF-Chem). In subsequent steps, the tracker employs Euclidean distance and overlap of objects to decide if an object is new, creating a new label, or if it's the same as a prior one, maintaining the label. This way, the watershed algorithm tracks pollution plume evolution, facilitating analysis of their lifetime characteristics. In summary, the Lagrangian tracker (i) traces objects from inception to cessation, enabling characterization during their city-to-city transit, (ii) fully relates sources to urban impacts, (iii) quantifies transported pollution volume from temperature hotspots to cities, and (iv) assesses how much city pollution increase stems from related emission sources. Notably, the tracker is adaptable to various gridded data (e.g., CAMS, ERA5) and can be customized for multiple variables (e.g., precipitation, temperature) and diverse applications.

2.3.3. Tracker application and analysis: CAMS reanalysis

This tracker-based approach holds promise for application with diverse gridded datasets, including the publicly available CAMS reanalysis dataset, showcasing its potential for global use. While the results and discussion section dives into specifics, it's notable that CAMS struggles to accurately represent NO and NO_2 in the NWSA, leading us to exclude them from tracker analysis here. The Lagrangian tracker requires a threshold to distinguish pollution plumes. For PM_{10} and $PM_{2.5}$, we employ the spatial and temporal 90th percentile (spatial and temporal for all pollutants) as a threshold, refined through multiple tests with the objective of capturing notable Amazonas and north Colombian events (Mogollón-Sotelo et al., 2021). Tracker assessment was conducted for both $PM_{2.5}$ and PM_{10}, but we present $PM_{2.5}$ results as they align significantly.

Similarly, the 90th percentile serves as the threshold for SO_2. Unlike other pollutants, SO_2 events are not well documented so multiple tests were applied with the idea of effectively capturing a significant number of events from volcanoes and wildfires. In contrast, O_3 analysis adopts the 70th percentile. This threshold accommodates on-land events and not only high Caribbean Sea O_3 plumes (Cooper et al., 2014), and was refined through tests aiming to capture well known events in *Bogotá*, *Medellín*, and *Bucaramanga* (Ballesteros-González et al., 2020; Mendez-Espinosa et al., 2020a, 2020b; Casallas et al., 2023a, 2023b). Our approach enables a comprehensive study using the CAMS reanalysis dataset for general events, supplemented by the WRF-Chem model for specific cases. In essence, this framework elucidates pollutant variations and demonstrates the tracker's versatility for diverse databases and purposes.

2.3.4. Tracker application and analysis: a WRF case study

To demonstrate the algorithm's applicability across various data sources and its suitability for in-depth analysis of distinct pollution events, we employ the WRF-Chem model Version 4.2 (Skamarock et al., 2019). The model incorporates the chemistry module by Grell et al. (2005), enhanced through subsequent advancements (e.g., Grell et al., 2010; Grell et al., 2011). Two nested domains, configured with a 60-second time step for hourly output (domain 2), were established. The outer domain spans Lon: −85.8821 to −62.3299 and Lat: −7.1406 to 16.1739, featuring an 18 km^2 spatial resolution. The inner domain spans Lon: −81.9838 to −66.0658 and Lat: −3.2591 to 12.5555, offering a spatial resolution of 6 km^2 (see Fig. SM2). The model is employed to simulate two highly polluted events, as detailed in Mogollón-Sotelo et al. (2021). The first event, occurring during March 27–28, 2019, is situated over the CVB and northern *Colombia*, regions prone to wildfires. The simulation spans March 20 to April 3, 2019. The second event unfolds over the Amazon, an ecologically significant area, during August 14–16, 2019, with the simulation period spanning August 10 to August 21, 2019. This approach aims to scrutinize diverse and substantial events that impacted *Bogotá* and *Medellín*, respectively, highlighting distinct environmental contexts.

Chemical parameterizations for *Bogotá* and *Medellín* were validated in previous studies (Kumar et al., 2016; Gonzalez et al., 2018). These studies employed the RADM2 chemistry scheme (Stockwell et al., 1990; Stockwell et al., 1997), GO-CART Aerosols scheme with boundary conditions from Chin et al. (2000) and Ginoux et al. (2001), and SORGAM volatile basis parameterization for secondary aerosols (Ahmadov et al., 2012). Gas chemistry followed Wang et al. (2015), while photolysis used the Madronich (1987) scheme. On the physical side, Thompson et al. (2008) microphysical scheme, enhanced by Thompson and Eidhammer (2014) and Iversen et al. (2021), was applied. Radiation employed the RRTMG scheme (Mlawer et al., 1997; Iacono et al., 2008) for Longwave and Shortwave radiation. Surface layer dynamics followed Monin-Obukhov similarity theory (Monin and Obukhov, 1954). The Yonsei University (YSU) Planetary Boundary Layer (PBL) scheme (Hong et al., 2006, based on Hong and Pan, 1996) was chosen for PBL dynamics due to its realism (Potvin et al., 2020). Surface fluxes were modeled using the Smagorinsky 2D scheme (Smagorinsky, 1963). The Grell 3D convective parameterization (Grell and Dévényi, 2002). The convective scheme is used since the inner-domain has a 6 km^2 resolution, being on the verge of the gray zone, this could be a source of error, but as shown by Freitas et al. (2020) the use of parameterized convection at 6 km^2 scales should not produce significant differences compared to a 3 km^2 simulation.

Meteorological boundary conditions stem from the Global Forecast System (GFS) (NCEP, 2015) at 0.25° resolution with 40 vertical levels. Chemical initial conditions adopt the global anthropogenic emissions inventory EDGAR v4.3.1 (Crippa et al., 2016) at 0.1° resolution. Refining the global inventory, local inventories from SDA (2020b) for *Bogotá* and *Medellín* were incorporated following Ballesteros-González et al. (2020). The biogenic inventory MEGAN v2.1 (Guenther et al., 2012) operates at 0.5° resolution. Fire emissions employ the FINN v1.5 database, available daily at 1 km^2 resolution (Wiedinmyer et al., 2011; Pfister et al., 2011). Model validation, although underestimating concentrations (possibly due to errors on the local inventories – Celis et al., 2022), aptly captures peaks and patterns of high pollution events in the CVB and Amazon (Section 3.3).

Application of the tracker centers on $PM_{2.5}$, PM_{10}, and O_3 from WRF-Chem model output, with a focus on $PM_{2.5}$ due to analogous behavior and source. The threshold, set at the 90th percentile, is chosen following tests for regional pollution event capture. The model's higher resolution reveals that events are smaller (compared to CAMS results) due to disintegration during transport, notably influenced by terrain features. Utilizing WRF-Chem output aims to showcase the tracker's functionality, repeating selected diagnostics computed for CAMS data. For *Bogotá* and *Medellín*, well-documented high pollution events and robust emission inventories (Gonzalez et al., 2018; Casallas et al., 2020) enable a detailed analysis of the change in pollution with time, decomposing it into local and advective contributions (Holton, 2004). Excluding deposition, as previous studies suggest LRT as the primary cause for Colombian city pollution peaks (Mendez-Espinosa et al., 2020a; Ballesteros-González et al., 2020; CCG22; Casallas et al., 2023a; Casallas et al., 2023b), the study calculates event contributions (Holton, 2004) from advection, total and local tendencies (Eq. (1)), quantifying LRT's role in pollution dispersion.

$$DC/Dt = \partial C/\partial t + (V \bullet \nabla C) \quad (1)$$

where, DC/Dt is the total tendency, with C being the pollutant concentration and t indicates the time. $\partial C/\partial t$ is the local tendency, and $V \bullet \nabla C$ is the advective term, in which V represents the wind vector and ∇ is an operator $(\partial/\partial x + \partial/\partial y)$.

3. Results and discussion

3.1. Air pollution distribution, trends, and policies

This section describes the CAMS reanalysis validation and the spatial distribution of atmospheric pollution in the NWSA. For this, two aspects are studied: CAMS is validated using statistical and categorical parameters, and describes the pollution "hotspots" of NWSA since pollution is not only located in the main cities but rather in many places that lack in-situ measurements. This section also analyzes the pollutant trends in different cities of NWSA and identifies where further research is needed.

3.1.1. Regional analysis and CAMS validation

In this section, we evaluate the CAMS reanalysis precision in representing six air pollutants and pinpoint locations in NWSA with elevated air pollution levels over the years. For brevity, only MB validation results are displayed (Fig. 1g–l, see Figs. SM3 to SM10 for statistical parameters results), as it captures notable changes (Duveiller et al., 2016) tied to high pollution events. Other statistics are found in the SM. While SO_2 exhibits notable errors, possibly improving in recent years (Inness et al., 2022), we deem it adequate for a detailed source and transport study, with acknowledgment of substantial uncertainties. Regarding NO and NO_2, significant discrepancies are evident (SM for details), potentially arising from limited surface measurements for these pollutants and considerable cloud cover in the region (Hernandez-Deckers, 2021; Casallas et al., 2023a), impeding robust satellite data (Godłowska et al., 2023) assimilation in CAMS.

However, pollutants other than NO_x (i.e., PM_{10}, $PM_{2.5}$, O3, and SO_2) are well-captured by the CAMS dataset for temporal trends. MB and other statistical parameters closely align with benchmarks in most instances. Therefore, we omit NO_x from further analysis in this research. Although spatial distribution of NO_x is described, since concentrations

Promote urban and rural electrification using clean energy sources to significantly cut pollutant emissions. Conduct research to account for potential O_3 increases due to this transition. (iv) Enhanced Early Warning Systems: Strengthen and expand early warning systems, especially in highly polluted areas vulnerable to wildfires and volcanic activity. Implement new systems in contaminated regions and update existing ones in larger cities, including alerts for air pollution due to volcanic activity and wildfires. (v) Improved Air-Quality Networks: Expand monitoring networks with additional stations measuring various pollutants (e.g., VOCs). Install stations in areas affected by orography and fires, such as east of the Eastern Cordillera, and in growing urban centers to address PM_{10} and $PM_{2.5}$ concerns.

In summary, all the phenomena discussed here merit additional studies. They could be used as a starting point for new studies and university courses in NWSA since new and fresh ideas are needed to understand the region's pollution characteristics. We highlight the need for more monitoring stations measuring NO, NO_2, O_3, SO_2, VOCs, and $PM_{2.5}$. This information will help to understand pollutant's spatial and temporal variations and to adapt, evaluate and implement abatement strategies. Our results prove that wildfires affect the region's principal cities and show the importance of tackling wildfires internationally. They also highlight the importance of studying high pollution events as compounds (e.g., Zscheischler et al., 2018; Zscheischler et al., 2020) of drought, heat, wildfires, and air pollution, since pollution episodes are a consequence of a combination of other hazards.

CRediT authorship contribution statement

Alejandro Casallas: Conceptualization, Data curation, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing – original draft. **Ailin Cabrera:** Data curation, Methodology, Software, Validation, Formal analysis, Writing – original draft. **Marco-Andrés Guevara-Luna:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **Adrian Tompkins:** Methodology, Software, Formal analysis, Validation, Visualization, Writing – original draft. **Yuri González:** Data curation, Software, Validation, Writing – original draft. **Juan Aranda:** Data curation, Software, Validation, Writing – original draft. **Luis Carlos Belalcazar:** Conceptualization, Methodology, Formal analysis, Writing – original draft. **Caroline Mogollon-Sotelo:** Data curation, Software, Validation, Visualization, Formal analysis, Writing – original draft. **Nathalia Celis:** Data curation, Software, Validation, Visualization, Formal analysis, Writing – original draft. **Ellie Lopez-Barrera:** Writing – original draft. **Carlos A. Peña-Rincon:** Writing – original draft. **Camilo Ferro:** Data curation, Software, Validation, Visualization, Formal analysis, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ellie Lopez-Barrera reports article publishing charges was provided by Sergio Arboleda University and by the Ibero-American Program of Science and Technology for Development. Juan Aranda reports administrative support was provided by University of La Sabana. Ellie Lopez-Barrera reports article publishing charges was provided by Ibero-American Program of Science and Technology for Development.

Data availability

I have shared the link to my data and some scripts.

Acknowledgments

First and foremost, we express our gratitude to one of the reviewers of the CCG22 work, whose comments on the HYSPLIT analyses served as a catalyst for the development and application of the Lagrangian framework presented in this study. We want to thank Edward Camelo and Leo Durst for their suggestions. We also thank Sebastian K. Muller for discussions on the Lagrangian tracker. Additionally, we are grateful to Yang Chen for discussing the significance of studying wildfires and air quality as compound events during the ICTP "Workshop on Climate Information for Risk Assessment and Regional Adaptation from Global Scale Climate Projections to Local Scale Climate Hazards". We would also like to thank two anonymous reviewers for their constructive comments. They were very important to improve and be more precise in the results of this manuscript. The first author (AC) received a fellowship from the Abdus Salam International Centre for Theoretical Physics – ICTP. The author, JA, would like to thank CAPSAB-Facultad de Ingeniería, Universidad de La Sabana (Project ID: ING-310-2023) for supporting him in this work. This work was partly supported by Sergio Arboleda University (No. EI.BG.086.21.001) and CYTED-RED PREMIA through the project "RED DE PREVENCIÓN, MITIGACIÓN Y REHABILITACIÓN DE ÁREAS AFECTADAS POR INCENDIOS FORESTALES (REDPREMIA)" with the identifier: 921PTE0124. Nevertheless, the funding sources have no involvement in the design or development of the research—special thanks to the NumPy, Matplotlib, Xarray, and Pandas developers' teams.

Statement

During the preparation of this work the author(s) used Grammarly and ChatGPT in order to improve the readability of the paper. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Data availability

We encourage the use of the Lagrangian tracker algorithm, which is available at the following link: https://github.com/Casicus94/Lagrangian_air_quality/tree/main. The other scripts utilized to develop this research are available upon reasonable request to the lead author. The CAMS reanalysis (Inness et al., 2019) data were downloaded from the Atmosphere Data Store, meanwhile the ERA5 reanalysis (Hersbach et al., 2023) data were downloaded from the Climate Data Store, hotspots data from MODIS satellite were downloaded from NASA (2023) and the three datasets are publicly available. The WRF-Chem output used would be available for 5 years in the following link: https://drive.google.com/drive/u/0/folders/1knt3OmPqSuzhHhqSOCRDOiC3LtFNi_8C. The in-situ station data can be downloaded from SISAIRE (2023) or upon request to the lead author. The vehicle counts need to be requested to: https://www.medellin.gov.co/movilidad/ for *Medellín* and to https://www.movilidadbogota.gov.co/web/ in the case of *Bogotá*.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.167350.

References

Adams, D.K., Gutman, S.I., Holub, K.L., Pereira, D.S., 2013. GNSS observations of deep convective time scales in the Amazon. Geophys. Res. Lett. 40, 2818–2823. https://doi.org/10.1002/grl.50573.

Adams, D.K., Fernandes, R.M.S., Holub, K.L., Gutman, S.I., Barbosa, H.M.J., Machado, L.A.T., Calheiros, A.J.P., Bennett, R.A., Kursinski, E.R., Sapucci, L.F., DeMets, C., Chagas, G.F.B., Arellano, A., Filizola, N., Amorim-Rocha, A.A., Araujo-Silva, R., Assuncao, L.M.F., Cirino, G.G., Pauliquevis, T., Portela, B.T.T., Sa, A., Sousa, J.M., Tanaka, L.M.S., 2015. The Amazon dense GNSS meteorological network: a new approach for examining water vapor and deep convection interactions in the tropics. Bull. Am. Meteorol. Soc. 96, 2151–2165. https://doi.org/10.1175/BAMS-D-13-00171.1.

Agudelo, D.C., Lopez-Barrera, E.A., Casallas, A., 2021. Evaluation of the ten-year plan of air decontamination for Bogotá (2010–2020) efficiency. In: 2021 Congreso