Environmental Science and Pollution Research (2020) 27:35993–36012
https://doi.org/10.1007/s11356-020-08268-4

URBAN AIR QUALITY, CLIMATE AND POLLUTION: FROM MEASUREMENT TO MODELING APPLICATIONS

On the mathematical modelling and data assimilation for air pollution assessment in the Tropical Andes

O. L. Quintero Montoya[1] · Elías D. Niño-Ruiz[2] · Nicolás Pinel[3]

Received: 3 October 2019 / Accepted: 27 February 2020 / Published online: 25 April 2020

Abstract
Air pollution assessment in the Tropical Andes requires a multidisciplinary approach. This can be supported from the understanding of the underlying biological dynamics and atmospheric behavior, to the mathematical approach for the proper use of all available information. This review paper touches on several aspects in which mathematical models can help to solve challenging problems regarding air pollution in reviewing the state-of-the-art at the global level and assessing the corresponding state of development as applied to the Tropical Andes. We address the complexities and challenges that modelling atmospheric dynamics in a mega-diverse region with abrupt topography entails. Understanding the relevance of monitoring and facing the problems of data scarcity, we call attention to the usefulness of data assimilation for uncertainty reduction, and how these techniques could help tackle the scarcity of regional monitoring networks to accelerate the implementation and development of modelling systems for air quality in the Tropical Andes. Finally, we suggest a cyberphysical framework for decision-making processes based on the data assimilation of chemical transport models, the forecast of scenarios, and their use in regulation and policy making.

Keywords Computer modelling and simulation · Data assimilation and data analysis · Biodiversity · Conservation

Introduction

The Tropical Andean countries (Venezuela, Colombia, Ecuador, Peru, and Bolivia) face deteriorating air quality. Their two megacities (Bogotá and Lima), and their corresponding countries, rank in the top 50 most polluted capital cities and countries of the world (Air Quality Index 2018), respectively. Most of the large cities in the region exceed the annual average $PM_{2.5}$ World Health Organization guideline of 10 $\mu g/m^3$ (WHO 2016). The peculiar topography (such as abrupt elevation changes, with narrow inter-mountain valleys over 1000 m deep, and transitions from sea level to over 5000 m in less than 200 km) and environmental conditions of the region (such as the bi-annual transit of the Inter Tropical Convergence Zone, and the irregular behavior of the El Niño Southern Oscillation) generate atmospheric conditions that impoverish air quality. Growing urban populations across the region are increasingly exposed to deteriorating air quality. The availability of monitoring stations for atmospheric pollutants in the region is limited. The geographical singularities of the Tropical Andean Region (TAR) present phenomenal challenges for modelling atmospheric chemical dynamics.

Air pollution modelling addresses complex systems. The questions pursued may range from understanding

Responsible Editor: Marcus Schulz

✉ O. L. Quintero Montoya
oquinte1@eafit.edu.co

Elías D. Niño-Ruiz
enino@uninorte.edu.co

Nicolás Pinel
npinelp@eafit.edu.co

[1] Mathematical Modelling Research Group, Universidad EAFIT, Medellín, Colombia

[2] Computer Science Department, Universidad del Norte, Barranquilla, Colombia

[3] Biodiversity Evolution and Conservation, Universidad EAFIT, Medellín, Colombia

broad-scale regional teleconnections through atmospheric transport (Yim et al. 2019) to providing estimates of day-to-day human exposure to atmospheric contaminants (Soares et al. 2014). The questions may also aim to understand nutrient balances and ecosystem integrity (Makowski et al. 2013; Duce et al. 2008; Boy et al. 2008), or they may look towards assessing the loss in agricultural production (Van Dingenen et al. 2009; Feng et al. 2019) due to plant's exposure to ground-level ozone (Silva and Heald 2018). Each one of those questions unveils new layers of complexity, each with its own modelling needs.

Atmospheric pollution in the TAR due to anthropogenic activities has been affecting local ecosystems since early colonial times (Uglietti et al. 2015; Hagan et al. 2011), reflecting dynamics of economic activity (Bandowe et al. 2018; Cooke et al. 2009). Pollutants deriving from domestic burning in the region can have local impacts with widespread consequences, such as the increase in melting of glaciers (Schmitt et al. 2015). Regional transport of contaminants can link distant locations, as exemplified by the affectations in air quality in major Colombian cities from fires in the Eastern savannahs of the country (Mendez-Espinosa et al. 2019), or the transboundary transport of particulate matter into Colombia from Venezuelan fires (Hernandez et al. 2019). Similarly, air quality in the Southern realm of the TAR can be strongly affected by fire activity in Southern Amazonia, a region responsible for 50–75% of fire-derived particulates in South America (Reddington et al. 2015) and whose particulates can deposit over the remote Pacific Ocean up to 4000 km away from the South American coast (Bourgeois et al. 2015).

A wave of air quality crises in the TAR are giving rise to a surging effort to implement and develop modelling and forecasting strategies to support air quality management and public policy to protect human, agricultural, and ecosystem health. This paper provides a critical review on the need for modelling and understanding dynamics at several scales based on the special dynamics in the TAR. To assess the state of air quality studies in the TAR, we compiled the indexed scientific literature retrieved by searching the SCOPUS database with the query "TITLE-ABS-KEY ((atmospheric AND pollution) OR (air AND quality) OR (chemical AND transport AND model)) AND TITLE-ABS-KEY (andes OR colombia OR venezuela OR ecuador OR peru OR bolivia)," which resulted in 607 matches. Manual curation of the results to exclude works outside of the geographic scope or subject of interest yielded 224 articles and 42 conference papers in English (217) or Spanish (49) spanning from 1977 up to January 31, 2020. Of the works represented in the publications, 118 were conducted in or focused on Colombia, 48 on Peru, 38 on Venezuela, 34 on Ecuador, and 18 on Bolivia, with the remaining 10 having a multi-country or regional scope. Of these works, only 25 involved mathematical modelling specific to the TAR. None of them included data assimilation into their modelling schemes. We present a brief summary of the state of the air quality modelling in the region, followed by a brief recounting of data assimilation, looking for the understanding on the mathematical techniques and offering the possibility to tackle via mathematical and computational approaches, some of the challenges for air pollution assessment and prediction. Finally, a decision-making process based on the *human in the loop* concept from control and systems theory is presented, offering the possibility to integrate all the scientific and technical areas looking for the consolidation of a community around the proper policy and regulations on the subject of air pollution.

Air quality modelling

Mathematical models let humans apprehend the dynamics of the natural world. Numerical weather modelling, and the forecasting it enables, has been hailed by renown meteorologist Cliff Mass as one of the great human achievements of the twentieth century ((Mass 2014); for evidence supporting such a grand claim, see figure 6.4.3 of Owens and Hewson (2018)). Chemical transport models (CTM) are mathematical representations of the reactions and processes that form, transform, and transport pollutants in the atmosphere. They are the workhorse of air quality modelling. CTM require initial and boundary conditions to reproduce reality accurately. The former depends mainly on the emissions inventory and land use/land cover data. The latter, provided by means of numerical weather models, offer information of the dominant physical processes in the atmosphere. Numerical models that combine weather forecasting and atmospheric chemistry are referred to as chemical weather forecasting models (CWFM). Studies in air pollution combine information from sensors, models, epidemiology, and emissions inventories.

Atmospheric pollution modelling efforts in Latin America have been pioneered by the Brazilian (Orsini et al. 1967; Andrade et al. 1994; Montero et al. 2001; Ulke and Andrade 2001; Sánchez-Ccoyllo and De Fátima Andrade 2002; Vivanco and Andrade 2006; de Miranda et al. 2012) and Chilean communities (Schmitz 2005), with most of the studies available to date focused on urban areas, and particularly in megacities such as São Paulo, Rio de Janeiro, Santiago de Chile, and Mexico City (Emmons et al. 2010). One example is the METROCLIMA MASP, an initiative of the São Paulo Metropolitan Area jointly tracking climate change and air quality. The scope of the project is in consonance with state-of-art research performed by international groups and the one recommended by the World Meteorological Organization through the use of local mea-

surements, satellite information, and atmospheric modelling for estimating the impact of the urban emission of climate pollutants.

The use of models in air pollution studies in Latin America at different resolutions is a positioned methodology. Another example is the research on fine particulate matter air pollution in the urban environment of São Paulo: relative contribution of regional wildfire and local transportation sources. The project aims to build a numerical modelling framework to understand the source of air pollution in the megacity of São Paulo (with possible later application to other megacities such as Houston). A novel ultra-high-resolution of regional chemistry transport model simulation for State of São Paulo region, with a three-box nested framework, will be developed. Specifically, the model set-up is with an outer domain to cover a large part of southern Brazil at 30 km resolution, an inner domain to cover the state of São Paulo at 10 km resolution, and the most inner to cover the city of São Paulo with 2.5 km resolution. The project is aiming to carefully design the balance of domain size and domain resolution with the explicit goal of attributing local and remote sources. The model simulations will be evaluated against the local chemical measurement obtained by the USP PI (Nogueira et al. 2020).

The first published air quality modelling study in the TAR focused on the Colombian region of the Aburrá Valley (Toro et al. 2006), the second largest population center in the country encompassing the city of Medellín and nine additional con-urban municipalities. The cited study used the EUMAC Zooming Model (EZM (Moussiopoulos 1995) to explore the changes in O_3 concentrations around Medellín in response to changes in the two main ozone precursors, NO_x and Volatile Organic Compounds (VOC). While the work represents a pioneering attempt at modelling in the TA, the authors conclusions regarding increases in VOC as a potential mitigating strategy to ozone formation are in contrast to more recent assessments of the dynamics between NO_x and VOC in other regions (e.g., Wang et al. (2019). Renewed efforts to implement CTM for the Aburrá Valley aim to provide tools for air quality management at urban (e.g., Rendón et al. (2017) and Rodriguez et al. (2017)) and regional (e.g., Pinel et al. (2017)) scales.

The US EPA's atmospheric dispersion model AERMOD has been used to assess the transport dynamics of particulate matter originating from open pit mines in Colombia (Huertas et al. 2012, 2014), and to evaluate the potential historical human exposure to atmospheric mercury resulting from metalliferous mining in Peru (Robins et al. 2012). The AERMOD modelling system does not include atmospheric chemical reactions. Thus, it is only of limited applicability to local and regional air quality modelling.

Most of the atmospheric modelling efforts in the TAR have applied the Weather Research and Forecasting (WRF) model with its atmospheric chemistry module (WRF-Chem), in conjunction with the EDGAR (Emissions Database for Global Atmospheric Research) emissions inventory. The earliest published use of WRF-Chem specifically in the TAR addressed the potential implications for tropospheric ozone concentrations around Quito from the isoprene emissions of oil palm plantations (Parra 2008). WRF-Chem has been employed for simulating the behavior of PM_{10} over the Bogotá metropolitan area (Kumar et al. 2016). In this case, results persistently underestimated by an order of magnitude the PM_{10} concentrations relative to the available measurements. The model WRF-Chem has also been applied to studding the behavior of O_3 over the medium-size, mountain city of Manizales (González et al. 2018) at a horizontal resolution of 1 km. The study compared the performance of the model using the EDGAR emissions inventory and a high-resolution emission inventory previously developed (Gonzalez et al. 2017), obtaining better results with the latter. WRF-Chem has also been applied to studies of carbon monoxide dynamics over Quito under different boundary layer schemes, achieving an agreement with measurements on up to 79% of the days within the simulation (Parra 2017). Similarly, WRF-Chem has been used in Ecuador to assess the dynamics of atmospheric contaminants in Cuenca (Parra 2018a), and of volcanic ash (Parra 2018b), both under various atmospheric boundary layer schemes. Lastly, the atmospheric dynamics of particulate matter in Bogotá have been modelled (Nedbor-Gross et al. 2018) with the Community Multi-scale Air Quality (CMAQ) chemical transport model (LeDuc and Fine 2002), implementing a locally constructed emissions inventory (Pachón et al. 2018). Despite the many efforts at the use of CTMs for studying air quality dynamics in the TAR, none of the published works assimilated available data in order to improve the performance of the models.

From sensors to humans

In the TAR, the vast majority of air quality monitoring stations are located inside urban areas. While they provide information on the general state of air quality, they are designed to monitor concentrations at spatial resolutions that are too coarse for assessing direct human exposure to contamination. For instance, the primary air quality monitoring stations in the urban area of Medellín (Colombia) and its con-urban municipalities consists of 19 $PM_{2.5}$ monitoring stations, at an average density of 8.25 km^2 over the entire area of the 10 municipalities. The integration of atmospheric pollutant data into CTM through data assimilation strengthens the ability to understand local and regional dynamics and fill in the gaps in spatial coverage (Andersson et al. 2007; Beltman et al. 2013; Hendriks et al. 2013). Even so, the

results exist in scales that do not yet explore the interaction human-contaminant.

Epidemiological studies provide insights into the impact that exposure to atmospheric contaminants has on human health. The epidemiological study of disease burden is a established method for understanding the impact of air pollution (Burnett et al. 2018; WHO 2016; Nasari et al. 2016). In the TAR, disease burden from air pollution in four Colombian cities was evaluated using ecological timelines and disease rates (Rodriguez-Villamizar et al. 2018). The incidence of acute respiratory illness in children from Quito was also evaluated in response to changes in air quality management policies (Estrella et al. 2019).

An approach complementary to epidemiological studies is presented by exposure models, which aim to link the fields of atmospheric modelling and environmental health. They recognize that the concentration of atmospheric contaminants may vary significantly in a matter of tens of meters. Exposure models may begin by refining the spatial estimates of contaminants through land use regression models (LURM (Briggs et al. 1997); reviewed by (Ryan and Lemasters 2007; Hoek et al. 2008; Amini et al. 2017b). Parting from high-resolution spatial urban data, LURM have been variously used for estimating potential exposure levels to NO_x (Kryza et al. 2011; Beelen et al. 2013; Wu et al. 2017), particulate matter (Eeftens et al. 2012; Habermann et al. 2015; Zhang et al. 2015), ultra-fine particles (Weichenthal et al. 2016; Van Nunen et al. 2017), organic volatiles (Amini et al. 2017a), and black carbon (Sanchez et al. 2018). A LURM was used to interpret the regional $PM_{2.5}$ concentration patterns between 2011 and 2014 (Londoño-Ciro and Cañón-Barriga 2018).

With high-resolution, spatially explicit data or estimates on contaminant concentration, exposure models can integrate human activity to approach estimates of daily exposure (Soares et al. 2014; Breen et al. 2015). Exposure models such as the one developed for Helsinki (Soares et al. 2014) exemplify the value in the integration of multiple sources of data to construct a model that connects dimensions spanning from atmospheric processes to human behavior. These type of models also highlight the need for solid data fusion approaches that may permit the integration of data from disparate sources.

Uncertainties and inaccuracies

Uncertainty in large-scale models, specially CTM, is a very complicated issue. Uncertainty may be reduced by increasing the accuracy of initial conditions, such as accurate representations of land cover or updated emissions inventories. The historical institutional weaknesses in TAR have delayed the availability of accurate inventories, especially at national scales. Several efforts in Latin America arose few years ago but they are still under development. Even under the idealistic scenario where state-of-the-art emissions inventories are available, population increases and changes in economic activities require that inventories be updated frequently. Consequently, uncertainty may be reduced, but never eliminated. Before describing approaches for adapting the models to the uncertainties inherent in the system through data assimilation in "Data assimilation of complex models and the lack of sensor networks in the Tropical Andes," we first address the situation of model and inventories uncertainties in the TAR.

Emissions inventories

Chemical transport models rely on up-to-date emission inventories for accurate reproduction of pollutant concentrations. While the majority of the available reports of atmospheric pollution modelling in the region rely on global emissions inventories (e.g., Crippa et al. (2012)), the work on the Colombian city of Manizales highlighted the relevance of using locally refined inventories (González et al. 2018).

Sources of atmospheric contaminants in TAR include the expected ones: industry, transportation, fires, and agriculture. While improving constantly, accounting efforts still suffer from imperfect records and loose regulatory oversight. Furthermore, illegal mining and illegal drug production are sources of pollutants that escape official inventories, and that must be incorporated into detailed inventories.

Interest in updating emissions inventories for Latin America gave life to the international project PAPILA (Prediction of Air Pollution in Latin America and the Caribbean). One of the developments expected from the project is the refinement of emissions inventories through coupling of high-resolution satellite data, such as the one being generated by the TROPOspheric Monitoring Instrument (TROPOMI) aboard the Copernicus Sentinel-5 Precursor satellite; with detailed Land Cover/Land Use (LULC). This approach will help fill the gap on emissions inventories related to activities that escape official accounting, such as illegal mining, but also highlight novel regional sources that hitherto have escaped monitoring attention.

Anthropogenic emissions inventories deficiencies are the main weakness for air quality simulations in the TAR, but additional knowledge gaps of potentially significant importance remain. One source of nitrogenous emissions rarely incorporated in air quality modelling is the nitrogen generated by lightning discharges. Lake Maracaibo in Venezuela ranks as the site with the largest number of lighting strikes per year, according to NASA global

lighting flash rate estimates (http://svs.gsfc.nasa.gov/3144). The same analysis lists three sites in Colombia among the top ten, which makes the TAR the region with the second highest lighting flash densities, after the Congo Basin. Refinements in lighting models and the associated tropospheric chemistry using satellite NO_2 observations give an estimate of 9 Tg of N produced globally by lighting each year (Nault et al. 2017). Recent advances in lighting parameterizations in the WRF-Chem model improve the correspondence between estimated NO_2 column concentrations and aircraft-acquired tropospheric NO_2 measurements over the USA (Zhu et al. 2019). The incorporation of lighting-derived chemistry into CTM simulations can lead to significant changes in atmospheric concentration estimates of ozone (Liaskos et al. 2015), particularly over the tropics. Despite the potentially limited importance for simulations aimed at air quality management, proper accounting of NO_x generated by lighting in the tropics can be of great impact for studies aimed at understanding nitrogen fluxes in the TAR, as recent TROPOMI NO_2 data suggests (personal observations).

Modelling biogenic emissions Emissions by plants and soils are a significant source of natural, biogenic emissions of volatile organic compounds (BVOCs, mainly from plant species) and nitrogen oxides (NO and N_2O, mainly from soils). VOCs play a central role in the formation of ground level ozone (O_3) (Monks et al. 2015) and secondary particulate matter (Carlton et al. 2009; Kurtén et al. 2011; Bianchi et al. 2016). Biogenic emissions can impact air quality significantly at local and regional scales (Ma et al. 2019; Wang et al. 2019, Mo et al. 2018; Zhang et al. 2017; Sartelet et al. 2012; Steinbrecher et al. 2009). Not surprisingly, the choice of biogenic emissions model can has considerable impacts in the pollution patterns produced by CTM simulations (Jiang et al. 2019).

Isoprene is the main VOC emitted by plants, which contributes over 0.5 Pg of carbon to the global annual carbon budget (Arneth et al. 2008; Guenther et al. 2012). Isoprene production is a stress response trait that enhances photo-chemical efficiency and protects plants against oxidative damage (Pollastri et al. 2014). Among tropical tree and liana species, the ability to emit isoprene is associated with increased temperature ranges for photosynthetic activity (Taylor et al. 2019). Nearly 20% of all perennial plant species are estimated to be isoprene producers (Loreto and Fineschi 2015), a fraction that could reach up to 40% among tropical tree species (Harley et al. 2004). Tropical biomes are one of the main sources of isoprene and other BVOC (Henrot et al. 2017). Isoprene emission dynamics is a key aspect of BVOC modelling in scenarios of air pollution under climate change, since peak isoprene emissions are associated with drought periods and other events of environmental stress. Agricultural transformations may also impact isoprene emission dynamics in the TAR; using WRF-Chem, Parra (2008) showed the potentially heavy impact in ozone formation in the Quito (Ecuador) and its Metropolitan District due to isoprene emission scenarios associated with the expansion of oil palm plantations.

The primary source for biogenic emissions data used in CTM simulations is the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al. 2006). MEGAN's most current published version is 2.1 (Guenther et al. 2012). (The site https://bai.ess.uci.edu/megan/history indicates that MEGAN 3.0 beta code is available upon request). MEGAN is a modelling framework that provides fluxes for 147 compounds in 19 categories (Guenther et al. 2012). MEGAN has been used to estimate the global fluxes of non-methane BVOC (Sindelarova et al. 2014). It has been integrated into LOTOS-EUROS (Manders et al. 2017), ECHAM6-HAMMOZ (Henrot et al. 2017), and MOZART-4 (Emmons et al. 2010), among many other CTM and Earth System Models. It is capable of spatial resolutions of up to 1 × 1 km, and species resolution that the user may chose at the individual species level (when available), to Plant Functional Types. CTM may improve on the flux estimates provided by MEGAN by adopting regionally specific models, such as the Europe-specific isoprene database (Beltman et al. 2013) used by LOTOS-EUROS for simulations over that continent (Manders et al. 2017). While detailed information is broadly available for temperate forests, for "tropical forests isoprene emission factors are based primarily on above canopy measurements due to the high species diversity" (Guenther et al. 2012).

Tropical rainforests are zonally heterogenous. The Amazon River Basin is estimated to contain over 16,000 different species of trees, and while just 227 of these accounted for nearly 50% of all trees, most of these "hyperdominant" species are dominant in only one of the five different regions distinguished within the Basin (ter Steege et al. 2013). The ability to tolerate seasonal drought is a major determinant in tree species distribution in the Neotropics (Esquivel-Muelbert et al. 2017), likely as a result of the capacity to emit isoprene in response to drought stress. Isoprene emission has been shown to be a key determinant in community structuring in tropical ecosystems (Taylor et al. 2018), and may confer a competitive advantage in scenarios of future climate change (Taylor et al. 2019).

Recent studies applying DA in CTM are trying to use satellite data to constrain BVOC fluxes to improve model performance (Worden et al. 2019). While isoprene cannot be directly observed by satellites, refinement of its emission dynamics can be accomplished through data assimilation (inverse analysis) of satellite observations of formaldehyde (HCHO), one of its primary oxidation products (Kaiser

et al. 2018). Because of the complex chemistry between isoprene and NOx, concurrent satellite observations of NOx and formaldehyde are needed to reduce biases in the estimations. Because biogenic emissions reflect physiological responses from the vegetation to environmental conditions, further improvements in BVOC fluxes can be achieved through data assimilation of meteorological satellite data into CTM, such as cloud location data (Zhang et al. 2018) or the amount of photosynthetically active radiation (PAR) reaching the surface (Zhang et al. 2017, 2018). Improvements in estimates of biogenic emissions can refine not only CTM simulations but also climate change studies, as BVOC can contribute to the atmospheric concentrations of carbon monoxide and tropospheric ozone.

The TA region is rich in ecosystem diversity. It hosts the Chocó-Darién and the Tropical Andes bioregions, two of the world's 25 biodiversity hot spots (Myers et al. 2000). Colombia alone recognizes 98 different general ecosystems, composed of over 8000 distinct biotic communities. It follows then that the first need in adapting CTM to the TA region is the recognition of the need for higher resolution in land cover categorization, and biogenic emission parameterization. For instance, the European Space Agency's Climate Change Initiative Land Cover product (Defourny et al. 2016) contains 38 categories based mostly on broad functional types. It contains no information regarding ecosystem-specific characteristics. Furthermore, once it is used as an input into a CTM, these categories are reduced to even fewer, broader types (9, in the case of LOTOS-EUROS). The diversity of ecosystem processes must be better portrayed in CTM in order to improve model performance in the TA and to afford proper recognition of the most relevant contributors to air pollutants.

Regionally cognizant databases and modelling frameworks are beginning to emerge (see for instance the BIGA initiative, and the associated characterization of VOC emissions in the city of Manizales (Díaz-Poveda 2019); see also Bolaño-Ortiz et al. (2015)); BVOC inventories had been produced for the city of Medellín (Toro et al. 2001), but they were constructed based on coarse land use categories and emission factors obtained from BEIS). Resources that fully integrate the particularities of BVOC emissions of tropical ecosystems into CTM and other air quality modelling efforts are sorely missing. While they are necessary to understand not only atmospheric chemistry but also ecosystem dynamics in the tropics, their construction will require time and resource. A lot of ground and speed could be gained by the deployment of sensors in strategic, periurban areas that monitor the ecotones surrounding urban areas, and through assimilation of the generated data into the models, start dissecting the dynamics of air quality in tropical systems.

Model evaluation and model ensembles

Established atmospheric modelling communities have come to adopt the combination of model results as the preferred approach, embracing a statistical representation of the ensemble to approximate the simulated reality. Ensembles of models can be constructed from a single model run with variations in its initial conditions, or more commonly in atmospheric modelling, from several similar models simulating the same situation (Galmarini et al. 2004). Multimodel ensembles (MME) are now a standard practice in climate modelling, as reviewed by Tebaldi and Knutti (2007). Mathematical formalism for representing ensembles, as well as techniques for finding optimal linear combinations of the results, has been developed (Potempski and Galmarini 2009). The construction of MME has evolved from simply averaging simulation results, as it has been shown that in such a scenario, the combination of the full set of models can be outperformed by specific combination of subsets (composition effect) (Solazzo et al. 2012b). More sophisticated approaches based on spectral decomposition and extraction of the best components from the member models have demonstrated the ability to reduce bias and reduce forecast error (Galmarini et al. 2013). MME are currently used for air quality forecasting in the European continent (Marècal et al. 2015), using an ensemble of seven state-of-the-art regional CTM. The ensemble provides up to 96-h forecasts for the ten most pertinent chemical species at horizontal resolutions of 10–20 km. Similar to the observation by Solazzo et al. (2012b), the evaluation of the MME forecast system conducted by Marècal et al. (2015) at the end of MACC-II (2014) showed that individual models could still outperform the ensemble. Kioutsioukis and Galmarani (2014) introduced a set of "best practices" in the construction of MME, for which they considered the concepts of ensemble diversity and accuracy. A beautiful development of MME was presented by the proposal of two-scale MME schemes that combine global and regional air quality MME (Galmarini et al. 2018), that implemented spectral decomposition and component selection among the individual model results. The hybrid approach resulted in improvements in accuracy over the mono-scale MME (13% over global MME, and 2–3% over regional MME). The authors attribute these improvements to the different strengths of ways in which global and regional models represent atmospheric processes.

Multimodel performance evaluations are frequently conducted by large international consortia. For example, under European Cooperation in Science and Technology (COST) ES0602 action, eighteen operational chemical weather forecasting models on regional and continental scales in Europe

are described and compared in 2012 (Kukkonen et al. 2012). This evaluation included discussions on the entire spectrum of air quality modelling, from how weather forecasting and atmospheric chemistry models were integrated into chemical weather forecasting systems, through strategies for conducting sensitivity analysis and model performance evaluation.

Perhaps the most famous model intercomparison and evaluation effort is the *Air Quality Model Evaluation International Initiative*, or AQMEII (Rao et al. 2011), focused on model performance over Europe and North America following the model evaluation framework proposed by Dennis et al. (2010). On its initial phase, ten state-of-the-science regional air quality (AQ) modelling systems were used for simulating the entire 2006. Standardized modelling outputs from each of the seventeen participating groups were made available on the web-distributed ENSEMBLE system, which allowed statistical and ensemble analyses to be performed by the greater community. As would be the case for the subsequent phases, in Solazzo et al. (2012a), the 1-year model simulations are inter-compared and evaluated with a large set of observations for ground-level particulate matter (PM_{10} and $PM_{2.5}$), as well as its chemical precursors such as SO_2 and NO_2. Analyses of PM_{10} yearly time series and mean diurnal cycle showed a large underestimation proposed as possibly related to errors in the emissions and meteorological inputs (e.g., wind speed and precipitation). The ability of the models to capture high PM concentrations was also evaluated at two separate $PM_{2.5}$ episodes in Europe and North America. Large variability was found among the model's predicted emissions, deposition rates, and concentrations of PM. No single model was found to match consistently the observations for the entire period or region, highlighting the challenges that remained to address (Solazzo et al. 2012a). Similar exercises were made for ozone (Solazzo et al. 2012b). The unprecedented scale of the exercise (two continents, one full year, seventeen modelling groups) marked a milepost that enabled a theretofore unparalleled detailed description of model skills and the uncertainty with respect to their predictions. The AQMEII effort is currently in its fourth phase, focusing on the intercomparison of deposition estimates. AQMEII3 (2016-2018) focused on the impacts of long-range transport of contaminants on regional air quality. One of the highlights among the products of the third phase was the development of methods for apportionment of error to the individual underlying processes (Solazzo et al. 2017). The results from this novel approach emphasized the strength of the interconnection among the meteorology and the atmospheric chemical species, and the differential sources of errors for different types of contaminants.

Model evaluation and intermodel comparisons like the ones described rely on the availability of extensive regional air quality monitoring stations. The quality of the data used for evaluation is of importance for proper performance assessment, as detailed by Solazzo and Galmarini (2015). Neither detailed and rigorous model performance evaluation nor intermodel comparisons are available for the TAR, in large part because of the lack of monitoring networks (particularly outside of the urban areas), but also in part because the modelling community in the region is not as consolidated as the European and North American modelling communities.

Evaluation of large-scale models continues being a difficult task. As reality is not always completely covered by a reliable network of sensors, CTM sometimes are compared against scarce measurements. Mathematically, the solution of differential equations and chemical reactions in a large dimensional grid requires robust measures instead of simplistic approximations of deviations.

The lack of good evaluation methods based on the complexity provides opportunities for further research. Functional data analysis (FDA) encompasses different numerical and statistical methodologies that could help separate climate variability aspects and explore their consequences, improving the system understanding. FDA is used to compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same functions, or different functions for a common set of replicates [Boada et al., *in progress*]. Comparison between observational and model functional data analyses would serve as a technique of model evaluation. Also, FDA could offer the basis to the development of bias correction methods.

From systems and control, the uncertainty has been modelled via identification of the non-linear and/or non-Gaussian dynamics. The phenomenology modelled by stochastic processes provides a rich environment for estimation of the states and parameters of the model. Some approaches came from the variational calculus by solving the initial conditions of the system looking for the trajectories of the system and their relationship with the measured data. Fisherian and Bayesian paradigms allow the description with maximum likelihood and maximum a posteriori of the probability distribution of both system states and outputs. Reducing the uncertainty by adding the observations to the model is not a recent idea. Kalman in early 1960s developed an algorithm solving the optimal filtering problem for linear and Gaussian systems. A wide range of filtering techniques extended the approach to non linear systems and non Gaussian solutions to chemistry-biological models (Quintero et al. 2008, 2009a, b). Their use for small-scale models was positioned among the practitioners and the understanding of certain properties (reversibility/irreversibility) led to close the control loop. Nevertheless, the story for large-scale complex dynamical

systems such as CTM was not the same. The physics underlying the transport phenomena condition the chemical reactions and the dimensionality of the problem represented a challenge to be faced.

Data assimilation of complex models and the lack of sensor networks in the Tropical Andes

Lets assume the atmospheric dynamics can be modelled with a mathematical model, usually written as a discrete time state-space representation:

$$\mathbf{x}_k = \mathcal{M}_{t_{k-1}\to t_k}(\mathbf{x}_{k-1}) \quad \text{for } 0 \le k-1 \le M, \text{ and } \mathbf{x}_k \in \mathbb{R}^{n\times 1}, \tag{1}$$

where $\mathcal{M} : \mathbb{R}^{n\times 1} \to \mathbb{R}^{n\times 1}$ is an imperfect numerical model which mimics the behavior of the ocean and/or the atmosphere.

Data assimilation (DA) is the process whereby an imperfect numerical forecast $\{\mathbf{x}_k^b\}_{k=0}^M$ is adjusted according to real-noisy observations $\{\mathbf{y}_k\}_{k=0}^M$ (Kalnay 2003; Lahoz et al. 2010; Reichle 2008) where $\mathbf{x}_k^b \in \mathbb{R}^{n\times 1}$ is well-known as the background state, k denotes time index where observations are available, for $0 \le k \le M$, n is the state size, M is the size of the assimilation window, $\mathbf{y}_k \in \mathbb{R}^{m\times 1}$ is a vector of observations, and m is the number of observations. The goal is then to estimate the actual state $\mathbf{x}_k^* \in \mathbb{R}^{n\times 1}$ of a (highly) non-linear system which approximately evolves according to some imperfect numerical model.

The scarcity of the sensor networks and spatio temporal variance of the data requires a mathematical, analytical, and computational effort moving the frontiers of knowledge in the field. This challenge was accepted and embraced by the community and variations of the EnKF and variational techniques appear to solve initial conditions for CTM in the assessment of ozone, particulate matter, nitrogenate species, and other chemical compounds in atmosphere in Europe and China.

A promise of bounty lies in the incorporation of DA into atmospheric pollution modelling, not only for its contribution to the reduction of uncertainty but also in opening the door to air quality forecasting. Forecasting with CTM presents us with interesting and complex challenges associated with the uncertainty in weather predictions, the lack of accurate emissions inventory, and the scarcity and sparsity of air quality monitoring networks. These challenges require creative solutions; these challenges represent opportunities for the advancement of knowledge.

DA in CTM is a relatively recent development (Wang et al. 2011; Kumar et al. 2012; Candiani et al. 2013; Joseph et al. 2013; Tilloy et al. 2013; Nguyen et al. 2018). Several efforts have been channeled to the development of an ensemble-based Kalman filter (EnKF) (Evensen 1994, 2003a, b, 2009) for DA in the LOTOS-EUROS model (Sebacher et al. 2013). Additional work looked for the estimation of states and parameters of the model using not only ground-based sensors but also satellite and remotely sensed data (Lu et al. 2016; 2015; 2016a, b; 2017; Jin et al. 2018, 2019).

The ensemble Kalman filter

The ensemble Kalman filter (EnKF) is a sequential Monte Carlo method for parameter and state estimation in highly non-linear models (Lorenc 2003; Evensen 2003a). The popularity of the EnKF obeys to his simple formulation and relatively ease implementation (Burgers et al. 1998). In the EnKF, an ensemble of model realizations is employed to estimate moments of the background error distribution (Houtekamer and Mitchell 1998; Stroud et al. 2018):

$$\mathbf{X}_k^b = \left[\mathbf{x}_k^{b[1]}, \mathbf{x}_k^{b[2]}, \ldots, \mathbf{x}_k^{b[N]}\right] \in \mathbb{R}^{n\times N} \tag{2}$$

where $\mathbf{x}_k^{b[e]} \in \mathbb{R}^{n\times 1}$ stands for the eth ensemble member, for $1 \le e \le N$, at time k, for $0 \le k \le M$. Then, the ensemble mean:

$$\overline{\mathbf{x}}_k^b = \frac{1}{N}\cdot\sum_{e=1}^{N}\mathbf{x}_k^{b[e]} \in \mathbb{R}^{n\times 1} \tag{3}$$

and the ensemble covariance matrix:

$$\mathbf{P}_k^b = \frac{1}{N-1}\cdot\mathbf{\Delta X}_k^b\cdot\left[\mathbf{\Delta X}_k^b\right]^T \in \mathbb{R}^{n\times n} \tag{4}$$

act as estimates of the background state $\mathbf{x}_k^b$ and the background error covariance matrix $\mathbf{B}_k$, respectively, where the matrix of member deviations reads:

$$\mathbf{\Delta X}_k^b = \mathbf{X}_k^b - \overline{\mathbf{x}}_k^b\cdot\mathbf{1}^T \in \mathbb{R}^{n\times N}. \tag{5}$$

Posterior members can be computed via the use synthetic observations:

$$\mathbf{X}_k^a = \mathbf{X}_k^b + \mathbf{\Delta X}_k^a, \tag{6}$$

where the analysis increments can be obtained via the solution of the next linear system:

$$\left[\left[\mathbf{P}_k^b\right]^{-1} + \mathbf{H}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{H}_k\right]\cdot\mathbf{\Delta X}_k^a = \mathbf{H}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{D}_k^s \in \mathbb{R}^{n\times N}, \tag{7}$$

and $\mathbf{D}_k^s \in \mathbb{R}^{m\times N}$ is the innovation matrix on the synthetic observations whose eth column reads $\mathbf{y}_k - \mathbf{H}_k\cdot\mathbf{x}_k^{b[e]} + \boldsymbol{\epsilon}_k^{[e]} \in \mathbb{R}^{m\times 1}$ with $\boldsymbol{\epsilon}_k^{[e]} \sim \mathcal{N}(\mathbf{0}_m, \mathbf{R}_k)$. In practice, model dimensions range in the order of millions while ensemble sizes are constrained by the hundreds and as a direct consequence, sampling errors impact the quality of analysis

increments. To counteract the effects of sampling noise, localizations methods are commonly employed (Greybush et al. 2011; Chen and Oliver 2010), in practice. For instance, methods such as covariance matrix localization (**B**-localization) (Lei et al. 2018), domain localization, and observation localization (**R**-localization) (Anderson 2001, 2019; Han et al. 2018) are employed under operational DA scenarios. Yet another possible choice is to make use of precision covariance matrix estimation. In this context, for instance, the use of the *spatial-predecessors* concept can be employed to obtain sparse estimators of precision matrices (Levina et al. 2008). The predecessors of model component i, from now on $P(i, r)$, for $1 \leq i \leq n$ and a radius of influence $r \in \mathbb{Z}^+$, are given by the set of components whose labels are lesser than that of the ith one. Of course, this will depend on the format employed to label components on a numerical grid. For instance, Fig. 1 shows an example for a two dimensional domain when $r = 1, i = 6$, and model components are labeled by using column-major format.

In the EnKF based on a modified Cholesky decomposition (EnKF-MC) (Nino-Ruiz et al. 2017b, 2018), the following estimator is employed to approximate the precision covariance matrix of the background error distribution (Bickel et al. 2008b):

$$\widehat{\mathbf{B}}_k^{-1} = \widehat{\mathbf{L}}_k^T \cdot \widehat{\mathbf{D}}_k^{-1} \cdot \widehat{\mathbf{L}}_k \in \mathbb{R}^{n \times n}, \tag{8}$$

where the Cholesky factor $\mathbf{L}_k \in \mathbb{R}^{n \times n}$ is a lower triangular matrix,

$$\left\{\widehat{\mathbf{L}}_k\right\}_{i,v} = \begin{cases} -\beta_{i,v,k} & , v \in P(i,r) \\ 1 & , i = v \\ 0 & , \text{otherwise} \end{cases}, \tag{9}$$

whose non-zero sub-diagonal elements $\beta_{i,v,k}$ are obtained by fitting models of the form,

$$\mathbf{x}_{[i]_k}^T = \sum_{v \in P(i,r)} \beta_{i,v,k} \cdot \mathbf{x}_{[v]_k}^T + \boldsymbol{\gamma}_{ik} \in \mathbb{R}^{N \times 1}, 1 \leq i \leq n, \tag{10}$$

where $\mathbf{x}_{[i]_k}^T \in \mathbb{R}^{N \times 1}$ denotes the ith row (model component) of the ensemble Eq. 2, components of vector $\boldsymbol{\gamma}_{ik} \in \mathbb{R}^{N \times 1}$ are samples from a zero-mean normal distribution with unknown variance σ_k^2, and $\mathbf{D}_k \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal elements read,

$$\{\mathbf{D}_k\}_{i,i} = \widehat{\mathbf{var}}\left(\mathbf{x}_{[i]_k}^T - \sum_{v \in P(i,r)} \beta_{i,v,k} \cdot \mathbf{x}_{[j]_k}^T\right)^{-1} \tag{11}$$

$$\approx \mathbf{var}\left(\boldsymbol{\gamma}_{ik}\right)^{-1} = \frac{1}{\sigma_k^2} > 0, \text{ with } \{\mathbf{D}_k\}_{1,1} = \widehat{\mathbf{var}}\left(\mathbf{x}_{[1]_k}^T\right)^{-1}, \tag{12}$$

where $\mathbf{var}(\bullet)$ and $\widehat{\mathbf{var}}(\bullet)$ denote the actual and the empirical variances, respectively. The analysis equations can then be written as follows:

$$\mathbf{X}_k^a = \mathbf{X}_k^b + \left[\widetilde{\mathbf{L}}_k^T \cdot \widetilde{\mathbf{D}}_k^{-1/2}\right]^{-1} \cdot \mathbf{E}_k \in \mathbb{R}^{n \times N}, \tag{13}$$

where

$$\begin{aligned} \widetilde{\mathbf{A}}_k^{-1} &= \widetilde{\mathbf{L}}_k^T \cdot \widetilde{\mathbf{D}}_k^{-1} \cdot \widetilde{\mathbf{L}}_k = \widehat{\mathbf{B}}_k^{-1} + \mathbf{H}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{H}_k \\ &= \widehat{\mathbf{L}}_k^T \cdot \widehat{\mathbf{D}}_k^{-1} \cdot \widehat{\mathbf{L}}_k + \mathbf{H}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{H}_k \in \mathbb{R}^{n \times n} \end{aligned} \tag{14}$$

is an estimate of the posterior precision covariance matrix while the columns of matrix $\mathbf{E}_k \in \mathbb{R}^{n \times N}$ are formed by samples from a standard normal distribution, $\widetilde{\mathbf{L}}_k^T \in \mathbb{R}^{n \times n}$ is a lower triangular matrix (with the same structure as $\widehat{\mathbf{L}}_k$), and $\widetilde{\mathbf{D}}_k^{-1} \in \mathbb{R}^{n \times n}$ is a diagonal matrix. Given the special structure of the left-hand side in Eq. 13, the direct inversion of the matrix $\widetilde{\mathbf{L}}_k \cdot \widetilde{\mathbf{D}}_k^{-1/2} \in \mathbb{R}^{n \times n}$ can be avoided, and instead backward substitutions can be employed to compute the analysis increments. Moreover, the Cholesky factorization of the innovation matrix Eq. 14 can be implicitly performed, for instance, one can update the prior precision factors $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{D}}^{-1}$ with the information brought by the data error correlations onto the model space $\mathbf{H}_k^T \cdot \mathbf{R}_k^{-1} \cdot \mathbf{H}_k \in \mathbb{R}^{n \times n}$ by using rank-one updates over such factors. The algorithm proposed in Nino-Ruiz et al. (2017a) and Nino-Ruiz (2017) (Algorithm 1) can be modified in such manner that a rank m update method can be formulated to estimate posterior factors of precision covariance matrices based on the prior ones, this can be seen in Algorithm 1. The computational effort of this method reads:

$$\mathcal{O}\left(n \cdot m \cdot \varphi^2\right),$$

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

(a) In blue, local box for the model component 6 when $r = 1$.

1	5	9	13
2	6	10	14
3	7	11	15
4	8	12	16

(b) In blue, predecessors of the model component 6 for $r = 1$.

Fig. 1 Local model components (local box) and local predecessors for the model component 6 when $r = 1$. Column-major ordering is utilized to label the model components

where $\varphi \approx N$ stands for the maximum number of non-zero elements per row across all rows in $\widehat{\mathbf{L}}_k$. Neither of matrices $\mathbf{B}_k^{-1}$ are held in memory but their $\widehat{\mathbf{L}}_k$ and $\widehat{\mathbf{D}}_k^{-1}$ factors are enough to perform analysis computations, respectively. Note that, once the posterior factors are estimated, the analysis Eq. 13 can be efficiently employed with no more than Nino-Ruiz (2017):

$$\mathcal{O}\left(n \cdot N + n \cdot \varphi\right) ,$$

long computations (multiplications and divisions), which is equivalent to the computational cost of matrix-free EnKF implementations which do not account for localization: $\mathcal{O}\left(n \cdot m \cdot N + n \cdot N^2\right)$ (Godinez and Moulton 2012).

Algorithm 1 Estimation of posterior factors of the precision covariance matrix $\mathbf{A}^{-1}$ via rank-m updates over the prior ones $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{D}}^{-1}$ with $\mathbf{B}^{-1} \approx \widehat{\mathbf{B}}^{-1} = \widehat{\mathbf{L}}^T \cdot \widehat{\mathbf{D}}^{-1} \cdot \widehat{\mathbf{L}}$.

function COMPUTE_POSTERIOR_FACTORS($\widehat{\mathbf{L}}$, $\widehat{\mathbf{D}}^{-1}$, $\mathbf{H}$, $\mathbf{R}^{-1}$)
  $\mathbf{W} \leftarrow \mathbf{H}^T \cdot \mathbf{R}^{-1/2}$ ▷ A square root approximation of $\mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}$
  for $j \leftarrow 1 \rightarrow m$ **do**
    $\mathbf{d} \leftarrow \mathbf{diag}\left(\widehat{\mathbf{D}}^{-1}\right)$ ▷ The diagonal elements of $\mathbf{D}^{-1}$ are obtained and vectorized
    $\mathbf{w} \leftarrow \mathbf{column}\left(\mathbf{W}, j\right)$ ▷ The j-th column of matrix $\mathbf{W}$ is obtained (and vectorized)
    $a \leftarrow 0$
    $\mathbf{p} \leftarrow \mathbf{0}$
    for $i \leftarrow n \rightarrow 1$ **do**
      $\{\mathbf{p}\}_i \leftarrow \{\mathbf{w}\}_i - \left\{\widehat{\mathbf{L}}_{i,\ell}^T \cdot \mathbf{p}_\ell\right\}_{\ell \in P(i,r)}$
      $\{\mathbf{d}\}_i \leftarrow (1-a) \cdot \{\mathbf{p}\}_i^2 + \{\mathbf{d}\}_i$
      $\eta_i \leftarrow \{\mathbf{d}\}_i^{-1} \cdot (1-a) \cdot \{\mathbf{p}\}_i$
      $a \leftarrow a + \{\mathbf{d}\}_i \cdot \eta_i^2$
    end for
    $\widetilde{\mathbf{L}} \leftarrow \mathbf{I}$
    for $i \leftarrow 1 \rightarrow n$ **do**
      for $v \in P(i,r)$ **do**
        $a \leftarrow \left\{\widehat{\mathbf{L}}_{i,\ell}^T \cdot \mathbf{p}_\ell\right\}_{\ell \in P(i,r)}$
        $\{\widetilde{\mathbf{L}}\}_{i,v} \leftarrow \{\widetilde{\mathbf{L}}\}_{i,v} + \eta_i \cdot \left[\{\mathbf{p}\}_v + a\right]$
      end for
    end for
  end for
  return $\widetilde{\mathbf{L}}$, $\widetilde{\mathbf{D}}^{-1}$
end function

Recent work by our group on the assimilation of the LOTOS-EUROS model for the Aburrá Valley in Colombia is pioneering novel uses for DA in the TAR. The assimilation of surface data from the official network of monitoring stations (SIATA) via EnKF with covariance localization improved the representation of PM_{10} and $PM_{2.5}$dynamics and their concentration estimates within the Valley [López-Restrepo et al., *submitted*]. Our results showed that for calibration of the radius for local analysis, the adjustment of the correlation time length τ, and the estimation of the observation covariance error matrix $\mathbf{R}$, all contributed to significantly improved performance of the model, reproducing the observations more closely than the model without DA.

The estimation of emissions correction factors via DA compensated for the scarcity in accurate and detailed emissions inventories for the region. Forecast performance was sensitive to time and inheritance scheme, demonstrating that the temporal dynamics of pollutant emissions associated with the diurnal and weekly patterns of human activity need to be taken into account in the development of forecast systems. Inheritance schemes cognizant of complex system attributes (e.g., rugged topography, spatially heterogeneous, and highly dynamic meteorology) may yield improved performance and increase the resolution and usability of air quality forecast systems. Improved schemes for covariance localization, necessitated by the complex topography of the Valley, are being developed. Robust estimators are being derived from the lessons in Nino-Ruiz (2017) and Nino-Ruiz et al. (2017b, 2018), with the concurrent development of their extension to CTM.

Variational methods

Another approach to incorporate measurements on the dynamical model is the variational techniques. They use the variational calculus and optimal control theory to estimate the best suitable initial states for an interval using the dynamical discrete model. The definition of such interval allows the derivation of a 3Dimensional variational data assimilation or 4Dimensional models taking into account the time variation of the dynamical 3Dimensional system.

The main idea is to minimize a cost function combining the knowledge of the system and the data. In four-dimensional variational (4D-Var) data assimilation methods (Huang et al. 2009), when observational errors and initial background errors are Gaussian, cost functions of the form (Tr'emolet 2006; Trémolet 2007):

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2} \cdot \left\|\mathbf{x}_0 - \mathbf{x}_0^b\right\|_{\mathbf{B}_0^{-1}}^2 + \frac{1}{2} \cdot \sum_{k=0}^{M-1} \left\|\mathbf{y}_k - \mathcal{H}_k\left(\mathbf{x}_k\right)\right\|_{\mathbf{R}_k^{-1}}^2 , \tag{15}$$

are employed to estimate the initial state $\mathbf{x}_0 \in \mathbb{R}^{n \times 1}$ which best fit the given data $\{\mathbf{y}_k\}_{k=0}^M$:

$$\mathbf{x}_0^a = \arg\min_{\mathbf{x}_0} \mathcal{J}\left(\mathbf{x}_0\right) \text{ , subject to } \mathbf{x}_k = \mathcal{M}_{t_{k-1} \rightarrow t_k}\left(\mathbf{x}_{k-1}\right) . \tag{16}$$

The solution to Eq. 16 requires an explicit expression of the model operator partial derivatives or their tangent linear model. As 4D-Var-based methods rely on tangent linear and adjoint models, the solution is labor-intensive to develop and computationally expensive to run (Ito et al. 2016). This is because the large number of states in a large-scale numerical model or chemical transport model tends to higher orders.

Typically, ensemble-based methods can be employed under 4D-Var operational scenarios to avoid the use of these kind of models (Harlim and Hunt 2007; Miyoshi and Kunii 2012). The main idea behind these approaches is to build ensemble sub-spaces onto which analysis increments can be estimated (Wang et al. 2007). The flow-dependency property of ensemble covariances can be exploited to ensure consistency (with regard to the numerical model dynamics) of the resulting analysis estimations (Gustafsson 2007). However, since ensemble members come at high computational costs, ensemble sizes are bounden by the hundreds while model resolutions are constrained by the billions (Yin et al. 2015).

As a direct consequence, analysis increments can be impacted by sampling noise which can result in poor estimation of posterior members. Although localization methods can be employed to counteract the effects of sampling noise, their use under 4D-EnKF formulations is not clear (in regard to their efficient implementation, for instance). We think that, the use of a modified Cholesky decomposition (Bickel et al. 2008a; Rothman et al. 2009) can be exploited in the context of 4D-Var methods to estimate a full-rank square root approximation of $\mathbf{B}$ at assimilation steps, the ranges of these estimations can be treated as new control spaces onto which analysis increments can be estimated. In this manner, we can mitigate the impact of sampling noise and even more, given the special structure of the employed estimator, efficient and practical formulations of hybrid methods can be feasible.

Hybrid methods

Inaccuracies in LULC data, while diminishing thanks to the increasing availability of satellite imagery, are still a source of uncertainty for CTM, specially since model categories and parameterizations do not reflect fully the realities on the ground. A possible strategy, as mentioned above in the section on inventories 3, is the use of chemical satellite and other remotely sensed data for DA through variational approaches. Variational assimilation of the highly complex systems represented by CTM faces the difficulty of formulating the *ad joint* of the linear tangent model. This hurdle is to overcome by means of ensemble-based variational techniques, as in van Velzen and Segers (2010). This is still an incipient field, a wide open area of study replete with opportunities for developing computational and numerical techniques to obtain alternative best estimates for the system's states and parameters.

The construction of an tangent linear and/or an adjoint model for real, large forecast models is an extremely labor-intensive process. For instance, the adjoint model of the high-resolution limited area modelling (HIRLAM) 4D-Var (Gustafsson and Bojarova 2014; Stengel et al. 2009) was developed in 10 years in which most of the time was spent to detect and to fix errors in the tangent and the adjoint models (Gustafsson 2007). In order to avoid the implementation of such models, one can employ ensemble-based methods to encapsulate the statistical evolution of initial errors and to build sub-space onto which initial analysis increments can be estimated. As is well-known, these methods naturally propagate flow-dependent background error covariance matrices. For instance, in Four-Dimensional Variational Ensemble Data Assimilation (4D-EnKF) (Liu et al. 2008; Lorenc et al. 2015), the model trajectory is constrained to the space spanned by the background ensemble members, this is:

$$\mathbf{x}_k = \overline{\mathbf{x}}_k^b + \mathbf{\Delta X}_k \cdot \mathbf{w}\,, \tag{17}$$

where $\mathbf{w} \in \mathbb{R}^{N\times 1}$ is a vector in redundant coordinates to be determined later. This equivalent to:

$$\mathbf{x}_k - \overline{\mathbf{x}}_k^b \in \mathbf{range}\left\{\mathbf{\Delta X}_k\right\} \approx \mathbf{range}\left\{\mathbf{B}_k^{1/2}\right\}\,.$$

Therefore, the analysis increments are computed onto the space given by a low-rank square root approximation of the background error covariance matrices at observation times. By replacing Eq. 17 into Eq. 15, one obtains:

$$\begin{aligned}\mathcal{J}(\mathbf{x}_0) &= \mathcal{J}\left(\overline{\mathbf{x}}_0^b + \mathbf{\Delta X}_0 \cdot \mathbf{w}\right)\\ &= \widehat{\mathcal{J}}(\mathbf{w}) = \frac{(N-1)}{2}\cdot\|\mathbf{w}\|^2\\ &\quad + \frac{1}{2}\cdot\sum_{k=0}^{M}\|\mathbf{d}_k - \mathbf{Q}_k\cdot\mathbf{w}\|^2_{\mathbf{R}_k^{-1}}\,,\end{aligned} \tag{18}$$

where $\mathbf{d}_k = \mathbf{y}_k - \mathbf{H}_k\cdot\overline{\mathbf{x}}_k^b \in \mathbb{R}^{m\times 1}$ is the innovation vector and $\mathbf{Q}_k = \mathbf{H}_k\cdot\mathbf{\Delta X}_k^b \in \mathbb{R}^{m\times N}$. The optimal values of the control variable $\mathbf{w}$ is then seek in order to estimate the initial analysis state:

$$\mathbf{w}^* = \arg\min_{\mathbf{w}}\widehat{\mathcal{J}}(\mathbf{w})\,. \tag{19}$$

The gradient of Eq. 18 is equal to:

$$\nabla_{\mathbf{w}}\widehat{\mathcal{J}}(\mathbf{w}) = (N-1)\cdot\mathbf{w} - \sum_{k=0}^{M}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot[\mathbf{d}_k - \mathbf{Q}_k\cdot\mathbf{w}]$$
$$= \left[(N-1)\cdot\mathbf{I} + \sum_{k=0}^{M}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{Q}_k\right]\cdot\mathbf{w}$$
$$- \sum_{k=0}^{M}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{d}_k \in \mathbb{R}^{N\times 1}, \tag{20}$$

and by setting this gradient to zero, the optimal weights Eq. 19 read:

$$\mathbf{w}^* = \left[(N-1)\cdot\mathbf{I} + \sum_{k=0}^{M}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{Q}_k\right]^{-1}$$
$$\cdot\sum_{k=0}^{M}\mathbf{Q}_k^T\cdot\mathbf{R}_k^{-1}\cdot\mathbf{d}_k \in \mathbb{R}^{N\times 1}, \tag{21}$$

from which the initial analysis state can be estimated:

$$\overline{\mathbf{x}}_0^a = \overline{\mathbf{x}}_0^b + \mathbf{\Delta}\mathbf{X}_0^b\cdot\mathbf{w}^*. \tag{22}$$

Since Eq. 17 represents an approximated solution rather than an exact solution, the initial analysis is only recovered and propagated in time in order to obtain an approximation of the optimal trajectory of Eq. 16. Note that all computations are performed onto the ensemble space Eq. 17 and therefore, the computational cost of estimating Eq. 22 is linearly bounded with regard to the model size n and the number of observations m (Ruiz and Sandu 2016).

Currently, our work is directed towards the implementation of the 4DEnVar technique, followed by an EnKS-MC ensemble Kalman smoother with modified Cholesky technique in order to compare the behavior of two ensemble-based techniques assimilating satellite measurements of the NO_2 from the TROPOMI instrument to assess the emission and deposition of nitrogen species for parameter estimation in natural protected areas over Colombia [Yarce et al., *under construction*]. Our work focuses on introducing a singular value decomposition of the ensemble members to estimate the dominant autovectors degeneracy due to the time propagation of the members, and a Cholesky decomposition for inverse covariance matrix background covariance matrix estimations in a hybrid adjoint-free data assimilation technique.

Mathematical models as decision-making tools

The field of control systems aims to design and implement control laws on dynamical systems. The main goal is to find a set of finite control actions that carry the system from point A to point B in finite time. In order to achieve the best behavior possible for the system, a reference point or trajectory is needed and a control law must be designed and calculated according to certain criteria such as stability, robustness, or performance.

The control law is designed based on knowledge about the system, often represented as a mathematical model. The complexity of the control law deployed increases according to the complexity of the model. Linear control, nonlinear control, robust control, predictive control, adaptive control, intelligent control, stochastic control, and many other branches in the field of control theory arose from the need to adapt to the particularities of both the model and the system.

The set of control actions must be not only feasible but also realizable and performed with finite energy. Consequently, the cost function includes particularities to bound the solution to the problem. By closing the loop between controller and system, the entire phenomenon changes and improves. This notion can be extended to the decision and regulatory policy making regarding air pollution. If so, how could it be described?

Assuming the correct development and implementation of data assimilation for CTM as techniques and developments mentioned in "Data assimilation of complex models and the lack of sensor networks in the Tropical Andes," we may thus claim that the models describe and forecast accurately air pollution dispersion scenarios.

The next step would be to design a finite set of control actions. The optimal control actions refer to the set of strategies for mitigation and control of emissions and pollutants expelled and transformed on the atmosphere. Two types of scenarios at different scales arise: (1) high-resolution scenarios for urban zones; (2) scenarios of lower resolution of regional development and ecosystem protection. Forecast and source apportionment, jointly with emission factors estimation, may be products of a careful data assimilation and model ensemble scheme. Some examples such as Marco Polo- Panda project exemplifies the main goal of our north western south America initiative. A natural way of thinking is the regulatory assessment based in evidence and science. A natural enemy may be the complex geodynamics; there is a lot of science to be done.

From fundamental research to operational environments, a long path must be followed (Solazzo et al. 2013). But, in the case of data assimilation in CTM, once the observations provide information to the model, forecast is an immediate step. Forecast of pollutants' dispersion requires also a good weather numerical model. The daily cycle must be depicted. Several realizations of the same process may produce a good ensemble. Numerical errors may challenge the nature of the stochasticity added to the model but every CTM may offer

the best of its reaction schemes to the ensemble (Anderson 2012, 2019).

Regulations based on modelling and scenarios and traffic models used as dynamic emissions serve as the core of the control law design. Several years ago, this type of approach was used for the telecommunications market in Ecuador; mathematical evidence was presented but political motivations declined the technical suggestions (Bastidas et al. 2014). Mitigation strategies may be studied and simulated and the effects for urban planning take place among the decision makers. The cyberphysical system when humans decide whether or not act on the systems is the next generation of tools for air pollution assessment in Latin America. Computers provide powerful visualization allowing the cognitive process of decision makers to evaluate qualitatively the effects of their choices and the sources to be mitigated by means of new technologies. Cyberphysical systems involve several new advanced techniques from control and systems world, human machine interfaces and human computer interaction techniques deploy to policy makers the effect of the social, economic, and health impacts.

But when the human closes the loop, information represents the milestone. A formal approach to decision theory involves Bayesian formulations and the relevance of air pollution assessments in public health, economy, agriculture, biodiversity, and society requires the deep study of the behavioral economics. The air pollution is not necessarily a field limited to environmental and chemical engineers, requires a multidisciplinary effort.

Decision-making for air pollution requires the reduction on the uncertainties for modelling, from emissions inventory to land use and land cover. Influence of human activities in meteorology and the complex terrain of the Tropical Andes provide challenges for numerical weather prediction. Mathematical models offer a suitable tool mixed with monitoring stations. The scarcity and spatio-temporal variance of observations provides difficulties for assimilation of the models, but the interdisciplinary scientific community keep working on formal, real scenario and generalized new data assimilation techniques for uncertainty reduction.

Data assimilation offers a dynamical driven alternative to diminish the lack of knowledge on the air pollution dynamics in short time scales. Adding surface, satellite, in situ, and laser-based remote sensing data to a model will enrich the knowledge for proper scenario simulation and on line decision-making. The vast amount or data also allows the possibility to design machine learning-based preprocessing data platforms pre-data assimilation.

As developed towards this paper, none of this is impossible to tackle. Evidence built from the last years supports the potential for a scientific and technical advances. Nevertheless, the humans continue being the ones deciding whether or not move forward to chaotic air quality scenarios.

Conclusions

The time is ripe for the coming together of the air quality modelling community to address the challenges and opportunities of modelling atmospheric chemical dynamics in the Tropical Andean region. The region lacks broad networks of monitoring stations such as those available in the European and North American regions. Because of the mega-diversity of the natural ecosystems present in the region, biogenic emissions are a major source of uncertainty for CTM simulations. Furthermore, because of the remoteness of some of the territories, and the sometimes illegal activities, certain sources of emissions are seen only from the air. To the time of submission of this review, a critical wild fires on the Amazon's region were advertised by remote sensing.

For adequate performance, CTM implemented in the TAR will require all the help they can get to make the most of the available data. International alliances such as PAPILA will help fill some of the gaps in the advancement of modelling in the region in addition to the emissions inventory.

Data assimilation for CTMs is not a new scientific research area. In Europe and China, we found evidence of the benefits of adding different sources of information to the models, improving their results. Nevertheless, data assimilation in chemical process is not a straightforward duty. It requires a set of mathematical and computational skills along with the understanding of the dominant physics and chemical reactions, leading to the enrichment of interdisciplinary teams. Sources of data from ground, in situ, and remote will help minimize uncertainty from initial and boundary conditions by means of DA. They also provide complexity to the new DA techniques and algorithms due to the spatio temporal scarcity of the data.

DA applied to CTM can also help the process of source apportionment. Source apportionment is the process whereby the in is a fundamental step in implementing (Nguyen et al. 2018).

By having a good model with reduced uncertainty, a natural step is forecasting of not only urban zones but also emission and deposition dynamics from protected areas and water sources. Agricultural activities may also be planned by a long-term prediction. Daily cycles analysis for urban pollution and scenarios of dynamic and static emissions sources modification arise as a good option of decision-making process.

We introduced and developed the idea of cyberphysical systems for air pollution management and opened insights

for future research. We plan to set an operative cyber-physical system based on assimilated models for citizens behavioral actions concerning air pollution.

Funding information This work was financially supported in part by award UN 2018-38 and by the Applied Math and Computer Science Lab at Universidad del Norte, Colombia, also partly supported from the research grant "Data Assimilation Schemes in Colombian Geodynamics - Cooperative Research Plan for 2017–2020 between Universidad EAFIT and TUDelft" made feasible the statements, ideas, and developments.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Air Quality Index (2018) World air quality index. IQAir Air Visual. https://www.airvisual.com/world-most-polluted-cities/world-air-quality-report-2018-en.pdf

Amini H, Hosseini V, Schindler C, Hassankhany H, Yunesian M, Henderson SB, Künzli N (2017a) Spatiotemporal description of BTEX volatile organic compounds in a middle eastern megacity: Tehran study of exposure prediction for environmental health research (Tehran SEPEHR). Environ Pollut 226:219–229

Amini H, Yunesian M, Hosseini V, Schindler C, Henderson SB, Künzli N (2017b) A systematic review of land use regression models for volatile organic compounds. Atmos Environ 171:1–16

Anderson JL (2001) An ensemble adjustment kalman filter for data assimilation. Mon Weather Rev 129(12):2884–2903

Anderson JL (2012) Localization and sampling error correction in ensemble Kalman filter data assimilation. Mon Weather Rev 140(7):2359–2371

Anderson JL (2019) A nonlinear rank regression method for ensemble kalman filter data assimilation. Mon Weather Rev 147(8):2847–2860

Andersson C, Langner J, Bergstrom R (2007) Interannual variation and trends in air pollution over Europe due to climate variability during 1958–2001 simulated with a regional CTM coupled to the ERA40 reanalysis. Tellus 59(3):77–98

Andrade F, Orsini C, Maenhaut W (1994) Relation between aerosol sources and meteorological parameters for inhalable atmospheric particles in Sao Paulo City, Brazil. Atmos Environ 28(14):2307–2315

Arneth A, Monson RK, Schurgers G, Niinemets Ü, Palmer PI (2008) Why are estimates of global terrestrial isoprene emissions so similar (and why is this not so for monoterpenes)? Atmos Chem Phys 8(16):4605–4620

Bandowe BAM, Frnkl L, Grosjean M, Tylmann W, Mosquera PV, Hampel H, Schneider T (2018) A 150-year record of polycyclic aromatic compound (PAC) deposition from high Andean Cajas National Park, southern Ecuador. Science of the Total Environment 621:1652-1663. https://doi.org/10.1016/j.scitotenv.2017.10.060

Bastidas M, Quintero O. L., Garcia J (2014) nteligencia De Mercados: Comportamientos Estratgicos Sobre Precios De Oferta En El Mercado Spot Elctrico Colombiano //(Market Intelligence: Strategic Behaviors with Prices in the Colombian Electricity Spot). Documentos de Economia y Fiannzas 16:1. http://dx.doi.org/10.2139/ssrn.2393225

Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, Pedeli X, Tsai MY, Künzli N, Schikowski T, Marcon A, Eriksen KT, Raaschou-Nielsen O, Stephanou E, Patelarou E, Lanki T, Yli-Tuomi T, Declercq C, Falq G, Stempfelet M, Birk M, Cyrys J, von Klot S, Nádor G, Varró MJ, Dedele A, Gražulevičiene R, Mölter A, Lindley S, Madsen C, Cesaroni G, Ranzi A, Badaloni C, Hoffmann B, Nonnemacher M, Krämer U, Kuhlbusch T, Cirach M, de Nazelle A, Nieuwenhuijsen M, Bellander T, Korek M, Olsson D, Strömgren M, Dons E, Jerrett M, Fischer P, Wang M, Brunekreef B, de Hoogh K (2013) Development of NO 2 and NO x land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. Atmos Environ 72(2):10–23

Beltman JB, Hendriks C, Tum M, Schaap M (2013) The impact of large scale biomass production on ozone air pollution in Europe. Atmos Environ 71:352–363

Bianchi F, Barmet P, Stirnweis L, El Haddad I, Platt SM, Saurer M, Lötscher C, Siegwolf R, Bigi A, Hoyle CR, DeCarlo PF, Slowik JG, Prévôt ASH, Baltensperger U, Dommen J (2016) Contribution of methane to aerosol carbon mass. Atmospheric Environment 141:41–47. https://doi.org/10.1016/j.atmosenv.2016.06.036

Bickel PJ, Levina E et al (2008a) Covariance regularization by thresholding. Ann Stat 36(6):2577–2604

Bickel PJ, Levina E et al (2008b) Regularized estimation of large covariance matrices. Ann Stat 36(1):199–227

Boy J, Rollenbeck R, Valarezo C, Wilcke W (2008) Amazonian biomass burning-derived acid and nutrient deposition in the north Andean montane forest of Ecuador. Global Biogeochemical Cycles 22(4):1–16. https://doi.org/10.1029/2007GB003158

Bolaño-Ortiz TR, Camargo-Caicedo Y, Vélez-Pereira AM (2015) Emisiones Biogénicas De Monoterpenos En El Parque Nacional Natural Tayrona, Santa Marta (Colombia). Luna Azul 40:102–116

Bourgeois Q, Ekman AML, Krejci R (2015) Aerosol transport over the andes from the amazon basin to the remote Pacific Ocean: A multiyear CALIOP assessment. Journal of Geophysical Research 120(16):8411–8425. https://doi.org/10.1002/2015JD023254

Breen MS, Long TC, Schultz BD, Williams RW, Richmond-Bryant J, Breen M, Langstaff JE, Devlin RB, Schneider A, Burke JM, Batterman SA, Meng QY (2015) Air pollution exposure model for individuals (EMI) in health studies: evaluation for ambient PM2.5 in central North Carolina. Environ Sci Tech 49(24):14184–14194

Briggs DJ, Collins S, Elliott P, Fischer P, Kingham S, Lebret E, Pryl K, Van Reeuwijk H, Smallbone K, Van Der Veen A (1997) Mapping urban air pollution using gis: a regression-based approach. Int J Geogr Inf Sci 11(7):699–718

Burgers G, Jan van Leeuwen P, Evensen G (1998) Analysis scheme in the ensemble kalman filter. Mon Weather Rev 126(6):1719–1724

Burnett R, Chen H, Szyszkowicz M, Fann N, Hubbell B, Pope CA, Apte JS, Brauer M, Cohen A, Weichenthal S, Coggins J, Di Q, Brunekreef B, Frostad J, Lim SS, Kan H, Walker KD, Thurston GD, Hayes RB, Lim CC, Turner MC, Jerrett M, Krewski D, Gapstur SM, Diver WR, Ostro B, Goldberg D, Crouse DL, Martin RV, Peters P, Pinault L, Tjepkema M, Van Donkelaar A, Villeneuve PJ, Miller AB, Yin P, Zhou M, Wang L, Janssen NA, Marra M, Atkinson RW, Tsang H, Thach TQ, Cannon JB, Allen RT, Hart JE, Laden F, Cesaroni G, Forastiere F, Weinmayr G, Jaensch A, Nagel G, Concin H, Spadaro JV (2018) Global estimates of mortality associated with longterm exposure to outdoor fine particulate matter. Proc Natl Acad Sci USA 115(38):9592–9597

Candiani G, Carnevale C, Finzi G, Pisoni E, Volta M (2013) A comparison of reanalysis techniques: applying optimal interpolation and ensemble kalman filtering to improve air quality monitoring at mesoscale. Sci Total Environ 458-460:7–14

Carlton AG, Wiedinmyer C, Kroll JH (2009) A review of Secondary organic aerosol (SOA) formation from isoprene. Atmospheric Chemistry and Physics 9(14):4987–5005. https://doi.org/10.5194/acp-9-4987-2009

Chen Y, Oliver DS (2010) Cross-covariances and localization for enkf in multiphase flow data assimilation. Comput Geosci 14(4):579–601

Crippa M, Guizzardi D, Muntean M, Schaaf E, Monni S, Doering U, Petrescu AMR (2012) EDGAR v4. 3. 2 global atlas of the three major greenhouse gas emissions for the period 1970-2012. Central Europe 2010:1–20

Cooke CA, Wolfe AP, Hobbs WO (2009) Lake-sediment geochemistry reveals 1400 years of evolving extractive metallurgy at Cerro de Pasco, Peruvian Andes. Geology 37(11):1019–1022. https://doi.org/10.1130/G30276A.1

de Miranda RM, de Fatima Andrade M, Fornaro A, Astolfo R, de Andre PA, Saldiva P (2012) Urban air pollution: a representative survey of PM 2.5 mass concentrations in six Brazilian cities. Air Quality. Atmosphere and Health 5(1):63–77

Defourny P, Kirches G, Brockmann C, Boettcher M, Peters M, Bontemps S, Lamarche C, Schlerf M, Santoro M (2016) Land cover CCI: product user guide Version 2

Dennis R, Fox T, Fuentes M, Gilliland A, Hanna S, Hogrefe C, Irwin J, Rao ST, Scheffe R, Schere K, Steyn D, Venkatram A (2010) A framework for evaluating regional-scale numerical photochemical modeling systems. Environ Fluid Mech 10(4):471–489

Díaz-Poveda VC (2019) Evaluación de los compuestos orgánicos volátiles (COV) biogénicos y antropogénicos del municipio de Manizales Masters thesis. Universidad Nacional de Colombia, Manizales

Duce RA, LaRoche J, Altieri K, Arrigo KR, Baker AR, Capone DG, Cornell S, Dentener F, Galloway J, Ganeshram RS, Geider RJ, Jickells T, Kuypers MM, Langlois R, Liss PS, Liu SM, Middelburg JJ, Moore CM, Nickovic S, Oschlies A, Pedersen T, Prospero J, Schlitzer R, Seitzinger S, Sorensen LL, Uematsu M, Ulloa O, Voss M, Ward B, Zamora L (2008) Impacts of atmospheric anthropogenic nitrogen on the open ocean. Science 320(5878):893–897

Eeftens M, Beelen R, De Hoogh K, Bellander T, Cesaroni G, Cirach M, Declercq C, Dedele A, Dons E, De Nazelle A, Dimakopoulou K, Eriksen K, Falq G, Fischer P, Galassi C, Gražulevičiene R, Heinrich J, Hoffmann B, Jerrett M, Keidel D, Korek M, Lanki T, Lindley S, Madsen C, Mölter A, Nádor G, Nieuwenhuijsen M, Nonnemacher M, Pedeli X, Raaschou-Nielsen O, Patelarou E, Quass U, Ranzi A, Schindler C, Stempfelet M, Stephanou E, Sugiri D, Tsai MY, Yli-Tuomi T, Varró MJ, Vienneau D, Klot SV, Wolf K, Brunekreef B, Hoek G (2012) Development of land use regression models for PM 2.5, PM 2.5 absorbance, PM 10 and PM coarse in 20 European study areas; Results of the ESCAPE project. Environ Sci Technol 46(20):11195–11205

Emmons LK, Walters S, Hess PG, Lamarque JF, Pfister GG, Fillmore D, Granier C, Guenther A, Kinnison D, Laepple T, Orlando J, Tie X, Tyndall G, Wiedinmyer C, Baughcum SL, Kloster S (2010) Description and evaluation of the model for ozone and related chemical tracers, version 4 (MOZART-4). Geosci Model Dev 3(1):43–67

Esquivel-Muelbert A, Baker TR, Dexter KG, Lewis SL, ter Steege H, Lopez-Gonzalez G, Monteagudo Mendoza A, Brienen R, Feldpausch TR, Pitman N, Alonso A, van der Heijden G, Peña-Claros M, Ahuite M, Alexiaides M, Álvarez Dávila E, Murakami AA, Arroyo L, Aulestia M, Balslev H, Barroso J, Boot R, Cano A, Chama Moscoso V, Comiskey JA, Cornejo F, Dallmeier F, Daly DC, Dávila N, Duivenvoorden JF, Duque Montoya AJ, Erwin T, Di Fiore A, Fredericksen T, Fuentes A, García-Villacorta R, Gonzales T, Guevara Andino JE, Honorio Coronado EN, Huamantupa-Chuquimaco I, Killeen TJ, Malhi Y, Mendoza C, Mogollón H, Jørgensen PM, Montero JC, Mostacedo B, Nauray W, Neill D, Vargas PN, Palacios S, Palacios Cuenca W, Pallqui Camacho NC, Peacock J, Phillips JF, Pickavance G, Quesada CA, Ramírez-Angulo H, Restrepo Z, Reynel Rodriguez C, Paredes MR, Sierra R, Silveira M, Stevenson P, Stropp J, Terborgh J, Tirado M, Toledo M, Torres-Lezama A, Umaña MN, Urrego LE, Vasquez Martinez R, Gamarra LV, Vela CI, Vilanova Torre E, Vos V, von Hildebrand P, Vriesendorp C, Wang O, Young KR, Zartman CE, Phillips OL (2017) Seasonal drought limits tree species across the Neotropics. Ecography 40(5):618–629

Estrella B, Sempértegui F, Franco OH, Cepeda M, Naumova EN (2019) Air pollution control and the occurrence of acute respiratory illness in school children of Quito, Ecuador. J Public Health Policy 40(1):17–34

Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. 99

Evensen G (2003a) The ensemble kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53(4):343–367

Evensen G (2003b) The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn 53(4):343–367

Evensen G (2009) The ensemble Kalman filter for combined state and parameter estimation. IEEE Control Syst 29(3):83–104

Feng Z, De Marco A, Anav A, Gualtieri M, Sicard P, Tian H, Fornasier F, Tao F, Guo A, Paoletti E. (2019) Economic losses due to ozone impacts on human health, forest productivity and crop yield across China. Environment International 131(May):104966. https://doi.org/10.1016/j.envint.2019.104966

Fu G, Heemink A, Lu S, Segers A, Weber K, Lin HX (2016a) Model-based aviation advice on distal volcanic ash clouds by assimilating aircraft in situ measurements. Atmos Chem Phys 16(14):9189–9200

Fu G, Lin H-X, Heemink A, Segers A, Prata F, Lu S (2016b) Satellite data assimilation to improve forecasts of volcanic ash concentrations. Atmos Chem Phys Discuss 18:1–22

Fu G, Lin HX, Heemink a. W., Segers a. J., Lu S, Palsson T (2015) Assimilating aircraft-based measurements to improve forecast accuracy of volcanic ash transport. Atmos Environ 115:170–184

Fu G, Prata F, Xiang Lin H, Heemink A, Segers A, Lu S (2017) Data assimilation for volcanic ash plumes using a satellite observational operator: a case study on the 2010 Eyjafjallajökull volcanic eruption. Atmos Chem Phys 17(2):1187–1205

Galmarini S, Bianconi R, Klug W, Mikkelsen T, Addis R, Andronopoulos S, Astrup P, Baklanov A, Bartniki J, Bartzis JC, Bellasio R, Bompay F, Buckley R, Bouzom M, Champion H, D'Amours R, Davakis E, Eleveld H, Geertsema GT, Glaab H, Kollax M, Ilvonen M, Manning A, Pechinger U, Persson C, Polreich E, Potemski S, Prodanova M, Saltbones J, Slaper H, Sofiev MA, Syrakov D, Sørensen JH, Auwera LVD, Valkama I, Zelazny R (2004) Ensemble dispersion forecasting - part I: concept, approach and indicators. Atmos Environ 38(28):4607–4617

Galmarini S, Kioutsioukis I, Solazzo E (2013) E pluribus unum: ensemble air quality predictions. Atmos Chem Phys 13(14):7153–7182

Galmarini S, Kioutsioukis I, Solazzo E, Alyuz U, Balzarini A, Bellasio R, Benedictow AM, Bianconi R, Bieser J, Brandt J, Christensen JH, Colette A, Curci G, Davila Y, Dong X, Flemming J, Francis X, Fraser A, Fu J, Henze DK, Hogrefe C, Im U, Vivanco MG, Jiménez-Guerrero P, Jonson JE, Kitwiroon N, Manders A, Mathur R, Palacios-Peña L, Pirovano G, Pozzoli L, Prank M, Schultz M, Sokhi RS, Sudo K, Tuccella P, Takemura T, Sekiya T, Unal A (2018) Two-scale multi-model ensemble: is a hybrid ensemble of opportunity telling us more? Atmos Chem Phys 18(12):8727–8744

Godinez HC, Moulton JD (2012) An efficient matrix-free algorithm for the ensemble kalman filter. Comput Geosci 16(3):565–575
Gonzalez CM, Gomez CD, Rojas NY, Acevedo H, Aristizabal BH (2017) Relative impact of on-road vehicular and point-source industrial emissions of air pollutants in a medium-sized Andean city. Atmos Environ 152:279–289
González CM, Ynoue RY, Vara-Vela A, Rojas NY, Aristizábal BH (2018) High-resolution air quality modeling in a medium-sized city in the tropical Andes: assessment of local and global emissions in understanding ozone and PM10 dynamics. Atmos Poll Res 2017:1–15
Greybush SJ, Kalnay E, Miyoshi T, Ide K, Hunt BR (2011) Balance and ensemble kalman filter localization techniques. Mon Weather Rev 139(2):511–522
Guenther A, Karl T, Harley P, Weidinmyer C, Palmer PI, Geron C (2006) Edinburgh research explorer estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature) and physics e of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases an. Atmos Chem Phys 6:3181–3210
Guenther AB, Jiang X, Heald CL, Sakulyanontvittaya T, Duhl T, Emmons LK, Wang X (2012) The model of emissions of gases and aerosols from nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. Geosci Model Dev 5(6):1471–1492
Gustafsson N (2007) Discussion on '4d-var or enkf?' Tellus A: Dynamic Meteorology and Oceanography 59(5):774–777
Gustafsson N, Bojarova J (2014) Four-dimensional ensemble variational (4d-en-var) data assimilation for the high resolution limited area model (hirlam). Nonlinear Processes Geophys 21(4):745–762
Habermann M, Billger M, Haeger-Eugensson M (2015) Land use regression as method to model air pollution. Previous results for Gothenburg/Sweden. Procedia Eng 115(0):21–28
Hagan N, Robins N, Hsu-Kim H, Halabi S, Morris M, Woodall G, Zhang T, Bacon A, Richter DdeB, Vandenberg J (2011) Estimating historical atmospheric mercury concentrations from silver mining and their legacies in present-day surface soil in Potosí, Bolivia. Atmospheric Environment 45(40):7619–7626. https://doi.org/10.1016/j.atmosenv.2010.10.009
Han Y, Zhang J, Sun D (2018) Error control and adjustment method for underwater wireless sensor network localization. Appl Acoust 130:293–299
Harley P, Vasconcellos P, Vierling L, Pinheiro CCDS, Greenberg J, Guenther A, Klinger L, De Almeida SS, Neill D, Baker T, Phillips O, Malhi Y (2004) Variation in potential for isoprene emissions among Neotropical forest sites. Glob Chang Biol 10(5):630–650
Harlim J, Hunt BR (2007) Four-dimensional local ensemble transform kalman filter: numerical experiments with a global circulation model. Tellus A: Dynamic Meteorology and Oceanography 59(5):731–748
Hendriks C, Kranenburg R, Kuenen J, van Gijlswijk R, Wichink Kruit R, Segers A, Denier van der Gon H, Schaap M (2013) The origin of ambient particulate matter concentrations in the Netherlands. Atmos Environ 69:289–303
Henrot AJ, Stanelle T, Schröder S, Siegenthaler C, Taraborrelli D, Schultz MG (2017) Implementation of the MEGAN (v2.1) biogenic emission model in the ECHAM6-HAMMOZ chemistry climate model. Geosci Model Dev 10(2):903–926
Hernandez AJ, Morales-Rincon LA, Wu D, Mallia D, Lin JC, Jimenez R (2019) Transboundary transport of biomass burning aerosols and photochemical pollution in the Orinoco River Basin. Atmospheric Environment 205(45):1–8. https://doi.org/10.1016/j.atmosenv.2019.01.051
Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, Briggs D (2008) A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos Environ 42(33):7561–7578
Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble kalman filter technique. Mon Weather Rev 126(3):796–811
Huang X-Y, Xiao Q, Barker DM, Zhang X, Michalakes J, Huang W, Henderson T, Bray J, Chen Y, Ma Z et al (2009) Four-dimensional variational data assimilation for wrf: formulation and preliminary results. Mon Weather Rev 137(1):299–314
Huertas JI, Huertas ME, Izquierdo S, Gonzlez ED (2012) Air quality impact assessment of multiple open pit coal mines in northern Colombia. Journal of Environmental Management 93(1):121-129. https://doi.org/10.1016/j.jenvman.2011.08.007
Huertas JI, Huertas ME, Cervantes G, Daz J (2014) Assessment of the natural sources of particulate matter on the opencast mines air quality. Science of the Total Environment 493:1047-1055. https://doi.org/10.1016/j.scitotenv.2014.05.111
Ito S-I, Nagao H, Yamanaka A, Tsukada Y, Koyama T, Kano M, Inoue J (2016) Data assimilation for massive autonomous systems based on a second-order adjoint method. Phys Rev E 94(4):043307
Jiang J, Aksoyoglu S, Ciarelli G, Oikonomakis E, El-Haddad I, Canonaco F, O'dowd C, Ovadnevaite J, Minguillón MC, Baltensperger U, Cruz M (2019) Effects of two different biogenic emission models on modelled ozone and aerosol concentrations in Europe. Atmos Chem Phys 19(6):3747–3768
Jin J, Lin HX, Heemink A, Segers A (2018) Spatially varying parameter estimation for dust emissions using reduced-tangent-linearization 4DVar. Atmospheric Environment. https://doi.org/10.1016/j.atmosenv.2018.05.060
Jin J, Lin HX, Segers A, Xie Y, Heemink A (2019) Machine learning for observation bias correction with application to dust storm data assimilation. Atmospheric Chemistry and Physics 19(15):10009–10026. https://doi.org/10.5194/acp-19-10009-2019
Joseph J, Sharif HO, Sunil T, Alamgir H (2013) Application of validation data for assessing spatial interpolation methods for 8-h ozone or other sparsely monitored constituents. Environ Pollut 178:411–418
Kalnay E (2003) Atmospheric modeling, data assimilation and predictability. Cambridge University Press, Cambridge
Kioutsioukis I, Galmarini S (2014) De praeceptis ferendis: good practice in multi-model ensembles. Atmos Chem Phys 14(21):11791–11815
Kryza M, Szymanowski M, Dore AJ, Werner M (2011) Application of a land - use regression model for calculation of the spatial pattern of annual NOx air concentrations at national scale: a case study for Poland. Proc Environ Sci 7:98–103
Kukkonen J, Olsson T, Schultz DM, Baklanov A, Klein T, Miranda AI, Monteiro A, Hirtl M, Tarvainen V, Boy M, Peuch V-H, Poupkou A, Kioutsioukis I, Finardi S, Sofiev M, Sokhi R, Lehtinen KEJ, Karatzas K, San José R, Astitha M, Kallos G, Schaap M, Reimer E, Jakobs H, Eben K (2012) A review of operational, regional-scale, chemical weather forecasting models in europe. Atmos Chem Phys 12(1):1–87
Kumar A, Jiménez R, Belalcázar LC, Rojas NY (2016) Application of WRF-chem model to simulate PM10 concentration over Bogota. Aerosol Air Qual Res 16(5):1206–1221
Kumar U, De Ridder K, Lefebvre W, Janssen S (2012) Data assimilation of surface air pollutants (O 3 and NO 2) in the regional-scale air quality model AURORA. Atmos Environ 60(2):99–108
Kurtén T, Zhou L, Makkonen R, Merikanto J, Räisänen P, Boy M, Richards N, Rap A, Smolander S, Sogachev A, Guenther A, Mann GW, Carslaw K, Kulmala M (2011) Large methane releases lead to strong aerosol forcing and reduced cloudiness. Atmospheric

Chemistry and Physics 11(14):6961–6969. https://doi.org/10.5194/acp-11-6961-2011
Lahoz W, Khattatov B, Menard R (2010) Data assimilation. Springer, Berlin
LeDuc S, Fine S (2002) Models-3/Community Multiscale Air Quality (CMAQ) Modeling System. In: Borrego C, Schayes G (eds) Air Pollution Modeling and Its Application XV. Springer US, Boston, MA, pp 307–310. https://doi.org/10.1007/0-306-47813-7_30
Lei L, Whitaker JS, Bishop C (2018) Improving assimilation of radiance observations by implementing model space localization in an ensemble kalman filter. J Adv Model Earth Syst 10(12):3221–3232
Levina E, Rothman A, Zhu J et al (2008) Sparse estimation of large covariance matrices via a nested lasso penalty. Ann Appl Stat 2(1):245–263
Liu C, Xiao Q, Wang B (2008) An ensemble-based four-dimensional variational data assimilation scheme. part i: technical formulation and preliminary test. Mon Weather Rev 136(9):3363–3373
Liaskos CE, Allen DJ, Pickering KE (2015) Journal of Geophysical Research : Atmospheres by replay simulations with GEOS-5. Journal of Geophysical Research Atmospheres 120:8512–8534. https://doi.org/10.1002/2014JD022987
Londoño-Ciro LA, Cañón-Barriga JE (2018) Metodología para la caracterización espacio-temporal de PM2. 5 en el área urbana de la ciudad de Medellín-Colombia Spatio-Temporal Characterization Methodology of PM2. 5 in the Urban Area of the City of Medellín-Colombia Abstract Metodologia para a cara. Revista EIA 574:113–132
Lorenc AC (2003) The potential of the ensemble kalman filter for nwp—a comparison with 4d-var. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 129(595):3183–3203
Lorenc AC, Bowler NE, Clayton AM, Pring SR, Fairbairn D (2015) Comparison of hybrid-4denvar and hybrid-4dvar data assimilation methods for global nwp. Mon Weather Rev 143(1):212–229
Loreto F, Fineschi S (2015) Reconciling functions and evolution of isoprene emission in higher plants. N Phytol 206(2):578–582
Lu S, Lin HX, Heemink AW, Fu G, Segers AJ (2016) Estimation of volcanic ash emissions using trajectory-based 4D-Var data assimilation. Mon Weather Rev 144(2):575–589
Ma M, Gao Y, Wang Y, Zhang S, Leung LR, Liu C, Wang S, Zhao B, Chang X, Su H, Zhang T, Sheng L, Yao X, Gao H (2019) Substantial ozone enhancement over the north China plain from increased biogenic emissions due to heat waves and land cover in summer 2017. Atmos Chem Phys Discuss 2019:1–27
Makowski GS, Rollenbeck R, Fabian P, Bendix J (2013) Complex topography influences atmospheric nitrate deposition in a neotropical mountain rainforest. Atmospheric Environment 79:385–394. https://doi.org/10.1016/j.atmosenv.2013.06.023
Manders AMM, Builtjes PJH, Curier L, Denier Van Der Gon HAC, Hendriks C, Jonkers S, Kranenburg R, Kuenen JJP, Segers AJ, Timmermans RMA, Visschedijk AJH, Kruit RJW, Addo W, Van Pul J, Sauter FJ, Van Der Swaluw E, Swart DPJ, Douros J, Eskes H, Van Meijgaard E, Van Ulft B, Van Velthoven P, Banzhaf S, Mues AC, Stern R, Fu G, Lu S, Heemink A, Van Velzen N, Schaap M (2017) Curriculum vitae of the LOTOS–EUROS (v2.0) chemistry transport model. Geosci Model Dev 10:4145–4173
Marécal V, Peuch V, Andersson C, Andersson S, Arteta J, Beekmann M, Benedictow A, Bergström R (2015) A regional air quality forecasting system over Europe : the MACC-II daily ensemble production. Geosci Model Dev 2:2777–2813
Mass C (2014) Cliff mass weather and climate blog
Mendez-Espinosa JF, Belalcazar LC, Morales Betancourt R (2019) Regional air quality impact of northern South America biomass burning emissions. Atmospheric Environment 203(January):131–140. https://doi.org/10.1016/j.atmosenv.2019.01.042
Miyoshi T, Kunii M (2012) The local ensemble transform kalman filter with the weather research and forecasting model: experiments with real observations. Pure Appl Geophys 169(3):321–333
Montero L, Vasconcellos PC, Souza SR, Pires MA, Sanchez-Ccoyllo OR, Andrade MF, Carvalho LR (2001) Measurements of atmospheric carboxylic acids and carbonyl compounds in São Paulo City, Brazil. Environ Sci Tech 35(15):3071–3081
Monks PS, Archibald AT, Colette A, Cooper O, Coyle M, Derwent R, Fowler D, Granier C, Law KS, Mills GE, Stevenson DS, Tarasova O, Thouret V, Von Schneidemesser E, Sommariva R, Wild O, Williams ML (2015) Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer. Atmospheric Chemistry and Physics 15(15):8889–8973. https://doi.org/10.5194/acp-15-8889-2015
Moussiopoulos N (1995) The EUMAC Zooming Model, a tool for local-to-regional air quality studies. Meteorology and Atmospheric Physics 57(1-4):115–133. https://doi.org/10.1007/BF01044157
Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403(6772):853–858
Nasari MM, Szyszkowicz M, Chen H, Crouse D, Turner MC, Jerrett M, Pope CA, Hubbell B, Fann N, Cohen A, Gapstur SM, Diver WR, Stieb D, Forouzanfar MH, Kim SY, Olives C, Krewski D, Burnett RT (2016) A class of non-linear exposure-response models suitable for health impact assessment applicable to large cohort studies of ambient air pollution. Air Qual Atmosp Health 9(8):961–972
Nault BA, Laughner JL, Wooldridge PJ, Crounse JD, Dibb J, Diskin G, Peischl J, Podolske JR, Pollack IB, Ryerson TB, Scheuer E, Wennberg PO, Cohen RC (2017) Lightning NOx Emissions: Reconciling Measured and Modeled Estimates With Updated NOx Chemistry. Geophysical Research Letters 44(18):9479–9488. https://doi.org/10.1002/2017GL074436
Nedbor-Gross R, Henderson BH, Pèrez-Peña MP, Pachón J. E. (2018) Air quality modeling in Bogotá Colombia using local emissions and natural mitigation factor adjustment for re-suspended particulate matter. Atmos Poll Res 9(1):95–104
Nguyen CV, Soulhac L, Salizzoni P (2018) Source apportionment and data assimilation in urban air quality modelling for NO2: the lyon case study. Atmosphere, 9
Nino-Ruiz E (2017) A matrix-free posterior ensemble kalman filter implementation based on a modified cholesky decomposition. Atmosphere 8(7):125
Nino-Ruiz ED, Mancilla A, Calabria JC (2017a) A posterior ensemble kalman filter based on a modified cholesky decomposition. Procedia Computer Science, 108:2049–2058. International Conference on Computational Science, ICCS 2017 12-14 June 2017, Zurich, Switzerland
Nino-Ruiz ED, Sandu A, Deng X (2017b) A parallel implementation of the ensemble kalman filter based on modified cholesky decomposition. Journal of Computational Science
Nino-Ruiz ED, Sandu A, Deng X (2018) An ensemble kalman filter implementation based on modified cholesky decomposition for inverse covariance matrix estimation. SIAM J Sci Comput 40(2):A867–A886
Nogueira T, Kumar P, Nardocci A, de Fatima Andrade M (2020) Public health implications of particulate matter inside bus terminals in Sao Paulo, Brazil. Science of the Total Environment 711:135064

Orsini CQ, Tabacniks MH, Artaxo P, Andrade MF, Kerr AS (1967) Characteristics of fine and coarse particles of natural and urban aerosols of Brazil. Atmos Environ 20(11):2259–2269

Owens RG, Hewson TD (2018) ECMWF forecast user guide. Reading: ECMWF. Technical report

Pachón JE, Galvis B, Lombana O, Carmona LG, Fajardo S, Rincón A, Meneses S, Chaparro R, Nedbor-Gross R, Henderson B (2018) Development and evaluation of a comprehensive atmospheric emission inventory for air quality modeling in the megacity of Bogotà. Atmosphere 9(2):1–17

Parra R (2008) Contribution of oil palm isoprene emissions to tropospheric ozone levels in the Distrito Metropolitano de Quito (Ecuador). WIT Transactions on Ecology and the Environment 116:95–104. https://doi.org/10.2495/AIR080111

Parra Rene (2017) Assessment of planetary boundary layer schemes of the WRF-Chem model in the simulation of carbon monoxide dispersion in the urban area of Quito, Ecuador. WIT Transactions on Ecology and the Environment 211:41–50. 10.2495/AIR170041

Parra René (2018) Performance studies of planetary boundary layer schemes in WRF-Chem for the Andean region of Southern Ecuador. Atmospheric Pollution Research 9(3):411–428. https://doi.org/10.1016/j.apr.2017.11.011, https://doi.org/10.1016/j.apr.2017.11.011

Parra René (2018) Influence of boundary layer schemes in modeling the dispersion and sedimentation of volcanic ash in Ecuador. WIT Transactions on Ecology and the Environment 230(1):83–94. https://doi.org/10.2495/AIR180081

Pinel N, Salazar JF, Posada JA, Rodríguez M, Rendón AM, Quintero OL, Yarce A (2017) Potential urban pollution impacts on protected areas in Colombia through atmospheric teleconnections. In: Santos JM, Andrade MDF, Toledo T (eds) Proceedings of the 3rd CMAS South America Air Quality Conference, pages 152–154, Vitória. Fundação Espírito Santense de Tecnologia

Pollastri S, Tsonev T, Loreto F (2014) Isoprene improves photochemical efficiency and enhances heat dissipation in plants at physiological temperatures. J Exp Bot 65(6):1565–1570

Potempski S, Galmarini S (2009) Est modus in rebus: analytical properties of multi-model ensembles. Atmos Chem Phys 9(24):9471–9489

Quintero M, Olga L, Amicarelli A, di Sciascio F, Scaglia G (2008) State estimation in alcoholic continuous fermentation of zymomonas mobilis using recursive bayesian filtering: A simulation approach. Bioresources 3(2):316–334

Quintero M, Olga L, Amicarelli A, Scaglia G, di Sciascio F (2009a) Control based on numerical methods and recursive Bayesian estimation in a continuous alcoholic fermentation process. BioResources 4(4):1372–1395

Quintero M, Olga L, Nieto J, Amicarelli A, Scaglia G, Luna T (2009b) Control engineering perspective of fermentation process from zymomonas mobilis: Modeling, state estimation and control

Rao ST, Galmarini S, Puckett K (2011) Air quality model evaluation international initiative (AQMEII): advancing the state of the science in regional photochemical modeling and its applications. Bull Am Meteorol Soc 92(1):23–30

Reddington CL, Butt EW, Ridley DA, Artaxo P, Morgan WT, Coe H, Spracklen DV (2015) Air quality and human health improvements from reductions in deforestation-related fire in Brazil. Nature Geoscience 8(10):768–771. https://doi.org/10.1038/ngeo2535

Reichle RH (2008) Data assimilation methods in the earth sciences. Adv Water Res 31(11):1411–1418

Rendón AM, Salazar JF, Wirth V, Quintero OL (2017) Proceedings of the 3rd CMAS South America Air Quality Conference, pages 146–148, Vitória. Fundação Espírito Santense de Tecnologia. In: Santos JM, Andrade MDF, Toledo T (eds)

Robins NA, Hagan N, Halabi S, Hsu-Kim H, Gonzales RDE, Morris M, Woodall G, Richter DD, Heine P, Zhang T, Bacon A, Vandenberg J (2012) Estimations of historical atmospheric mercury concentrations from mercury refining and present-day soil concentrations of total mercury in Huancavelica, Peru. Science of the Total Environment 426:146-154. https://doi.org/10.1016/j.scitotenv.2012.03.082

Rodriguez M, Yarce A, Rendón AM, Quintero O, Pinel N (2017) Characterization and analysis of sattelite and ground data available for the Aburrá Valley (Medellín Metropolitan Area) as inputs for air quality models. In: Proceedings of the 3rd CMAS South America air quality conference, pp 149–151, Vitoria. Fundação Espírito Santense de Tecnologia

Rodriguez-Villamizar LA, Rojas-Roa NY, Blanco-Becerra LC, Herrera-Galindo VM, Fernàndez-Niño JA (2018) Short-term effects of air pollution on respiratory and circulatory morbidity in Colombia 2011–2014: A multi-city, time-series analysis. Int J Environ Res Public Health 15(8):E1610

Rothman AJ, Levina E, Zhu J (2009) Generalized thresholding of large covariance matrices. J Am Stat Assoc 104(485):177–186

Ruiz EDN, Sandu A (2016) A derivative-free trust region framework for variational data assimilation. J Comput Appl Math 293:164–179

Ryan PH, Lemasters GK (2007) A review of land-use regression models for characterizing intraurban air pollution exposure. Inhal Toxicol 19(SUPPL. 1):127–133

Sanchez M, Ambros A, Milà C, Salmon M, Balakrishnan K, Sambandam S, Sreekanth V, Marshall JD, Tonne C (2018) Development of land-use regression models for fine particles and black carbon in peri-urban South India. Sci Total Environ 634:77–86

Sànchez-Ccoyllo OR, De Fátima Andrade M (2002) The influence of meteorological conditions on the behavior of pollutants concentrations in São Paulo, Brazil. Environ Pollut 116(2):257–263

Sartelet KN, Couvidat F, Seigneur C, Roustan Y (2012) Impact of biogenic emissions on air quality over Europe and North America. Atmos Environ 53:131–141

Schmitt C. G., All J. D., Schwarz J. P., Arnott W. P., Cole R. J., Lapham E., Celestian A. (2015) Measurements of light-absorbing particles on the glaciers in the Cordillera Blanca, Peru. Cryosphere 9(1):331–340. https://doi.org/10.5194/tc-9-331-2015

Schmitz R (2005) Modelling of air pollution dispersion in Santiago de Chile. Atmos Environ 39(11):2035–2047

Sebacher B, Hanea R, Heemink A (2013) A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). Comput Geosci 17(5):813–832

Silva SJ, Heald CL (2018) Investigating dry deposition of ozone to vegetation. J Geophys Res Atmos 123(1):559–573

Sindelarova K, Granier C, Bouarar I, Guenther A, Tilmes S, Stavrakou T, Müller JF, Kuhn U, Stefani P, Knorr W (2014) Global data set of biogenic VOC emissions calculated by the MEGAN model over the last 30 years. Atmos Chem Phys 14(17):9317–9341

Soares J, Kousa A, Kukkonen J, Matilainen L, Kangas L, Kauhaniemi M, Riikonen K, Jalkanen JP, Rasila T, Hànninen O, Koskentalo T, Aarnio M, Hendriks C, Karppinen A (2014) Refinement of a model for evaluating the population exposure in an urban area. Geosci Model Dev 7(5):1855–1872

Solazzo E, Bianconi R, Pirovano G, Matthias V, Vautard R, Moran MD, Appel KW, Bessagnet B, Brandt J, Christensen JH, Chemel C, Coll I, Ferreira J, Forkel R, Francis XV, Grell G, Grossi P, Hansen AB, Miranda AI, Nopmongcol U, Prank M, Sartelet KN, Schaap M, Silver JD, Sokhi RS, Vira J, Werhahn J, Wolke R, Yarwood G, Zhang J, Rao ST, Galmarini S (2012a) Operational model evaluation for particulate matter in europe and

north america in the context of aqmeii. Atmos Environ 53:75–92. AQMEII: An International Initiative for the Evaluation of Regional-Scale Air Quality Models - Phase 1

Solazzo E, Bianconi R, Pirovano G, Moran MD, Vautard R, Hogrefe C, Appel KW, Matthias V, Grossi P, Bessagnet B, Brandt J, Chemel C, Christensen JH, Forkel R, Francis XV, Hansen AB, McKeen S, Nopmongcol U, Prank M, Sartelet KN, Segers A, Silver JD, Yarwood G, Werhahn J, Zhang J, Rao ST, Galmarini S (2013) Evaluating the capability of regional-scale air quality models to capture the vertical distribution of pollutants. Geosci Model Dev 6(3):791–818

Solazzo E, Bianconi R, Vautard R, Appel KW, Moran MD, Hogrefe C, Bessagnet B, Brandt J, Christensen JH, Chemel C, Coll I, Denier van der Gon H, Ferreira J, Forkel R, Francis XV, Grell G, Grossi P, Hansen AB, Jeričević A, Kraljević L, Miranda AI, Nopmongcol U, Pirovano G, Prank M, Riccio A, Sartelet KN, Schaap M, Silver JD, Sokhi RS, Vira J, Werhahn J, Wolke R, Yarwood G, Zhang J, Rao TS, Galmarini S (2012b) Model evaluation and ensemble modelling of surface-level ozone in Europe and North America in the context of AQMEII. Atmos Environ 53:60–74

Solazzo E, Galmarini S (2015) Comparing apples with apples: using spatially distributed time series of monitoring data for model evaluation. Atmos Environ 112:234–245

Solazzo E, Hogrefe C, Colette A, Garcia-Vivanco M, Galmarini S (2017) Advanced error diagnostics of the CMAQ and Chimere modelling systems within the AQMEII3 model evaluation framework. Atmos Chem Phys 17(17):10435–10465

Steinbrecher R, Smiatek G, Köble R, Seufert G, Theloke J, Hauff K, Ciccioli P, Vautard R, Curci G (2009) Intra- and inter-annual variability of VOC emissions from natural and semi-natural vegetation in Europe and neighbouring countries. Atmos Environ 43(7):1380–1391

Stengel M, Undén P, Lindskog M, Dahlgren P, Gustafsson N, Bennartz R (2009) Assimilation of seviri infrared radiances with hirlam 4d-var. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 135(645):2100–2109

Stroud JR, Katzfuss M, Wikle CK (2018) A bayesian adaptive ensemble kalman filter for sequential state and parameter estimation. Mon Weather Rev 146(1):373–386

Taylor TC, McMahon SM, Smith MN, Boyle B, Violle C, van Haren J, Simova I, Meir P, Ferreira LV, de Camargo PB, da Costa AC, Enquist BJ, Saleska SR (2018) Isoprene emission structures tropical tree biogeography and community assembly responses to climate. N Phytol 220(2):435–446

Taylor TC, Smith MN, Slot M, Feeley KJ (2019) The capacity to emit isoprene differentiates the photosynthetic temperature responses of tropical plant species. Plant Cell Environ 42(8):2448–2457

Tebaldi C, Knutti R (2007) The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society A: Mathematical. Phys Eng Sci 365(1857):2053–2075

ter Steege H, Pitman NCA, Sabatier D, Baraloto C, Salomão RP, Guevara JE, Phillips OL, Castilho CV, Magnusson WE, Molino J-F, Monteagudo A, Núñez Vargas P, Montero JC, Feldpausch TR, Coronado ENH, Killeen TJ, Mostacedo B, Vasquez R, Assis RL, Terborgh J, Wittmann F, Andrade A, Laurance WF, Laurance SGW, Marimon BS, Marimon B-H Jr, Al E (2013) Hyperdominance in the amazonian tree flora. Science 342:1243092–1–9

Tilloy A, Mallet V, Poulet D, Pesin C, Brocheton F (2013) BLUE-based NO <inf>2</inf> data assimilation at urban scale. J Geophys Res Atmos 118(4):2031–2040

Toro MV, Cremades LV, Calbó J (2006) Relationship between VOC and NOxemissions and chemical production of tropospheric ozone in the Aburrá Valley (Colombia). Chemosphere 65(5):881–888

Toro GMV, Cremades OLV, Ramirez BJJ (2001) Inventario de emisiones biogenicas en el valle de Aburrà. Revista Ingenierí,a y Gestión Ambiental, 17(32)

Tr'emolet Y (2006) Accounting for an imperfect model in 4d-var. Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography 132(621):2483–2504

Trémolet Y. (2007) Incremental 4d-var convergence study. Tellus A: Dynamic Meteorology and Oceanography 59(5):706–718

Uglietti C, Gabrielli P, Cooke CA, Vallelonga P, Thompson LG (2015) Widespread pollution of the south american atmosphere predates the industrial revolution by 240 y. Proceedings of the National Academy of Sciences of the United States of America 112(8):2349–2354. https://doi.org/10.1073/pnas.1421119112

Ulke AG, Andrade MF (2001) Modeling urban air pollution in Sao Paulo, Brazil: sensitivity of model predicted concentrations to different turbulence parameterizations. Atmos Environ 35(10):747–1763

Van Dingenen R, Dentener FJ, Raes F, Krol MC, Emberson L, Cofala J (2009) The global impact of ozone on agricultural crop yields under current and future air quality legislation. Atmos Environ 43(3):604–618

Van Nunen E, Vermeulen R, Tsai MY, Probst-Hensch N, Ineichen A, Davey M, Imboden M, Ducret-Stich R, Naccarati A, Raffaele D, Ranzi A, Ivaldi C, Galassi C, Nieuwenhuijsen M, Curto A, Donaire-Gonzalez D, Cirach M, Chatzi L, Kampouri M, Vlaanderen J, Meliefste K, Buijtenhuijs D, Brunekreef B, Morley D, Vineis P, Gulliver J, Hoek G (2017) Land use regression models for ultrafine particles in six European areas. Environ Sci Tech 51(6):3336–3345

van Velzen N, Segers AJ (2010) A problem-solving environment for data assimilation in air quality modelling. Environ Model Softw 25(3):277–288

Vivanco MG, Andrade MDF (2006) Validation of the emission inventory in the Sao Paulo Metropolitan Area of Brazil, based on ambient concentrations ratios of CO, NMOG and NOx and on a photochemical model. Atmos Environ 40(7):1189–1198

Wang X, Hamill TM, Whitaker JS, Bishop CH (2007) A comparison of hybrid ensemble transform kalman filter–optimum interpolation and ensemble square root filter analysis schemes. Mon Weather Rev 135(3):1055–1076

Wang X, Mallet V, Berroir JP, Herlin I (2011) Assimilation of OMI NO2 retrievals into a regional chemistry-transport model for improving air quality forecasts over Europe. Atmos Environ 45(2):485–492

Wang N, Lyu X, Deng X, Huang X, Jiang F, Ding A (2019) Aggravating O3 pollution due to NOx emission control in eastern China. Science of the Total Environment 677:732–744. https://doi.org/10.1016/j.scitotenv.2019.04.388

Weichenthal S, Van Ryswyk K, Goldstein A, Shekarrizfard M, Hatzopoulou M (2016) Characterizing the spatial distribution of ambient ultrafine particles in Toronto, Canada: a land use regression model. Environ Pollut 208:241–248

WHO (2016) Ambient air pollution: a global assessment of exposure and burden of disease. Technical report, World Health Organization

Worden HM, Bloom AA, Worden JR, Jiang Z, Marais E, Stavrakou T, Gaubert B, Lacey F (2019) New constraints on biogenic emissions using satellite-based estimates of carbon monoxide fluxes. Atmos Chem Phys Discuss 30:1–19

Wu H, Reis S, Lin C, Heal MR (2017) Effect of monitoring network design on land use regression models for estimating residential NO2 concentration. Atmos Environ 149:24–33

Yin J, Zhan X, Zheng Y, Hain CR, Liu J, Fang L (2015) Optimal ensemble size of ensemble kalman filter in sequential soil moisture data assimilation. Geophys Res Lett 42(16):6710–6715

Yim SHL, Gu Y, Shapiro M, Stephens B (2019) Air quality and acid deposition impacts of local emissions and transboundary air pollution in Japan and South Korea. Atmospheric Chemistry and Physics Discussions. 1–23. https://doi.org/10.5194/acp-2019-175

Zhang XL, Li QB, Su GF, Yuan MQ (2015) Ensemble-based simultaneous emission estimates and improved forecast of radioactive pollution from nuclear power plant accidents: Application to ETEX tracer experiment. J Environ Radioact 142:78–86

Zhang R, Cohan A, Pour Biazar A, Cohan DS (2017) Source apportionment of biogenic contributions to ozone formation over the United States. Atmospheric Environment 164:8–19. https://doi.org/10.1016/j.atmosenv.2017.05.044

Zhang R, White AT, Pour B, Arastoo M, Richard T, Cohan DS (2018) Incorporating GOES Satellite Photosynthetically Active Radiation (PAR) Retrievals to Improve Biogenic Emission Estimates in Texas. Journal of Geophysical Research: Atmospheres 123(2):1309–1324. https://doi.org/10.1002/2017JD026792

Zhu Q, Laughner JL, Cohen RC (2019) Lightning NO2 simulation over the contiguous US and its effects on satellite NO2 retrievals. Atmospheric Chemistry and Physics 19(20):13067–13078. https://doi.org/10.5194/acp-19-13067-2019