On the Utilization of Real-Time Activity and Air Quality Sensor Data in a Local-Scale Operative Dispersion Model in Helsinki

Lasse Johansson, Age Poom, Jarkko Niemi, Hilkka Timonen, and Ari Karppinen

Abstract We presented a prototype of an operative air quality modelling system for Helsinki region that utilizes measurement data of variable quality (stations and sensors). The system continuously taps into various sources of real-time activity data to support the modelling. Being an updated version of FMI-Enfuser and developed during UIA-HOPE project we refer to this new modelling system as HOPE-Enfuser. Using public cloud storage, the modelling output is freely accessible for 3rd party applications. This is also the first time when we provide modelled Lung Deposited Surface Area (LDSA), and black carbon (BC) concentration predictions with our modelling system. For shipping, we present a novel near-real-time approach for the dynamic modelling of shipping emissions. HOPE-Enfuser uses data fusion to adjust the modelling and learn from the measurement evidence. The Covid-19 pandemic caused disruptions in the behavior of people and in the patterns of local emission sources. In this paper we will also discuss how these disruptions were perceived and captured by the model.

Keywords Dispersion modelling · Sensor networks · Data fusion · Shipping emissions · Covid-19

1 Methods

The HOPE-Enfuser is an operative local scale air quality model (a combination of Gaussian Puff & Plume) used in the Helsinki Metropolitan area in Finland. The previous versions of the Enfuser model (Johansson et al., 2015; Mensink & Volker, 2021) has also been used in foreign installation sites such as Nanjing, China and

L. Johansson (✉) · H. Timonen · A. Karppinen
Finnish Meteorological Institute, P.O. Box 503, 00101 Helsinki, Finland
e-mail: lasse.johansson@fmi.fi

A. Poom
Helsinki Institute of Urban and Regional Studies (Urbaria), Helsinki, Finland

J. Niemi
Helsinki Region Environmental Authority HSY, Ilmalantori 1, 00240 Helsinki, Finland

C. Mensink and O. Jorba (eds.), *Air Pollution Modeling and its Application XXVIII*, Springer Proceedings in Complexity, https://doi.org/10.1007/978-3-031-12786-1_6

Delhi, India. The traditional modelled pollutant species are NO_2, NO, O_3, $PM_{2.5}$ and PM_{10} for which the model provides hourly average concentrations at breathing height of 2 m above ground. These model predictions are updated several times per day, each time including a "now-casting" period with measurements (up to 24 h in the past and a forecasting period to the future (up to 48 h) in a resolution of 13×13 m. The key source of information to support the modelling are (a), NWP meteorological data (HIRLAM, HARMONIE and GFS) (b), AQ measurement data for data fusion (c), GIS-data to describe the modelling environment (e.g., digital surface maps, land-use and road network (OpenStreetMap), Satellite data from Copernicus Sentinel-2, population data from Global Human Settlement) (d), regional scale AQ forecasts from FMI-SILAM (Sofiev et al., 2015) (e), local emission inventories and (f), supporting activity data. The most notable modelled emission source categories are (a), traffic (hourly emissions computed for each road separately addressing hourly flows of cars, heavy vehicles and buses (b), household combustion emissions (c), shipping emissions via FMI-STEAM (Johansson et al., 2017) (d), power plant emissions as elevated point sources and (f), the regional background.

Since the adoption, the model has been developed further during the on-going UIA-HOPE project in Helsinki, especially by connecting to new information sources to support the modelling. This extended version is also an operational service and provides online predictions to several use-cases. However, due to its novelty and the amount of new implemented features, the service is an unofficial one and is mainly used by internal partner of the UIA-HOPE project. There are several notable additions in the extended model version currently being developed. We have included LDSA, black carbon (BC) and CO to the list of modelled species. We also continuously extract several new types of supporting activity data (real-time shipping, traffic flow data and road weather measurements to name a few), which we focus more on Sect. 1.2. A real-time connection to a complementary AQ sensor network has also been added (Sect. 1.1).

Finally, as a technical novelty the modelling data is pushed to an Amazon S3 cloud storage. This storage a public access and a rolling archive of hourly modelling datasets is kept there for the past 2 weeks from where 3rd party application can access the model predictions. As an example of an application that utilizes the HOPE-Enfuser modelling data we present the Green Paths tool by University of Helsinki (Poom et al., 2020) (Fig. 1). This open-source tool assesses multiple environmental exposures (e.g., AQ and noise) for navigation paths and finds exposure-optimised routes for the users. For AQ related costs, the pathing algorithm uses Air Quality Index (AQI), that is provided by HOPE-Enfuser based on the pollutant species concentrations.

1.1 The Online Measurement Networks

HOPE-Enfuser utilizes data fusion with AQ measurements during its use. This includes online sensor data, being the first time where we have utilized such sensor data in an operational air quality modelling service. This can be regarded as a more

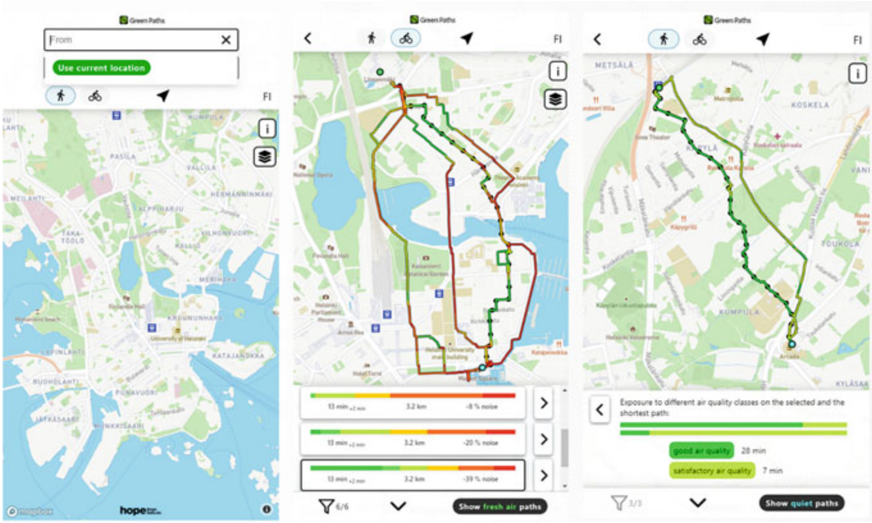

Fig. 1 Caption of the Green Paths—navigation application's user interface. The service can be accessed from: https://green-paths.web.app/

challenging approach than an offline-based demonstration, which the author of this study presented previously in Mensink and Volker (2021).

Briefly put, in the data fusion we apply adjustments to hourly emission factors to a set of emission source categories (e.g., traffic, households) to adapt the hourly modelling to a higher agreement with measurement evidence. For the background concentration a flat correction offset is also applied. In addition to a stronger hourly correction these adjustments also gradually contribute to a longer-term learning pattern, for which we provide a concrete example in Sect. 2. The longer-term corrections captured by the model can also adjust diurnal patterns for emission sources if needed.

The overall measurement network (as well as the modelling domain) configuration has been shown in Fig. 2. The base figure is a caption of monthly NO_2 average concentration (March 2021), automatically generated by the HOPE-Enfuser modelling service. Within the base figure we have highlighted three sensor focus areas of Pakila, Vallila and Jätkäsaari, in which most of the 25 AQ sensors (Vaisala AQT530) have been installed to. The official AQ measurement station network is more evenly distributed. It should be noted that for O_3 there are only 4 measurement sites, for BC there are 7 and for CO there are none. The network of LDSA-sensors (by Pegasor) is mostly concentrated on Pakila focus area (locations not shown in Fig. 1). The whole network is maintained by local authority (HSY), which has also provided emission inventories (households) for the Helsinki modelling area. Finally, it should be noted that during the data fusion we must address the measurement quality differences (i.e., preliminary weights) and we have assigned an order of magnitude lower quality estimate for the sensors with respect the refence stations.

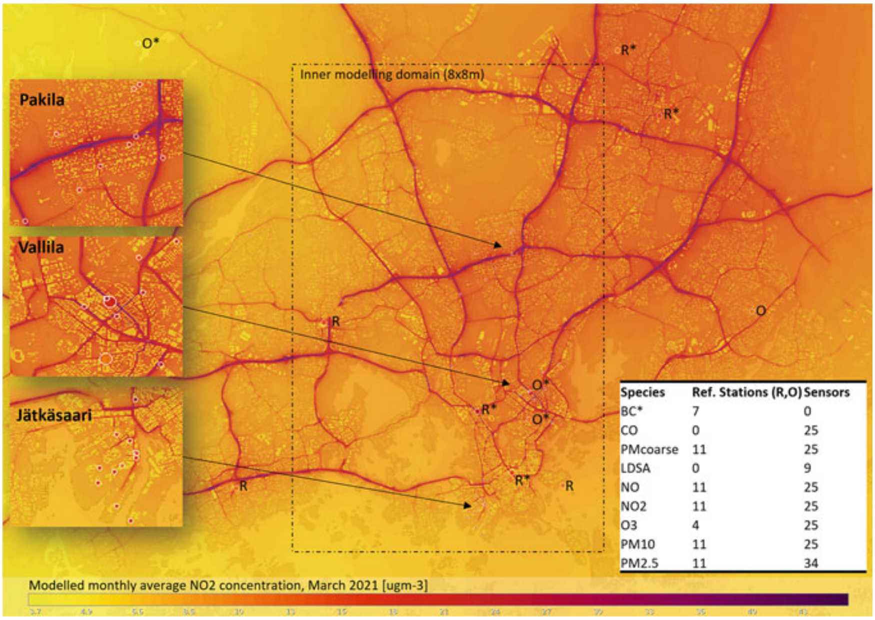

Fig. 2 Illustration of measurement network and modelling domain configuration, using a monthly average modelling result as an example. Coloured 'dots' correspond to observed monthly average NO_2 concentrations but more importantly shows the measurement location. For sensors the dots are smaller. * = station that also measures BC. O = station that measures O_3. All stations also measure $PM_{2.5}$, NO_2 and PM_{10}

This topic of measurement reliability and weighting is further discussed in Mensink and Volker (2021).

1.2 Real-Time Activity Data

For a reactive high-resolution AQ modelling system it is often not enough to rely on static emission inventories. This in mind HOPE-Enfuser continuously extracts information from online sources to get a more dynamic overview on the modelling area. As an example, we extract and use data from:

- Road weather measurements (more than 20 locations) to improve our road dust modelling capabilities. Source: DigiTraffic (https://tie.digitraffic.fi/en/).
- Traffic congestion data from Here.com. Main use-case is to address the effect of notable traffic jams and to analyse longer term traffic patterns in the area.
- Hourly traffic flow measurements from 55 monitoring sites, including a split between directions and vehicle categories. Source: DigiTraffic.

- Real-time shipping movement patterns via a local AIS-receiver coupled to the AISHub service (https://www.aishub.net/).

The most ambitious addition of these is the real-time shipping data and its derivation into continuous shipping emissions, which we focus on next. Helsinki centre has a busy passenger shipping port, which is also close to the Jätkäsaari focus area. To model the shipping emissions as they occur, we have adopted the following solution:

- We contribute with our own local AIS-receiver and connect to AISHub service to access real-time shipping activity data (AIS) globally but more importantly, for the Helsinki coastal area.
- On an hourly basis we use the archived recent AIS-data as input to the FMI-STEAM model, which creates a special kind of hourly emission inventory dataset. This is basically text-based line-data that describes emission release rates (g/s), time, location and an estimate for release height (m) for the emissions.
- Produced hourly emission inventory packages (the line-data text-files) by the STEAM-model are pushed to Amazon S3 cloud storage.
- Whenever HOPE-Enfuser modelling task occurs these fresh shipping emissions are accessed from the cloud storage (Fig. 3) and dispersion modelling occurs separately for each line of activity.

The special format of the emission inventory makes it possible to represent the shipping emissions at any chosen resolution by modelling the emissions as moving point sources. Due to technical limitations, we have settled in representing the shipping emissions as a gridded inventory (for Gaussian Puff releases) with a resolution of approximately 100 m and 5 min. For the forecasting period for the model, we simply use the shipping movements that occurred one week before.

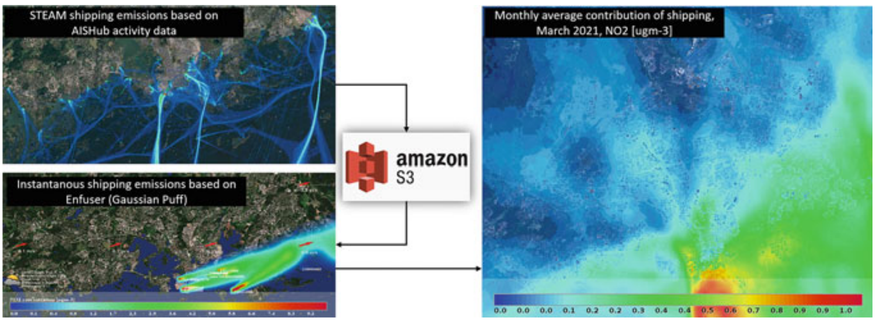

Fig. 3 The concept of near-real-time shipping emissions modelling in HOPE-Enfuser. AIS-activity data is used as input for FMI-STEAM that predicts shipping emissions. HOPE-Enfuser, while accessing AWS S3, uses this emission inventory data as input for dispersion modelling. In Right, an example of the shipping contribution to NO_2 concentrations near Helsinki is shown (March 2021, using the same geographic area as in Fig. 2)

2 Results and Discussion

The HOPE-Enfuser service has been operational during the Covid-19 epidemic, and this has provided the possibility to inspect the indirect effects of the epidemic to air quality in Helsinki. In Fig. 4 a comparison of diurnal average NO_2 concentration during Spring (March to May, 2018 and 2020) is shown for an urban traffic measurement site. The differences are drastic, and similar strong reductions in traffic-related concentrations are also visible in other urban measurement sites (not shown here). During early 2020, we also observe the traffic emission factors (being coupled to measurement evidence via the data fusion) started to decline gradually in HOPE-Enfuser. The hourly traffic flow measurement data, which we have formed into a generalized index (percent of average hourly patterns observed in 2017), also declined strongly for passenger cars at the same time. It should be noted that the decline in longer-term emission factor corrections is slower than the actual rapid decline in traffic flow counts, however the floating emission factor correction eventually stabilizes to a fixed level in July.

The modelled shipping emissions according to FMI-STEAM also reduced notably near the coastal areas of Helsinki, however the reductions occur later than with road traffic (after May 2020). For example, $PM_{2.5}$ shipping emissions in January-March 2020 near Helsinki are 120% with respect to the emissions in the previous year of 2019, but this relation is reduced to 73% during May–July respectively.

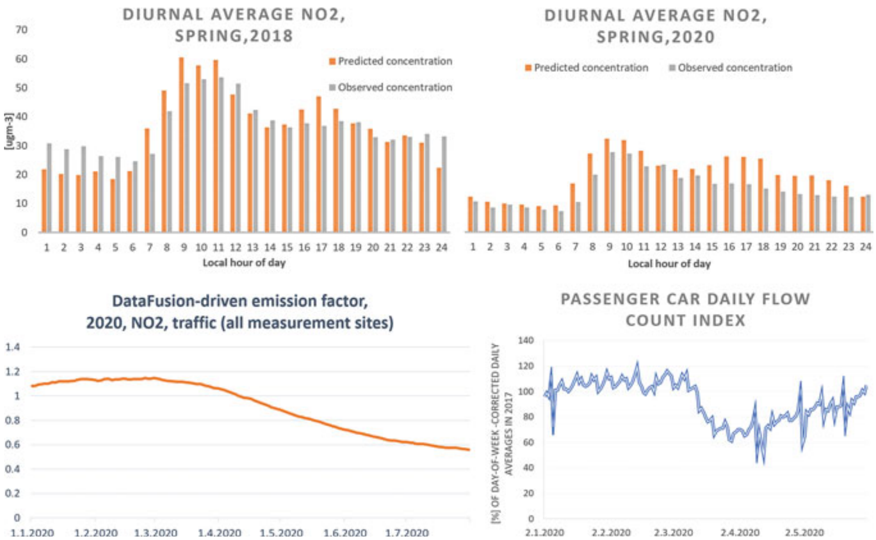

Fig. 4 Indirect effects of the pandemic during early 2020. Up, a comparison of diurnal average NO_2 concentrations at a selected urban traffic site (60.16964 N, 24.93924 E) is shown. Below, the data-fusion driven longer-term adjustment to traffic emissions is shown, as well as a derived passenger car flow count index based on 55 traffic flow count measurement sites in the area

As shown, the modelling system attempts to learn and adjust the modelling based on the measurement evidence. It has been designed to automatically adapt to changes in the modelling environment, with respect to local emission sources. The Covid-19 pandemic caused a notable disruption in the behavior of local emission sources in Helsinki and therefore provided a useful testing case for the performance of the modelling system. According to preliminary results, we have seen that the modelling system has been able to adapt to the changes in traffic patterns and to the changes in shipping emissions in the area. For traffic it should be noted that the corrections done to emission factors cannot yet account for geographical differences in the changes and the automatic adjustments are applied similarly to the whole modelling area. Interestingly, the traffic flow data shows a rapid recovery, while the AQ concentrations (and the traffic-related emission factors) remain low long afterwards with respect 2019, or 2018.

One of the key features of the modelling system is the simultaneous use of AQ data from both reference stations and a complementary network of sensors. This is a topic in which we concentrate more in further studies. Briefly put, there are 'pros and cons' related to the use of sensor data in such a way. Theoretically, an extended network of measurement devices makes the data fusion methodology more robust and facilitates more detailed input for the modelling system that attempts to learn via the measurement evidence. On the other hand, we have seen that when (a), the group of sensors is large and (b), their measurement biases are strongly correlated and (c), the group of reference stations is small, then the sensors can in fact harm the modelling performance. As an example, we currently observe this detrimental effect with our O_3 modelling results.

The evaluation of the modelling performance of our new pollutant species LDSA and BC will also be addressed in future papers. For LDSA there are a couple of challenges to solve: first, the LDSA sensors are located on suburban residential areas, some sensors being very close to households. This means that individual households may cause sudden peaks in LDSA concentrations, which are not possible to be captured by the model. Also, there is no LDSA background information available from the regional scale model we use; we simply proxy it from $PM_{2.5}$ and rely on data fusion -driven corrections. In our recent rolling 2-month assessment (Leave-one-out validation, July–August 2021) the hourly LDSA has an R2 of 0.49 against the measurements. There are similar issues with BC, however the rolling 2-month average R2 shows a higher value of 0.58. Again, we have no direct source of information for long range transportation of BC and we instead use elemental carbon mass (EC) from FMI-SILAM. The BC values are small (majority being lower than 1 ugm-3) in Helsinki up to the point that the sensitivity of the BC measurement stations may affect the correlation between model predictions.

The dynamic modelling of shipping emissions makes it possible to explain some of the high concentrations the stations and sensor measure, especially near the coastal regions. If done successfully then this in turn aids the data fusion method as the model can associate high measured concentrations to their proper source of origin. However, we also observe that the dynamic modelling of shipping emissions can cause large margin of error from time to time, as the dispersion of elevated strong emission plumes is highly challenging near the coast. For example, HOPE-Enfuser may model a false-positive strong shipping emission plume which are not captured by the nearby measurement network. The margin of error is especially noticeable for stationary ships at port, for which the modelling of auxiliary engine use is challenging.

References

Johansson, L., Epitropou, V., Karatzas, K., Karppinen, K., Wanner, L., Vrochidis, S., Bassoukos, A., Kukkonen, J., & Kompatsiaris, I. (2015). Fusion of meteorological and air quality data extracted from the web for personalized environmental information services. *Environmental Modelling & Software, 64*, 143–155. Elsevier.

Johansson, L., Jalkanen, J.-P., & Kukkonen, J. (2017). Global assessment of shipping emissions in 2015 on a high spatial and temporal resolution. *Atmospheric Environment, 167*, 403–415. https://doi.org/10.1016/j.atmosenv.2017.08.042

Mensink, C., & Volker, M. (2021). *Air pollution modeling and its application XXVII*. Springer International Publishing. https://doi.org/10.1007/978-3-662-63760-9

Poom, A., Helle, J., & Toivonen, T. (2020). Journey planners can promote active, healthy and sustainable urban travel. *Urbaria Summaries Series.*

Sofiev, M., Vira, J., Kouznetsov, R., Prank, M., Soares, J., Alves Antunes, J., & Genikhovich, E. (2015). Construction of the SILAM Eulerian atmospheric dispersion model based on the advection algorithm of Michael Galperin. *Geoscientific Model Development, 8*, 3497–3522. https://doi.org/10.5194/gmd-8-3497-2015

https://www.aishub.net/. Visited 15.4.2021.