Validation and optimization of the ATMO-Street air quality model chain by means of a large-scale citizen-science dataset

Hooyberghs, H.; De Craemer, S.; Lefebvre, W.; Vranckx, S.; Maiheu, B.; Trimpeneers, E.; Vanpoucke, C.; Janssen, S.; Meysman, F. J.R.; Fierens, F.

DOI
10.1016/j.atmosenv.2022.118946

Publication date
2022

Document Version
Final published version

Published in
Atmospheric Environment

Citation (APA)
Hooyberghs, H., De Craemer, S., Lefebvre, W., Vranckx, S., Maiheu, B., Trimpeneers, E., Vanpoucke, C., Janssen, S., Meysman, F. J. R., & Fierens, F. (2022). Validation and optimization of the ATMO-Street air quality model chain by means of a large-scale citizen-science dataset. *Atmospheric Environment*, *272*, Article 118946. https://doi.org/10.1016/j.atmosenv.2022.118946

Atmospheric Environment 272 (2022) 118946

Contents lists available at ScienceDirect

Atmospheric Environment

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

Validation and optimization of the ATMO-Street air quality model chain by means of a large-scale citizen-science dataset

H. Hooyberghs [a,*], S. De Craemer [a,b], W. Lefebvre [a], S. Vranckx [a], B. Maiheu [a], E. Trimpeneers [c,d], C. Vanpoucke [c,d], S. Janssen [a], F.J.R. Meysman [b,e], F. Fierens [c,d]

[a] Unit Environmental Modelling, VITO, Boeretang 200, 2400, Mol, Belgium
[b] Department of Biology, University of Antwerp, Universiteitsplein 1, 2610, Wilrijk, (Antwerpen), Belgium
[c] VMM Vlaamse Milieumaatschappij, Kronenburgstraat 45, 2000, Antwerpen, Belgium
[d] Belgian Interregional Environment Agency, Gaucheretstraat 92-94, 1030, Brussels, Belgium
[e] Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629, HZ Delft, the Netherlands

HIGHLIGHTS

- Street-level air quality models can substantially benefit from a validation using a one-off widespread spatial monitoring campaign.
- In the case of NO_2, such widespread spatial data collection is possible through mass-scale citizen science using low-cost passive samplers.
- The availability of the extensive spatial dataset enables a "deep validation", which can result in substantially improved model skill.
- The optimized ATMO-Street model chain passes the FAIRMODE model quality threshold, substantiating its suitability for policy support.

ARTICLE INFO

Keywords:
Air pollution
Spatial variation
Dispersion modelling
Street level modelling
Citizen science
Model validation
Model optimization
FAIRMODE Model Quality Objective
Semi-variogram analysis

ABSTRACT

Detailed validation of air quality models is essential, but remains challenging, due to a lack of suitable high-resolution measurement datasets. This is particularly true for pollutants with short-scale spatial variations, such as nitrogen dioxide (NO_2). While street-level air quality model chains can predict concentration gradients at high spatial resolution, measurement campaigns lack the coverage and spatial density required to validate these gradients. Citizen science offers a tool to collect large-scale datasets, but it remains unclear to what extent such data can truly increase model performance. Here we use the passive sampler dataset collected within the large-scale citizen science campaign CurieuzeNeuzen to assess the integrated ATMO-Street street-level air quality model chain. The extensiveness of the dataset (20.000 sampling locations across the densely populated region Flanders, ~1.5 data points per km^2) allowed an in-depth model validation and optimization. We illustrate generic techniques and methods to assess and improve street-level air quality models, and show that considerable model improvement can be achieved, in particular with respect to the correct representation of the small-scale spatial variability of the NO_2-concentrations. After model optimization, the model skill of the ATMO-Street chain significantly increased, passing the FAIRMODE model quality threshold, and thus substantiating its suitability for policy support. More generally, our results reveal how a "deep validation" based on extensive spatial data can substantially improve model performance, thus demonstrating how air quality modelling can benefit from one-off large-scale monitoring campaigns.

1. Introduction

Air pollution remains a key environmental problem in most European cities (WHO, 2016; EEA, 2019), and so an accurate assessment of air pollution patterns and abatement strategies is vitally important to reduce the impact on human health. Many of the associated policy questions are addressed using air quality models: models have been successfully applied to interpolate pollution levels in between measurement locations (Thunis et al., 2016), estimate the population exposure on regional and urban scales (Jerrett et al., 2005; Hoek, 2017;

* Corresponding author.
E-mail address: hans.hooyberghs@vito.be (H. Hooyberghs).

https://doi.org/10.1016/j.atmosenv.2022.118946
Received 7 September 2021; Received in revised form 24 December 2021; Accepted 5 January 2022
Available online 8 January 2022

Xie et al., 2017), and quantify the health impact related to long-term exposure (Hoek et al., 2013; Faustini et al., 2014; EEA, 2019). Additionally, air quality models are essential tools to develop and evaluate policy scenarios (Miranda et al., 2015; Brusselen et al., 2016; Thunis et al., 2016).

Nitrogen dioxide (NO_2) is one of the important air pollutants in urban environments. More than 90% of the urban population in the EU is exposed to concentrations that exceed the guidelines put forward by the World Health Organization (WHO), leading to approximately 70.000 premature deaths every year (EEA, 2019). When quantifying the population exposure and health impacts of NO_2, a particular challenge is the spatial heterogeneity of the concentration field. Because of street-canyon effects and the proximity to main emission sources, the NO_2-concentrations vary strongly over short distances (Marshall et al., 2008; Cyrys et al., 2012; Lefebvre et al., 2013b; Jensen et al., 2017). To attain suitable model skill, air quality models should adequality capture this short-scale spatial variation, and reliably predict the concentration field on a scale of tens of meters.

Validation (and subsequent model improvement) are essential when models are used for regulatory purposes. Model simulated pollution maps need to be validated at the proper spatial and temporal scales. Street level models that target prediction of within-street variation of NO_2 at high spatial resolution, should hence be validated using measurement campaigns that have a suitably dense sampling grid. Measurements in streets with different traffic loads are required to capture the small-scale spatial variability of NO_2-concentrations. Because of logistical and financial constraints, such a high sampling density cannot be obtained using official telemetric stations (Vardoulakis et al., 2011). As an alternative, wind tunnel experiments have been used (Ketzel et al., 2000; Baker and Hargreaves, 2001). Although these validation campaigns provide an opportunity to validate air quality models in a controlled environment (e.g. controlled boundary conditions) (Vardoulakis et al., 2003), one of the main challenges in field campaigns is to handle all the variability in boundary conditions and the way long-term averages are achieved.

Mass-scale citizen science offers an innovative way to generate the large datasets required for such a validation campaign (Irwin, 2018; Van Brussel and Huyse, 2019; De Craemer et al., 2020a; Meysman et al., 2022; Bo et al., 2020), but it is presently unclear to what extent such datasets can truly generate improved model performance. There is an important trade-off in this respect. While citizen science has the advantage of generating data at high spatial resolution, one typically uses passive sampler measurements, and so the resulting data is generally less accurate and of lower quality than those collected via official telemetric stations. Citizen science has clear benefits in terms of raising awareness about air pollution (Van Brussel and Huyse, 2019), but to what extent can the resulting high-resolution data truly support the improvement of air quality models?

To address this question, we validate and optimize the ATMO-Street model chain (Lefebvre et al., 2013b) using the extensive NO_2-dataset collected within the CurieuzeNeuzen citizen science project (https://2018.curieuzeneuzen.be/). While this article makes a case study of one particular model chain, many of our findings, methods and techniques are readily and generically applicable to other (street-level) models, and so the conclusions are highly relevant for air quality models in general.

ATMO-Street is an integrated model chain (Lefebvre et al., 2013b) that models air quality at high, street-level resolution (i.e. 10 m), and hence representative for the class of high-resolution, state-of-the-art models that is operated by Environment Agencies across the world for planning and policy purposes. ATMO-Street is used by the Flanders Environment Agency (Vlaamse Milieumaatschappij, VMM) to assess the air pollution at the street level scale for Flanders, a densely populated region in Northwestern Europe (13,522 km^2, 485 inhabitants km^{-2}; total population 6,552,000). In addition, ATMO-Street is the default tool used for planning purposes, evaluating the impact of regional and local air quality plans and health impact assessments in Flanders. The model chain has been previously validated via several dedicated measurement campaigns, focusing both on spatial patterns and time series (Lefebvre et al. 2011, 2013b). However, these validation campaigns focused on a relatively small number of sampling locations, with at most a few dozens of locations distributed among a single urban region.

In 2018, the citizen science project *CurieuzeNeuzen Vlaanderen* project engaged 20.000 citizens across Flanders to measure NO_2 concentrations in front of their house using a low-cost sampler design (Meysman et al., 2022). This measurement campaign was internationally unprecedented in terms of coverage and spatial density: 20.000 sampling kits containing NO_2 diffusion samplers were distributed (~1% of all households in Flanders), thus allowing measurements across a wide urbanized region (~250 km × 50 km) at high spatial density (~1.5 sites on average per km^2). The resulting extensive dataset is used here for a detailed validation case study of the ATMO-Street model chain.

We develop a generic three-step methodology to validate and optimize the model chain by means of the CurieuzeNeuzen dataset. Firstly, the original ATMO-Street model chain is validated against the NO_2 data using validation plots and statistical techniques. The extensiveness of the measurement dataset allows us to perform an in-depth model performance analysis by evaluating the concentrations based on different aspects (type of location, concentration class etc.). In the second step, we introduce improvements and optimizations to the model chain based on the findings of the validation. The effectiveness of the optimization is again verified by validating the results of the optimized model chain against the NO_2 data. Finally, we evaluate the remaining discrepancies between the modeled and measured concentrations and provide an outlook for further improvement of air quality modelling. In this last step, we pay special attention to the ability of ATMO-Street to capture the short-scale spatial variation of the NO_2-concentrations.

2. Methods

2.1. Measurement dataset

The measurement campaign will be only briefly summarized here, and is discussed in more detail in (De Craemer et al., 2020b; Meysman et al., 2022). In the CurieuzeNeuzen Vlaanderen citizen science campaign, 20.000 sampler kits were distributed to individual citizens, schools, companies, social organizations and municipalities to measure outdoor NO_2 concentration at streetside locations. At the front of the house (facing the street), two passive NO_2 samplers of the Palmes diffusion tubes type were strapped to a real estate sign panel and attached to a window pane. This set-up standardized air turbulence conditions near the Palmes tubes across all sampling points. Measurements were conducted preferably on the first floor or otherwise ground floor to constrain the height effect on NO_2 concentrations.

A four-week measurement was performed from 11 a.m. April 28th to 1 p.m. May 26th, 2018. Note that the duration of the Palmes tubes campaigns should be limited to approximately one month, to avoid saturation of the tubes. The exact time period for the campaign has been chosen to maximize the legitimacy of the validation results: background concentrations in May closely resemble the annual mean background concentrations, and May is one of the few months without a long vacation period, eliminating the need for time-specific corrections to the traffic data. The meteorological conditions during the measurement period were, however, somewhat atypical. The average temperatures during May 2018 were significantly higher than during a typical month May in the climatological baseline considered by the National Meteorological Agency (1981–2010), with a profound gradient in the bias from the west (coastline) to the east of the region (bias of approximately 3 °C in the east, and approximately 1.5 °C at the coastline). Moreover, there has been much less precipitation (30% less on average), much smaller wind speeds, and also the prevailing wind direction was clearly different. During May 2018, the prevailing wind directions were north-

north-west and north-east, while on average winds from the southwest are dominant in Flanders.

Duplicate samplers showed good precision (root mean square error 1.7 μg/m^3 between replicates, relative standard deviation < 5%). These raw NO_2 data were calibrated by simultaneous deployment of passive samplers at 24 EPA reference monitoring stations dispersed across the measurement region, and averaged across the two duplicates, thus resulting in mean NO_2 concentration over the 4-week measurement period. After a quality control, 17886 measurement locations were retained for the model validation campaign (Meysman et al., 2022). Assuming errors are random and uncorrelated, the addition of the standard deviations of the passive sampler measurement (1.7 μg/m^3) and calibration (2.2 μg/m^3) resulted in a total standard deviation of 3.9 μg/m^3, thus providing a relative uncertainty of 10% at the WHO-guideline value of 40 μg/m^3.

2.2. The ATMO-Street model chain

2.2.1. General overview

Street-level nitrogen dioxide concentrations are modeled using a model chain that captures the different scales of urban air quality. The ATMO-Street model chain (Lefebvre et al., 2013b) consists of the land-use based interpolation model RIO determining background concentrations (Janssen et al., 2008a), the bi-gaussian plume dispersion model IFDM accounting for the impact of local emissions from traffic and industry (Lefebvre et al., 2011), and the street-canyon module OSPM that calculates the in-street increment resulting from street-canyon effects (Berkowicz et al., 1997). Road traffic emissions are computed by the traffic emission model FASTRACE (Veldeman et al., 2016). The model chain calculates hourly concentrations at a number of irregularly spaced receptors, which are subsequently gridded to a regular raster with a 10 m resolution. The flowchart of the model chain is provided in Fig. 1.

For verification purposes, the simulations by the full ATMO-Street model chain were compared to versions that use only part of the model chain. In one type of sensitivity analysis, only the background concentrations from the RIO-model were considered, thus evaluating the predictive capability of only using wide-scale land use regression. In another sensitivity analysis, we used the RIO-IFDM combination, which combines the background concentrations with local contributions from traffic and industry, but neglects the street-canyon increment. The remainder of this section explains the three model components and their coupling in more detail.

2.2.2. Components

FASTRACE is a traffic emission model that calculates geographically explicit emissions for road transport, based on (1) emission factors (i.e. emissions per vehicle type per speed per kilometer) (2) fleet data (i.e. number of vehicles and mileages), and (3) mobility data (i.e. vehicle counts on a network). Emission factors were obtained from region specific calculations with the COPERT-tool, which is EU-wide used to calculate emission inventories for road transport (Ntziachristos et al., 2009). FASTRACE calculates yearly total emissions for each road segment, which are subsequently combined with daily, weekly and monthly traffic intensity profiles, to obtain hourly emissions for each road segment.

Background concentrations are modeled using RIO (Hooyberghs et al., 2006; Janssen et al., 2008b), a land use regression model for the interpolation of hourly pollutant concentrations as measured by the official telemetric network. The model is based on a residual kriging interpolation scheme using a land use derived covariate. A polynomial regression determines the statistical relationship (trend functions) between the long-term averaged concentrations at each hour of the day and the underlying land use parameter. RIO produces hourly concentration maps for NO_2, NO, and O_3 on a 4 × 4 km^2 grid, which are subsequently used as background concentrations for the IFDM and OSPM components of ATMO-Street chain.

Local open-street concentrations due to traffic emissions and industrial point sources are modeled by the bi-Gaussian plume model IFDM (Immission Frequency Distribution Model) (Lefebvre et al., 2013a). IFDM is a receptor grid model: air pollutant concentrations are computed for an abundance of receptor locations. Instead of a regular grid, we use a point source and road-following grid. This approach ensures that more receptor points are available where the largest concentration gradients are expected (Lefebvre et al., 2011). Since the model uses an hourly time resolution, we assume that the chemical equilibrium in the NO_x-O_3 reaction is reached. We take this chemical reaction into account using the fast-ozone-chemistry scheme (Berkowicz et al. 1997, 2008), which relies on temperature and solar height data. To avoid double-counting of the emission sources, a specific coupling between the regional model and the urban-scale model has been developed (Lefebvre et al., 2011).

To calculate the effect of buildings on the street level concentrations, the IFDM model is coupled to the Operational Street Pollution Model (OSPM) (Berkowicz et al., 1997; Ottosen et al., 2015; Jensen et al., 2017). Street level concentrations due to road traffic emissions are calculated using a combination of a plume model for the direct contribution and a box model for the recirculating part of the pollutants in the street. In the current set-up for OSPM, a receptor location is placed every 20 m on each road with a row of buildings adjacent to the road (i.e. at a maximum distance of 50 m to the middle of the road). The concentrations at the receptor locations of the IFDM and OSPM models are

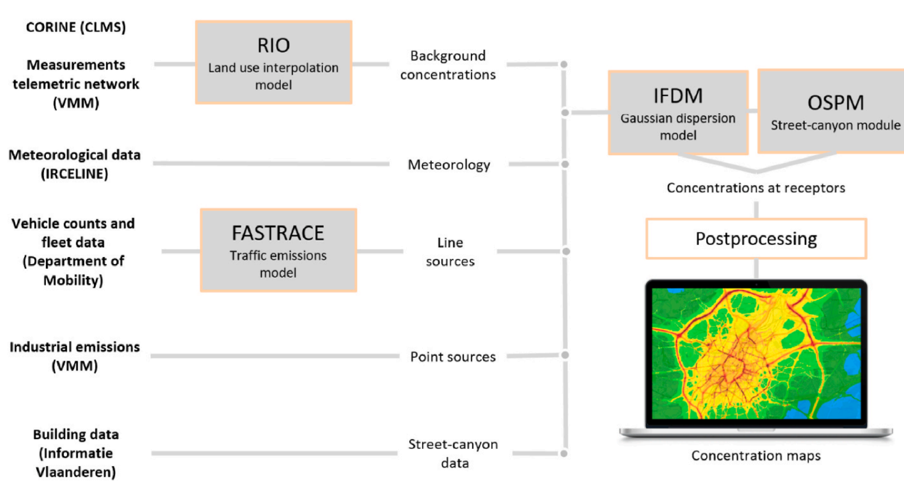

Fig. 1. Flowchart of ATMO-Street model.

eventually combined and gridded via a three-step postprocessing module. At first, IFDM results are gridded using Delaunay triangulation to obtain gridded open street concentrations. Secondly, we grid the OSPM results using nearest-neighbour interpolation. In the final step, both gridded maps are combined into a map with a 10 m resolution, by using the OSPM results at locations where buildings are adjacent to the road, and the IFDM results at all other locations.

A priori, we expect large deviations between the measurements and the modeled data for the background model RIO. Because of the coarse resolution, there will be a lot of scatter, a large underestimation of the results (especially close to busy roads) and not much correlation between the measurements and model values. Adding the Gaussian dispersion model IFDM should improve the results, especially for open locations, which should also significantly improve the scatter and the correlation. However, the RIO-IFDM model chain neglects the recirculation of pollution at locations with buildings adjacent to the road, hence a large bias is still expected. Adding the OSPM module should resolve this issue, but it also increases the susceptibility of the model chain to input errors. Because the concentration field at locations with recirculation is very sensitive to many parameters describing the setting (traffic emissions, vehicle speed and numbers influencing the traffic-induced turbulence, detailed building configuration in the immediate surroundings of the location) and many of these parameters are only approximatively known, we expect a lot of scatter for the locations where the OSPM model is applied.

2.2.3. Set-up for the validation campaign

In this study, the ATMO-Street model chain was applied to the same 4-week period as the citizen science measurement campaign. Input data stem from official datasets of the regional authorities. The regional background model RIO has been set up using the data from the telemetric network of the Flanders Environment Agency (VMM) and the Corine Land Cover of the Copernicus Land Monitoring Service (CLMS) as land-use input. Vehicle fleet and traffic data for the major roads in Flanders are provided by the Flemish Department for Mobility. Minor roads are only sparsely represented in these traffic data, and these roads are thus not considered in the present air quality assessment. There are also some known issues with the traffic data for urban locations, as recent mobility plans (e.g. low-traffic zones in city centers) are not always correctly represented. Point sources stem from the official emission inventory for industry of VMM. Building data has been retrieved from the official building dataset for Flanders (Informatie Vlaanderen).

The Gaussian dispersion model internally computes stability classes based on the Bultynck-Malet parametrization (Bultynck and Malet, 1972), and thus only requires surface temperature, wind speed and wind direction as meteorological input. These parameters have been composed by the Belgian Interregional Environment Agency (IRCEL - CELINE) by assimilating Copernicus C3S ERA5 reanalysis data (Copernicus Climate Change Service, 2017) with measurements at several meteorological stations, yielding surface wind and temperature fields with a 1 km resolution. Because the model uses input data for the actual time frame of the measurements, we do not expect an influence of the atypical meteorological conditions during the measurement period on the final conclusions of the study.

The coordinates of the measurement locations were recorded with a precision of 2 m. The corresponding model concentrations are determined by the concentration for the pixel (10 m × 10 m) in the gridded map that contains the measurement location. Note that in this way, the coordinates of the measurement locations are not used when defining the receptor grid. Model results are always reported at 1.5 m height, even in cases where the measurements were done at higher locations. Differences between the measurements and the model result at their location in this manuscript thus include uncertainties in the measurements, errors in the model input data, model errors and errors from the postprocessing.

3. Model optimization

3.1. Initial validation

Firstly, we focus on the validation statistics of the original setup of the ATMO-Street model chain before optimization (the so-called "original model"). Table 1 provides the validation statistics obtained by comparing the CurieuzeNeuzen data with the model values (see the appendix for a mathematical definition of the statistical quantities). The Pearson correlation coefficient (0.58) points at a reasonable correlation between the measurements and the model results, and is in line with results obtained in previous validation campaigns (Lefebvre et al., 2013b). The bias of the original model is substantial and negative (−4.1 μg/m^3, −20%). This indicates that the model underestimates the NO_2 concentrations in general, which is also reflected by the shift in histograms (see Fig. 2). Especially for the lower concentrations (<25 μg/m^3), the modeled distribution is shifted to lower concentrations in comparison with the measured contribution. The other statistics, such as the Bias Corrected Root Mean Square Error (BCRMSE 4.6 μg/m^3) and fraction of model values within a factor of two of the observation (Fac2: 99%) are more in line with the results of previous validation studies.

Additional insight into the discrepancy between modeled and measured concentrations is obtained by considering the bias and RMSE per concentration class. For this purpose, we grouped the locations in ten classes according to the deciles for the measured concentrations (Fig. 3). Apart from the 10th decile, the original model shows the highest relative biases (up to −22%) for the lower deciles. Similarly, the (relative) RMSE is larger for the third to sixth decile than for the seventh to ninth decile. As locations with higher-than-average concentrations are the most sensitive to issues with the Gaussian dispersion model or the street canyon module (or one of their input datasets), which would introduce larger (relative) deviations for the higher deciles, these findings therefore point at issues with the background model, which underestimates the background concentrations for the lower deciles. If we plot the bias across the spatial domain (Fig. 4), we indeed observe that underestimations are mainly occurring in rural locations (i.e. in the less densely populated areas), whereas the bias is much smaller for the urban locations. Finally, this underestimation of the background concentrations is also observed in the histogram (Fig. 2).

An in-depth analysis of the deviations between models and measurements uncovered a second issue with the original model, which concerned to the coupling of the Gaussian dispersion model (IFDM) and the street-canyon module (OSPM). When gridding the final concentration maps (i.e. during the postprocessing as discussed in Section 2.2), the grid cells with street-canyon module increments do not always correspond to the street side location where citizens put their diffusive

Table 1
Validation statistics for the original and the optimized ATMO-Street model chain. The validation statistics for the separate buildings blocks of the optimized model chain (RIO and RIO-IFDM) are also provided. Statistics include the bias, the Root Mean Square Error (RMSE), the bias-corrected RMSE (BCRMSE), the Pearson R^2 coefficient, the FAIRMODE model quality indicator (MQI) and the fraction of model values within a factor of two of observations (Fac2). A mathematical definition of the statistical quantities is provided in the Appendix.

Statistic	ATMO-Street Original model	ATMO-Street Optimized model	RIO-IFDM Optimized model	RIO Optimized model
Bias (μg/m^3)	−4.1	−2.7	−4.3	−4.5
RMSE (μg/m^3)	5.5	5.2	6.2	7.0
BCRMSE (μg/m^3)	4.6	4.4	4.4	5.3
Pearson R^2	0.58	0.58	0.51	0.33
MQI	0.96	0.80		
Fac2 (%)	99.0	99.4	98.0	96.2

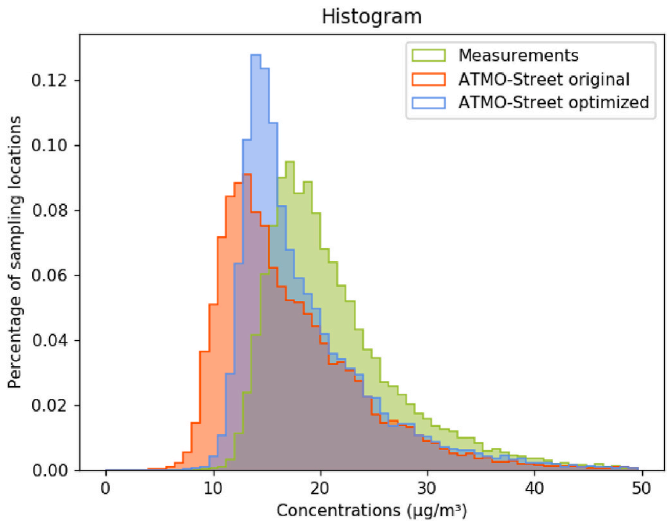

Fig. 2. Comparison between measured and modeled NO_2 concentrations at the 17.886 measurement locations with high quality data. Histograms for the measurement data (green), the updated model chain (blue) and the original model chain (red). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

indicated an overestimation at these locations in the original model.

After a detailed analysis, relying on land use data for the Curieuzeneuzen sampling locations, we found that the problem was linked to the land-use parameterization that is used in the RIO module. In this module, NO_2 data are assimilated from reference stations of the environmental monitoring agencies across the whole of Belgium. Yet, there is a strong north-south difference in urbanization in Belgium, which makes that there are relatively few reference stations in rural areas of Flanders (the northern part of Belgium). As a result, the land use parameterization applied in the original RIO module was heavily influenced by the observations at EPA reference stations in rural areas of Wallonia (the southern part of Belgium). Since background NO_2 values are lower in Wallonia (which is less densely populated and less industrialized), this caused an underestimation of background concentrations in Flanders, which was uncovered for the first time thanks to the Curieuzeneuzen sampler data.

Guided by the citizen science data, the RIO module was adapted by improving the parameterization of the different rural land use classes, yielding an optimized relation between the concentrations and the land use parameters (trend function). These optimizations principally consisted of a decoupling of the rural land uses classes (namely forests, natural areas and arable land) in the northern urbanized part and southern non-urbanized part of Belgium. Note that the finetuning required the availability of an abundance of measurement data at many

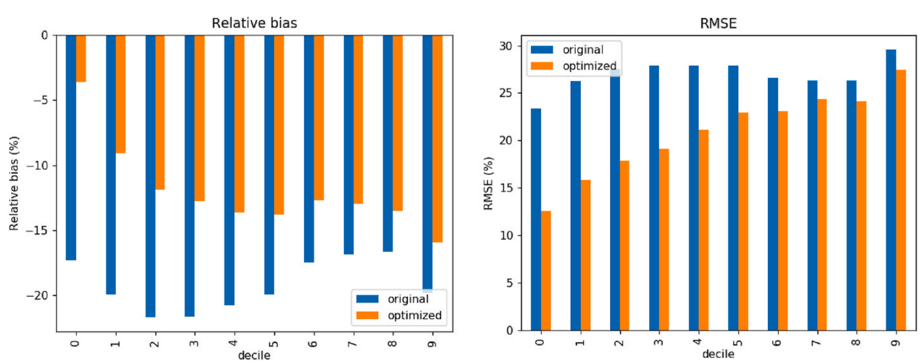

Fig. 3. Comparison between measured and modeled NO_2 concentrations per concentration class (10 deciles of the measured concentrations – decile 0 contains lowest concentrations). Relative bias (left) and RMSE (right) per decile. The panels provide the results for the original model chain (blue) and the updated model chain (orange). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

samplers. This was especially apparent for streets that are diagonal with respect to the north-south axis, as illustrated in Fig. 5. These locations are therefore incorrectly assigned the results of the Gaussian dispersion model, instead of the results of the street-canyon module. Although this situation only occurred at a limited number of locations, large underestimations were obtained at these locations, which hence significantly influence the bias for the largest decile in Fig. 3.

Note that the two issues discussed above (background underestimation, incorrect street canyon postprocessing assignment) could only be detected due to the extensiveness of the Curieuzeneuzen dataset. While ATMO-Street has previously been validated with smaller datasets, this analysis has been unable to reveal the background concentration issue, as a large and spatially widespread dataset is required for the type of analysis presented in Figs. 3 and 4. Additionally, the postprocessing issue was only observed for a small number of locations, and so the issue can only surface in suitably large datasets (the probability to include such locations in a dataset increases with the sampling size).

3.2. Model optimization

Guided by the results of the initial validation, model optimizations were implemented. A first improvement targeted the rural background concentrations, as the results of the Curieuzeneuzen campaign clearly rural locations with different land uses in their surroundings, and the Curieuzeneuzen measurements have thus been indispensable.

The citizen science data were only used to determine an improved land use parameterization, but they are not directly used in the spatial interpolation itself. The model results hence remain independent of the measurements, and thus independent validation of the optimized model using the citizen science data is still possible.

A second correction targets the incorrect street canyon postprocessing assignment, by adjusting the GIS-tools that determine the locations where the street-canyon concentrations are applied. To this end, we modified the parameter that determines the maximal extent of the street canyon concentrations (expressed as the distance to the middle of the road), to make sure the OSPM results are used for all locations where the concentration is significantly influenced by the presence of the buildings. Moreover, in the optimized model version, the concentrations of the OSPM street-canyon module are now also used for half-open locations, i.e., for roads with a continuous row of houses at one side of the street.[1] Due to these two modifications, the (higher) street-canyon

[1] Because the size of the recirculation vortex is only dependent on the *upwind* building in the OSPM model, the model can also be used for locations with a continuous row of buildings at one side of the road.

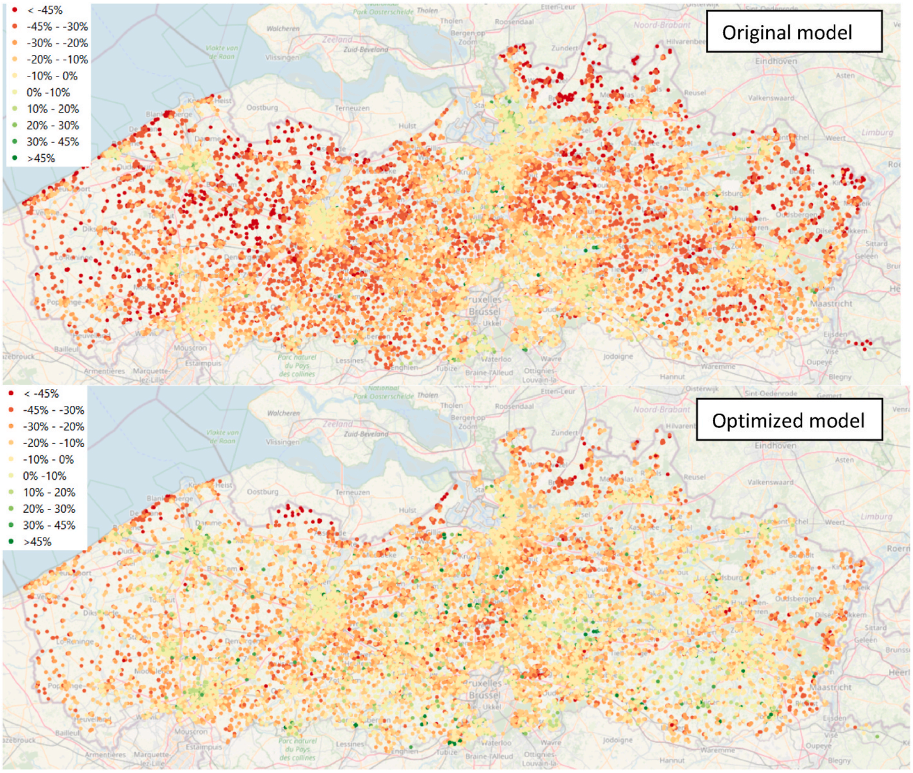

Fig. 4. Map of the relative difference between the measured and the modeled concentrations (in %) across the region of Flanders (17.886 measurement locations with high quality data). Negative (red) values indicate model underestimation, positive (green) values signify model overestimation. The top panel shows results for the original model, the bottom panel for the optimized model. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

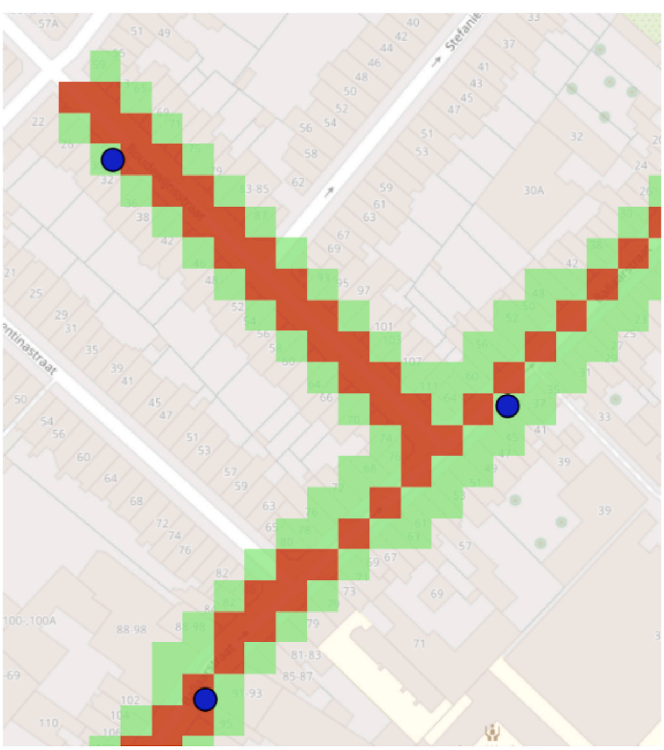

Fig. 5. Illustration of the issue concerning the coupling between the Gaussian dispersion and the streetcanyon model. In the original model chain, the streetcanyon concentrations are only used for the red colored grid cells. For some of the sampling locations (blue dots), the Gaussian dispersion results are hence applied. In the optimized model, the streetcanyon contribution is also used for the green grid cells, and hence for all three sampling locations in the domain of the figure. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

contributions are attributed to more sampling locations, leading to an increase in the mean NO_2 concentration across all sampling locations.

3.3. Analysis of the optimizations

3.3.1. Basic analysis

Table 1 provides the validation statistics for the optimized model. Fig. 2 shows the histogram and Fig. 3 depicts the bias and RMSE per decile. The optimized model outperforms the original model in many aspects. The bias and RMSE are markedly lower for the optimized model. The largest improvements in the bias and RMSE are observed for the lower deciles, as shown by Fig. 3, and as expected because of the nature of the optimizations. Moreover, also the Fac2 increases from 99% to 99.4%, which implies that the number of sampling sites for which the modeled data deviates more than a factor of two of the observations decreases with 40% from 1% to 0.6%. The relative difference map of the original and the optimized model (Fig. 4) highlights moreover the reduced (relative) bias in rural locations. For many of the locations in the rural areas, the relative difference between the modeled and measured data is reduced to less than 10%.

To facilitate the interpretation of these results, we compare the statistics of the validation reported in this Paper (based on more error-prone citizen science data) with those of more traditional studies (with more controlled measurements), for ATMO-Street and similar street-level models used for policy support in Europe. Typically, the bias is somewhat higher for the study at hand, compared to a bias of −0.7% for the ADMS-Urban map in London (Hood et al., 2018) and 2% for the DEHM/UBM/OSPM map in Copenhagen (Jensen et al., 2017). On the other hand, the correlation, RMSE and Fac2 are more in line with those found in the traditional campaigns (e.g. correlation 0.6 in Aarhus and 0.7 in Copenhagen (Jensen et al., 2017), correlation 0.7 in London (Hood et al., 2018), and RMSE 6 $\mu g/m^3$ in Antwerp (Lefebvre et al., 2013b)). A detailed comparison is, however, complicated, as both the set-up of the measurement campaigns and the model chains vary significantly among the studies.

The optimization does, however, not remove all differences between measurements and modeled concentrations. The histogram (Fig. 2) indicates that the underestimation for many locations with low-to-middle concentrations is reduced, but has not completely disappeared, and that the distribution of the model values for the optimized model still deviates from the distribution for the measured values. In addition, also, the Pearson R^2 is the same for the optimized and the original model (see Table 1), indicating that the correlation between the measured and modeled data does not improve.

3.3.2. MQI

As an additional benchmark for the model quality, we focus on the model quality index (MQI) as proposed by the Forum for Air Quality Modelling in Europe (FAIRMODE). This indicator describes the discrepancy between measurements and modelling results linked to the RMSE (Thunis et al., 2016; Pisoni et al., 2019; Janssen et al., 2020). The MQI is a quality indicator that is specifically designed to assess the performance of a model as a policy support tool for official assessments and EU reporting. The FAIRMODE model quality objective (MQO) states that air quality models can be used for official assessment purposes if the MQI is less or equal to one.

The original ATMO-Street model just meets the MQO objective, as the MQI is equal to 0.96 (Table 1). The ensuing model optimization however decreased the RMSE and bias, which resulted in a substantial decrease of the MQI to 0.80, and the optimized model thus satisfies the objective with a far greater margin, indicating that the optimized model is better suited for policy support.

3.3.3. Spatial variation

Street-level air quality models are designed to simulate street-level concentration fields with high spatial resolution. We use a semi-variogram analysis to test how well different models represent the spatial heterogeneity of the concentration field. The semi-variogram visualizes the degree of spatial variation of a set of observations by quantifying the differences between observations at a given distance through the semi-variance (Cressie, 1992)

$$\gamma(h) = \frac{1}{2N(h)} \sum_{N(h)} (c_i - c_j)^2$$

Here, the sum is over all pairs of locations that are a certain distance h apart, c_i and c_j are the concentrations at these two locations, and $N(h)$ is the number of pairs that are considered. The semi-variance $\gamma(h)$ quantifies the difference between two observations separated by a distance h, with larger values indicating larger spatial variations. The technique is only effective if the (spatial) resolution of the sampling dataset is in line with the actual spatial scale of the gradients in the concentration field. An application for NO_2 concentrations thus requires a dataset with a dense sampling, like the citizen science dataset analyzed here.

Fig. 6compares the semi-variogram for the citizen science data to the ATMO-Street model results, for both the original and the optimized model chain. As NO_2 tends to vary over short spatial scales, we only consider measurement locations that are less than 2 km apart. The figure highlights how the short scale spatial variations are clearly better represented by the optimized model. The underlying reason is the optimization in street canyon postprocessing (i.e., improvement of the coupling between the Gaussian dispersion and the street canyon module), which yields a much better representation of the spatial gradients on a local scale. The analysis shows how the optimized model chain adequately captures the short-scale spatial variation of the concentration field, whereas the original model underestimates the spatial variation. Although over- and underestimations of the spatial heterogeneity can still occur at specific locations, the mean heterogeneity of the NO_2-concentrations in Flanders is trustworthily explained by the optimized model chain.

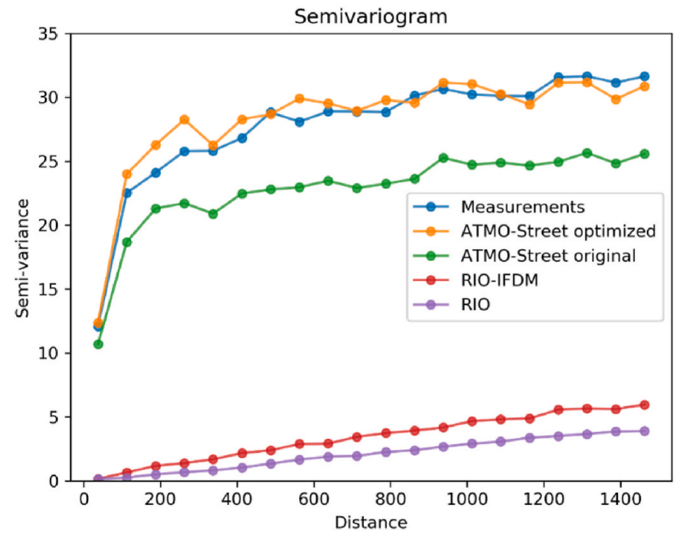

Fig. 6. Semi-variogram. The blue line indicates the semi-variance for the measurements of NO_2, while the other lines provide model results for NO_2 for three different types of models. The purple line provides results for the optimized RIO, the red line for the optimized RIO-IFDM and the orange line for the optimized ATMO-Street chain. The green line shows the results for the original ATMO-Street chain. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.3.4. Conclusions regarding the optimization

In summary, we conclude that the optimization resulted in substantial improvement of model performance, as substantiated by increased validation statistics, an improved MQI and a better representation of the short-scale spatial variations. This extensive analysis was made possible by exploitation of the large-scale data of the citizen science campaign. Although the optimization procedure presented here is specific to the ATMO-Street model, the underlying methodology and resulting conclusions are of wider interest for the air quality modelling community. Our "in-depth" validation of the ATMO-Street model relies on a statistical analysis that is applicable for any large-scale model validation, whereas the techniques to improve the RIO-model are applicable to any land use regression (LUR) model. Moreover, the observation that model shortcomings remain hidden when validation is done with limited data and only revealed through suitable large spatial datasets, is particularly relevant to the whole field of air quality monitoring and modelling.

Although the optimizations greatly improve many aspects of the model chain, they do not remove all differences between measurements and modeled concentrations, as e.g. indicated by the unchanged correlation coefficient and the updated histogram. In the next sections, we elaborate further on the validation of the optimized model and focus on the remaining discrepancies between the modeled and the measured concentrations.

4. In-depth validation of the optimized model

4.1. Analysis of the submodels

Environmental agencies use a range of different air quality model types for policy purposes, with different spatial resolution. Some models are solely based on land use regression, while others include more computationally intense approaches that explicitly include point and line emissions and simulate the ensuing atmospheric dispersion of the emitted pollutants and/or account for street canyon effects. The three different model components of the ATMO-Street chain reflect this cumulative complexity and increasing spatial detail. To examine the importance of the different components, Fig. 7 and Table 1 compare the

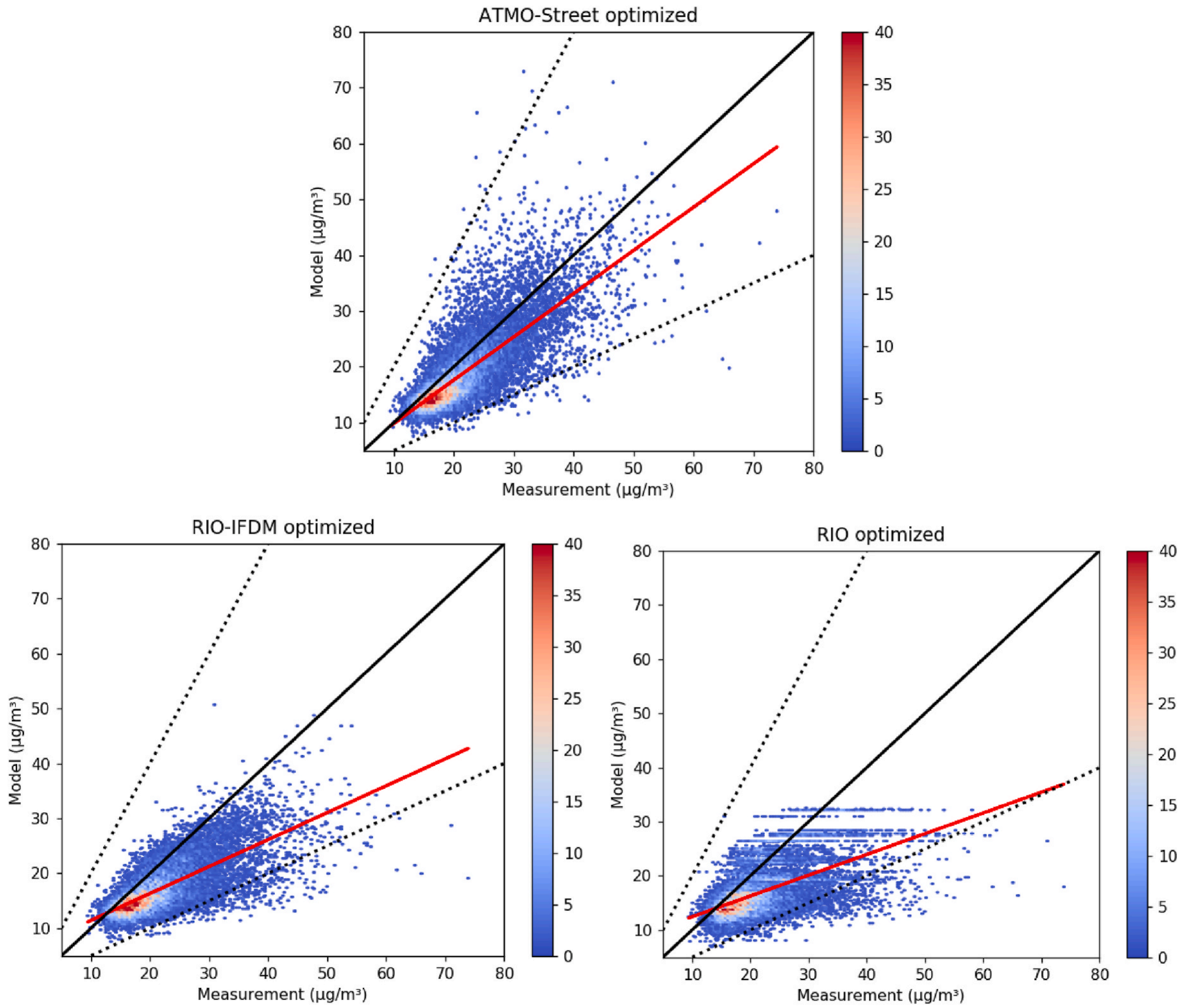

Fig. 7. Scatterplots showing the modeled concentration as a function of the measured concentrations for all sampling locations Different panels depict the full ATMO-Street model chain (top), the Gaussian dispersion model (RIO-IFDM, bottom left) and the background model only (RIO, bottom right). To improve the visibility, the scatter points have been binned per 0.5 µg/m^3. The colorscale indicates the number of points in each bin. The dashed lines indicate the upper and lower boundary of the interval [0.5 * measurements; 1.5* measurements].

validation statistics of the full optimized ATMO-Street chain with the background model only (RIO), and the combination of the background model with the Gaussian dispersion model (RIO-IFDM, i.e. ATMO-Street without street-canyon increments).

The background model only substantially underestimates the measured concentrations (Fig. 7). As substantiated by the linear regression coefficient and the scatterplot, the highest concentrations are particularly underestimated. The RIO model provides background concentrations on a 4 by 4 km resolution, and the highest roadside peaks in traffic dense streets are clearly missed. The addition of the Gaussian dispersion model IFDM considerably decreases the model-data discrepancy at these locations, and, consequently, the correlation and linear coefficient substantially improve. There is however still a significant bias, which is due to an underestimation of the street-canyon locations. Only the complete ATMO-Street model chain appropriately captures the recirculation of the pollution at these locations, yielding a much smaller bias.

The results for the RMSE, bias and correlation are in line with the expectations regarding the model components (see Section 2.2.2). The (absolute) bias and RMSE are large and the correlation low for the RIO-model. The RIO-IFDM model substantially improves on the RMSE and the correlation, but still has a large bias. When OSPM is added, the bias and the RMSE further decrease, but the improvement in RMSE is exclusively due to the decrease in bias, as indicated by the BCRMSE.

We conclude that only a model chain that takes the street-canyon increments explicitly into account manages to adequately assess the NO_2-concentrations. These findings emphasize the importance of the street-canyon contributions, and are in line with the results observed in other studies concerning modelling of air quality at street-level scale (Vardoulakis et al., 2003; Lefebvre et al., 2013b; Jensen et al., 2017).

We furthermore analyze the effect of the different submodels on the spatial variation of the modeled concentration field. In addition to the results of the full ATMO-Street chain, Fig. 6 also shows the spatial variation modeled by the RIO and RIO-IFDM submodels. Clearly, the RIO-background model largely underestimates the spatial variation observed in the citizen science data. This is not unsurprising given the coarse resolution of the RIO model (4 km × 4 km). The Gaussian dispersion model IFDM adds the open-street concentrations due to the road traffic and point sources, and as a result, the spatial variation increases. However, the semi-variance of the RIO-IFDM model still falls widely below that of the citizen science data. Adding the street canyon module OSPM greatly improves the representation of spatial variation: the spatial heterogeneity now closely approximates that observed by the measurements. The semi-variogram analysis thus demonstrates that street level air quality models like ATMO-Street are capable of capturing the general spatial heterogeneity of the NO_2 concentration field, if street canyon increments are included in the model chain.

4.2. Breakdown by location type

To gain some further insight in the remaining discrepancies between the modeled concentrations and the measurements, and the impact of the input data on these, we analyze the validation for some specific location types.

First, we group the sampling locations based on the model that is applied at the sampler location. We divide all the locations in two groups: locations where RIO-IFDM is applied (labeled 'IFDM'), or locations where RIO-IFDM-OSPM is applied (labeled 'OSPM'). The former set are typically locations with isolated buildings, whereas the second set consists of (more complex) locations with a row of buildings adjacent to a road. The validations statistics for the full ATMO-Street model chain at these two groups of locations are provided in Table 2. The results indicate a slightly lower bias for the locations where OSPM has been used, but also a much larger scatter (RMSE) and much lower correlation for these locations, as expected, because of the larger sensitivity to input errors for the OSPM locations (see 2.2.2).

Secondly, we split the sampling locations according to the availability of traffic data for the nearest road to the measurement location. The traffic dataset contains traffic flows for a limited number of streets (the major roads). The (absolute value of the) bias is lower for the locations for which traffic data is available (absolute bias −2.4 μg/m^3, relative bias −10%) compared to the locations without known traffic data (bias −2.9 μg/m^3, −15%) (see Table 2). Note that the mean measured concentration is higher for the locations for which traffic data is available (23.4 μg/m^3 versus 19.6 μg/m^3). As the relative bias increases with increasing concentration, we would expect the (relative) bias to be higher for the locations close to the roads. Since we observe the opposite, we definitely detect underestimations for sampling locations near roads where traffic data is lacking. Note, on the other hand, that the scatter is larger for the locations with traffic data (as quantified by a lower R^2 and larger BCRMSE). The underlying reason is the abundance of OSPM locations for the samplers in the vicinity of roads with traffic data (for 76% of the locations with traffic data OSPM has been used, while OSPM is not used for locations without traffic data). As the results in Table 2 indicate, the scatter is much larger for the OSPM locations, which is also reflected in a larger scatter for the locations with traffic data.

Finally, we make a comparison between cities in Flanders. We consider three groups of locations: Flanders' largest city Antwerp (500.000 inhabitants; 1002 samplers), Flanders' second largest city Ghent (250.000 inhabitants; 800 samplers), and samplers located in the other 8 largest cities (60.000–120.000 inhabitants; 2549 samplers). Validation statistics are provided in Table 2. The city of Ghent stands out from the other. This is because a new mobility plan has been introduced in 2017, which led to the introduction of new pedestrian streets, and modified traffic flows in the nearby streets, thus altering traffic flows within the historical city center. However, the available traffic data do not (yet) account for this new condition, and so the traffic data used for Ghent in the model set-up are less accurate than those for other cities. The validation statistics reflect these shortcomings in the traffic data. The BCRMSE in Ghent (22%) is higher than in Antwerp (16%) and the other cities (17%), while similarly, the correlation coefficient in Ghent (0.40) is lower than in Antwerp (0.52) and the other cities (0.46). When we only focus on the 200 samplers in the inner city of Ghent, where the largest impact of the new circulation plan is observed, the validation statistics become even worse. The correlation coefficient decreases to 0.27, and the relative BCRMSE increases to 25%.

4.3. Open issues

As the validation indicates, there are some remaining discrepancies between the modeled concentrations and the measurements, indicating some room for further improvement of the model and its input data.

An important issue concerns the quality of the mobility data that is used as input. Firstly, the traffic dataset only contains traffic flows for a limited number of streets. The validation substantiates that the NO_2-concentrations are, as expected, more adequately modeled for sampling locations near the roads that are included in the traffic data. Furthermore, the spatial pattern of the traffic data is outdated, which has an impact on the model quality for locations at which new mobility plans have recently been introduced (e.g Ghent). These findings clearly highlight the importance of up-to-data traffic data for air quality modelling at the local scale. Our analysis hence reveals that Environmental Protection Agencies should invest in the collection of traffic data, and keep these datasets also up to date, in order to support their air quality policies.

The statistics per city hint at a remaining issue with the optimized model, related to the urban background concentrations in Flanders' largest city, Antwerp. The bias in the largest city, Antwerp (−3.9 μg/m^3, −12%), is significantly larger than the bias in Ghent (−1.1 μg/m^3, −4%) and the other cities (−2.0 μg/m^3, −8%) (see Table 2). These results hint at a strong underestimation of the urban background concentration in Antwerp. Note, however, that previous validation studies have not observed the current underestimation: in a dedicated campaign focusing on Antwerp, a bias of −2 μg/m^3 has been observed (Lefebvre et al., 2013b), which is more in line with the bias observed in this work for the other urban locations. Therefore, the underlying reason of the issue is unclear, as it could either be related to the measurements (e.g. the calibration of the sampling results is mostly based on mid-range concentrations, whereas higher concentrations are mainly observed in Antwerp), or the model set-up (e.g. because the trend function of the land use regression model RIO may be unable to adequately represent the concentration in the dense urban area in Antwerp).

5. Conclusions

We have validated and optimized the high resolution ATMO-Street air quality model chain using the data of a large-scale citizen science measurement campaign. The extensiveness of the measurement dataset allows us to perform an in-depth model validation and optimization. We have evaluated the modeled concentrations by clustering the sampling sites by different aspects (type of location, concentration class etc.), thereby paying special attention to the small-scale spatial variability of the NO_2-concentrations. Optimizations guided by the data increased the

Table 2
Validation statistics for the optimized ATMO-Street model, with sampling locations clustered by location type. The table provides the bias, the relative bias, the relative bias-corrected RMSE (BCRMSE) and the Pearson R^2 coefficient. Columns 2 and 3 are related to the breakdown based on the model applied at the sampler location, columns 4 and 5 to the breakdown based on the availability of traffic data and the remaining columns to the breakdown based on the Flemish cities. More details on the binning are provided in the main text.

Statistic	IFDM locations (isolated building)	OSPM locations (multiple buildings)	Traffic data available	Traffic data unavailable	Antwerp	Ghent	Other Cities
Bias (μg/m^3)	−2.83	−2.43	−2.4	−2.9	−3.9	−1.1	−2.0
Relative Bias (%)	−14	−10	−10	−15	−12	−4	−8
Relative BCMRSE (%)	16	25	19	15	16	22	17
Pearson R^2	0.61	0.47	0.50	0.64	0.52	0.4	0.46

model performance and enhanced the capability of the model to correctly capture the spatial variation of the air pollution. The ATMO-Street model chain attains the FAIRMODE model quality objective, substantiating that the model is suited for policy support.

Our detailed model validation and optimization study reveals methodologies and insights that are of wider importance for the air quality monitoring and modelling community. Foremost, it demonstrates how the availability of an extensive spatial dataset enables a “deep validation”, which can result in substantially improved model skill. Secondly, the validation also highlights the importance of the street-canyon contributions. Only a model chain that takes the street-canyon increments caused by the recirculation of pollution explicitly into account, manages to adequately assess the NO_2-concentrations in Flanders. Thirdly, a model is only as good as the input it receives. Gaussian dispersion models and street-canyon modules are very sensitive to the availability and quality of the traffic data. Our analysis shows that the performance of the model chain is significantly reduced at locations where the traffic flows are outdated or locations which lack traffic data. Therefore, in order to improve the predictive power of street-level air quality models, a clear policy recommendation is to invest in the collection of accurate, up-to-date traffic data across the whole road network (i.e. not solely focusing on the major roads).

Finally, the most important lesson learnt is that street-level air quality models can substantially benefit from a validation using a one-off widespread spatial monitoring campaign. Such a detailed and rigorous validation of air quality models with large datasets is presently not a standard practice. Currently, the monitoring strategy of environmental monitoring agencies is focused on capturing temporal variability (i.e., high frequency monitoring at telemetric reference stations), while devoting far less attention to a profound documentation of spatial variability. As a result, model validation studies must typically focus on a small number of sampling sites. The analysis presented here, such as the semivariogram analysis regarding the spatial variation of concentrations, however highlights the importance of such large-scale measurement datasets with a high spatial resolution. In the case of NO_2, such widespread spatial data collection is possible through mass-scale citizen science using low-cost passive samplers. As such, citizen science offers not only a tool to increase awareness about air quality, but also removes a critical bottleneck to ascertain and improve the quality of air quality models.

CRediT authorship contribution statement

H. Hooyberghs: Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **S. De Craemer:** Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing. **W. Lefebvre:** Methodology, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **S. Vranckx:** Formal analysis, Writing – original draft. **B. Maiheu:** Methodology, Formal analysis. **E. Trimpeneers:** Writing – review & editing. **C. Vanpoucke:** Writing – review & editing. **S. Janssen:** Supervision. **F.J.R. Meysman:** Conceptualization, Supervision, Project administration, Writing – review & editing. **F. Fierens:** Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by funding for the citizen science project CurieuzeNeuzen Vlaanderen. We thank Prof. R. Blust at University of Antwerp, M. Naert and I. Renson at the newspaper De Standaard and Mr. M. Van Peteghem at Vlaamse Milieumaatschappij for enabling the CurieuzeNeuzen project. Foremost, we gratefully thank all 20.000 citizens of the CurieuzeNeuzen project for their enthusiastic participation and data collection.

Appendix. definition of the validation statistics

In this appendix, we provide an overview of the validation statistics that have been used. We henceforth assume that the difference between data set X (with values x_i) and data set Y (with values y_i) is studied. $< . >$ indicates the mean of a dataset, f.e. x is the mean of X.

- **Bias**: The bias indicates the relative difference between both data sets. Here the bias is indicated relative to the mean of the measurements.

$$Bias = \langle X \rangle - \langle Y \rangle$$

- **Root mean square error (RMSE)**: The RMSE is the sample standard deviation of the differences between predicted values and observed values. Both the absolute and relative RMSE are used. The absolute RMSE is

$$RMSE = \sqrt{\frac{1}{N}\sum_i (x_i - y_i)^2}$$

while the relative RMSE is

$$RMSE = \frac{\sqrt{\frac{1}{N}\sum_i (x_i - y_i)^2}}{\langle X \rangle}$$

- **Bias corrected root mean square error (BCRMSE)**: The BCRMSE is the RMSE of the unbiased data sets.

$$BCRMSE = \sqrt{\frac{1}{N}\sum_i ((x_i - \langle X \rangle) - (y_i - \langle Y \rangle))^2}$$

- **The Pearson correlation coefficient** quantifies is a measure of linear correlation between two sets of data. We always report the square of the Pearson coefficient, R^2, where R is defined as

$$R = \frac{COV(X, Y)}{\sigma(X)\sigma(Y)}$$

- **Factor2 (FAC2):** the FAC2 indicates the percentage of modeled points that lies within a factor two of the measured values, i.e. the percentage of data points that satisfies $\frac{1}{2}y_i < x_i < 2\, y_i$.

References

Baker, C.J., Hargreaves, D.M., 2001. Wind tunnel evaluation of a vehicle pollution dispersion model. J. Wind Eng. Ind. Aerod. 89, 187–200. https://doi.org/10.1016/S0167-6105(00)00061-1.

Berkowicz, R., Hertel, O., Larsen, S., Sørensen, N., Nielsen, M., 1997. Modelling traffic pollution in streets. Natl. Environ. Res. Inst., Roskilde, Denmark 10129, 20.

Berkowicz, R., Ketzel, M., Lofstrom, P., Rordam, H., 2008. NO2 Chemisty Scheme in OSPM and Other Danish Models.

Bo, M., Salizzoni, P., Pognant, F., Mezzalama, R., Clerico, M., 2020. A combined citizen science—modelling approach for NO2 assessment in Torino urban agglomeration. Atmosphere 11. https://doi.org/10.3390/ATMOS11070721, 721 11:721, 2020.

Brusselen, D Van, Oñate, WA de, Maiheu, B., Vranckx, S., Lefebvre, W., Janssen, S., Nawrot, T.S., Nemery, B., Avonts, D., 2016. Health impact assessment of a predicted air quality Change by moving traffic from an urban ring road into a tunnel. The case of Antwerp, Belgium. PLoS One. https://doi.org/10.1371/journal.pone.0154052. Accepted:1–19.

Bultynck, H., Malet, L.M., 1972. Evaluation of atmospheric dilution factors for effluents diffused from an elevated continuous point source. Tellus 24, 455–472. https://doi.org/10.1111/j.2153-3490.1972.tb01572.x.

Copernicus Climate Change Service, 2017. ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47.

Cressie, N., 1992. Statistics for spatial data. Terra. Nova 4, 613–617. https://doi.org/10.1111/j.1365-3121.1992.tb00605.x.

Cyrys, J., Eeftens, M., Heinrich, J., Ampe, C., Armengaud, A., Beelen, R., Bellander, T., Beregszaszi, T., Birk, M., Cesaroni, G., Cirach, M., de Hoogh, K., De Nazelle, A., de Vocht, F., Declercq, C., Dedele, A., Dimakopoulou, K., Eriksen, K., Galassi, C., Graulevičiene, R., Grivas, G., Gruzieva, O., Gustafsson, A.H., Hoffmann, B., Iakovides, M., Ineichen, A., Krämer, U., Lanki, T., Lozano, P., Madsen, C., Meliefste, K., Modig, L., Mölter, A., Mosler, G., Nieuwenhuijsen, M., Nonnemacher, M., Oldenwening, M., Peters, A., Pontet, S., Probst-Hensch, N., Quass, U., Raaschou-Nielsen, O., Ranzi, A., Sugiri, D., Stephanou, E.G., Taimisto, P., Tsai, M.Y., E, Vaskovi, Villani, S., Wang, M., Brunekreef, B., Hoek, G., 2012. Variation of NO2 and NOx concentrations between and within 36 European study areas: results from the ESCAPE study. Atmos. Environ. 62, 374–390. https://doi.org/10.1016/j.atmosenv.2012.07.080.

De Craemer, S., Vercauteren, J., Fierens, F., Lefebvre, W., Hooyberghs, H., Meysman, F., 2020a. CurieuzeNeuzen: Monitoring Air Quality Together with 20.000 Citizens.

De Craemer, S., Vercauteren, J., Fierens, F., Lefebvre, W., Meysman, F., 2020b. Using large-scale NO2 data from citizen science for air quality compliance and policy support. Environ. Sci. Technol. 54, 11070 https://doi.org/10.1021/acs.est.0c02436.

EEA, 2019. Air quality in Europe 2019. https://www.eea.europa.eu/publications/air-quality-in-europe-2019. (Accessed 10 August 2020).

Faustini, A., Rapp, R., Forastiere, F., 2014. Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. Eur. Respir. J. 44, 744–753. https://doi.org/10.1183/09031936.00114713.

Hoek, G., 2017. Methods for assessing long-term exposures to outdoor air pollutants. Curr. Environ. Heal. Rep. 4, 450–462. https://doi.org/10.1007/s40572-017-0169-5.

Hoek, G., Krishnan, R.M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., Kaufman, J.D., 2013. Long-term air pollution exposure and cardio-respiratory mortality: a review. Environ. Heal. A Glob. Access Sci. Sourc. 12, 43. https://doi.org/10.1186/1476-069X-12-43.

Hood, C., Mackenzie, I., Stocker, J., Johnson, K., Carruthers, D., Vieno, M., Doherty, R., 2018. Air quality simulations for London using a coupled regional-to-local modelling system. Atmos. Chem. Phys. 18, 11221–11245. https://doi.org/10.5194/acp-18-11221-2018.

Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., 2006. Spatial interpolation of ambient ozone concentrations from sparse monitoring points in Belgium. J. Environ. Monit. 8, 1129. https://doi.org/10.1039/b612607n.

Irwin, A., 2018. No PhDs needed: how citizen science is transforming research. Nature 562, 480–482. https://doi.org/10.1038/d41586-018-07106-5.

Janssen, S., Dumont, G., Fierens, F., Mensink, C., 2008a. Spatial interpolation of air pollution measurements using CORINE land cover data. Atmos. Environ. 42, 4884–4903. https://doi.org/10.1016/j.atmosenv.2008.02.043.

Janssen, S., Fierens, F., Dumont, G., Mensink, C., 2008b. Rio: a novel approach for air pollution mapping. In: Hrvatski Meteoroloski Casopis, pp. 172–176.

Jensen, S.S., Ketzel, M., Becker, T., Christensen, J., Brandt, J., Plejdrup, M., Winther, M., Nielsen, O.K., Hertel, O., Ellermann, T., 2017. High resolution multi-scale air quality modelling for all streets in Denmark. Transport. Res. Transport Environ. 52, 322–339. https://doi.org/10.1016/j.trd.2017.02.019.

Jerrett, M., Arain, A., Kanaroglou, P., Beckerman, B., Potoglou, D., Sahsuvaroglu, T., Morrison, J., Giovis, C., 2005. A review and evaluation of intraurban air pollution exposure models. J. Expo. Anal. Environ. Epidemiol. 15, 185–204.

Ketzel, M., Berkowicz, R., Lohmeyer, A., 2000. Comparison of numerical street dispersion models with results from wind tunnel and field measurements. Environ. Monit. Assess. 651 65, 363–370. https://doi.org/10.1023/A:1006460724703, 2000.

Lefebvre, W., Degrawe, B., Beckx, C., Vanhulsel, M., Kochan, B., Bellemans, T., Janssens, D., Wets, G., 2013a. Presentation and evaluation of an integrated model chain to respond to traf fi c- and health-related policy questions. Environ. Model. Software 40, 160–170. https://doi.org/10.1016/j.envsoft.2012.09.003.

Lefebvre, W., Van Poppel, M., Maiheu, B., Janssen, S., Dons, E., 2013b. Evaluation of the RIO-IFDM-street canyon model chain. Atmos. Environ. 77, 325–337. https://doi.org/10.1016/j.atmosenv.2013.05.026.

Lefebvre, W., Vercauteren, J., Schrooten, L., Janssen, S., Degraeuwe, B., Maenhaut, W., de Vlieger, I., Vankerkom, J., Cosemans, G., Mensink, C., Veldeman, N., Deutsch, F., Van Looy, S., Peelaerts, W., Lefebre, F., 2011. Validation of the MIMOSA-AURORA-IFDM model chain for policy support: modeling concentrations of elemental carbon in Flanders. Atmos. Environ. 45, 6705–6713. https://doi.org/10.1016/j.atmosenv.2011.08.033.

Marshall, J.D., Nethery, E., Brauer, M., 2008. Within-urban variability in ambient air pollution: comparison of estimation methods. Atmos. Environ. 42, 1359–1369. https://doi.org/10.1016/j.atmosenv.2007.08.012.

Meysman, F., De Craemer, S., Lefebvre, W., Vercauteren, J., Sluydts, V., Dons, E., Hooyberghs, H., Van den Bossche, J., Trimpeneers, E., Fierens, F., Huyse, H., 2022. Citizen Science Reveals the Spatial Structure of NO2 Traffic-Related Air Pollution. In preparation.

Miranda, A., Silveira, C., Ferreira, J., Monteiro, A., Lopes, D., Relvas, H., Borrego, C., Roebeling, P., 2015. Current air quality plans in Europe designed to support air quality management policies. Atmos. Pollut. Res. 6, 434–443. https://doi.org/10.5094/APR.2015.048.

Ntziachristos, L., Gkatzoflias, D., Kouridis, C., Samaras, Z., 2009. COPERT: A European Road Transport Emission Inventory Model. Springer, Berlin, Heidelberg, pp. 491–504.

Ottosen, T.-B., Kakosimos, K.E., Johansson, C., Hertel, O., Brandt, J., Skov, H., Berkowicz, R., Ellermann, T., Jensen, S.S., Ketzel, M., 2015. Analysis of the impact of inhomogeneous emissions in a semi-parameterized street canyon model. Geosci. Model Dev. Discuss. (GMDD) 8, 935–977. https://doi.org/10.5194/gmdd-8-935-2015.

Thunis, P., Miranda, A., Baldasano, J.M., Blond, N., Douros, J., Graff, A., Janssen, S., Juda-Rezler, K., Karvosenoja, N., Maffeis, G., Martilli, A., Rasoloharimahefa, M., Real, E., Viaene, P., Volta, M., White, L., 2016. Overview of current regional and local scale air quality modelling practices: assessment and planning tools in the EU. Environ. Sci. Pol. 65, 13–21. https://doi.org/10.1016/j.envsci.2016.03.013.

Van Brussel, S., Huyse, H., 2019. Citizen science on speed? Realising the triple objective of scientific rigour, policy influence and deep citizen engagement in a large-scale citizen science project on ambient air quality in Antwerp. J. Environ. Plann. Manag. 62, 534–551. https://doi.org/10.1080/09640568.2018.1428183.

Vardoulakis, S., Fisher, B.E.A., Pericleous, K., Gonzalez-Flesca, N., 2003. Modelling air quality in street canyons: a review. Atmos. Environ. 37, 155–182. https://doi.org/10.1016/S1352-2310(02)00857-9.

Vardoulakis, S., Solazzo, E., Lumbreras, J., 2011. Intra-urban and street scale variability of BTEX, NO2 and O3 in Birmingham, UK: implications for exposure assessment. Atmos. Environ. 45, 5069–5078. https://doi.org/10.1016/J.ATMOSENV.2011.06.038.

Veldeman, N., Maiheu, B., Lefebvre, W., Viaene, P., Deutsch, F., Janssen, S., Vanhulsel, M., Janssen, L., Peelaerts, W., Driesen, G., Van Looy, S., Hooyberghs, H., 2016. Rapport activiteiten in 2015 uitgevoerd in kader van de referentietaak 12 "Kenniscentrum Luchtkwaliteitmodellering".

WHO, 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease.

Xie, X., Semanjski, I., Gautama, S., Tsiligianni, E., Deligiannis, N., Rajan, R., Pasveer, F., Philips, W., 2017. A review of urban air pollution monitoring and exposure assessment methods. ISPRS Int. J. Geo-Inf. 6, 389. https://doi.org/10.3390/ijgi6120389.