Improved Bottom-up Residential Inventory

Edenderry and Dungarvan Study Cases

LIFE19 GIE/IE/001101

Action B1.1: An Improved Residential Solid Fuel Inventory

Partner responsible for this report: VITO

Prepared By: Jorge Sousa, Hans Hooyberghs, Lisa Blyth, Siemen Timmermans,

Version No.: 2.2

Date: 23 January 2024

DOCUMENT VERSION RECORD

Date of Issue	Version	Section	Who	Comment
16/03/2023	V1.0	All	Jorge Sousa	Initial draft
03/04/2023	V1.1	Chapters 1 to 3	Jorge	Updates further to input from Hans Hooyberghs (B3 lead) .
17/07/2023	V2.0	Chapters 1 to 3	Jorge	The results of the Dungarvan test case were added.
21/08/2023	V2.1	All	Hans, Jorge	Review and inputs from Hans and subsequent updates
01/12/2023	V2.1.1	Chapter 3	Jorge	Described the proposed approach to extrapolate the emissions based on further analysis.
14/12/2023	V2.1.2	Chapter 4	Jorge	Finalization of the recommendations and conclusions
21/12/2023	V2.1.3	All	Lisa	Initial review of all chapters with suggested changes, mainly to Chapter 4.
21/12/2023	V2.1.4	Chapter 4	Jorge	Update of Chapter 4.
17/01/2024	V2.1.5	Chapter 4	Lisa, Jorge	Final review of all chapters with minor changes & comments.
23/01/2024	V2.2	Chapter 4	Jorge	Resolved final comments
09/07/2024	V2.3	All	Dermot	Final version

TABLE OF CONTENTS

1 Chapter 1 Introduction 1
2 Chapter 2 The Numerical Methodology 3
2.1 The Data Assimilation Approach 3
2.2 Preparation of the ATMO-Street IFDM Model and Data for Optimization of the Data Assimilation methodology 5
2.2.1 Time factors 6
2.2.2 Background concentrations: 7
2.2.3 Aggregation of the Emission data: 7
3 Chapter 3 Test case: Edenderry 11
3.1 Experimental Campaign 11
3.2 Results 13
3.2.1 Initial model evaluation 13
3.2.2 Results of the data assimilation procedure 15
3.2.3 Spatial Pattern 18
3.2.4 Temporal Pattern 19
4 Chapter 4 – Test case: Dungarvan 21
4.1 Experimental campaign and survey 21
4.2 Initial model evaluation 25
4.2.1 Results of data Assimilation 27
4.2.2 Validation with updated emissions and time factors 30
4.2.3 Proposed methodology for emission extrapolation 33
5 Chapter 5 - Conclusions and Recommendations 36
5.1 Summary 36
5.2 Conclusions regarding the MapEIre residential emission maps and subsequent policy decisions 37
5.3 Conclusions for air quality modelling purposes 38
5.4 Conclusions for FAIRMODE WG4 38
5.5 Recommendations for future research studies after LIFE Emerald 38
5.6 Implications for Action B3 and the ATMO-Street Maps 39

1 Chapter 1 Introduction

The ATMO-Street air quality model is used to calculate the annual average PM_{10} and $PM_{2.5}$ concentrations for Ireland under Action B3. One of the key inputs for production of these maps is the residential heating emissions data which is taken from the MapEIre gridded emissions dataset for 2019. [1]

The accuracy of the PM maps is highly dependent on the reliability of the emission data. Therefore, it is of paramount importance to minimize the uncertainties associated with the emission inventory. In terms of $PM_{2.5}$, residential solid fuel heating is one of the key sources across Ireland (responsible for over 50% in Ireland). However, estimating the emissions associated with solid fuel heating is a challenge as the information available on the use and emissions of solid fuel (wood, coal and peat) burning across Ireland is incomplete given the existence of unreported use as a secondary heating source and the associated unknown quantity of used fuel. Hence, estimating residential emissions comes with a large uncertainty, which translates into a high uncertainty in the $PM_{2.5}$ concentration maps, particularly during the heating season.

Given the difficulty in directly estimating the residential emissions and thus improving the MapEIre residential emissions, a possible (complementary) approach to reduce the uncertainties is to use an inverse data assimilation methodology that corrects the emission strengths used in the emission input data, based on the mismatch between the predicted and measured concentrations. In this deliverable we describe two studies where a data assimilation methodology was applied to the Action B.3 modelling set-up, using observation data from a network of low-cost sensors which have been established within Action A.2. Edenderry was selected as the location for the first study due the relevance of solid heating and the lack of other major $PM_{2.5}$ sources. A total of 13 sensors were deployed between February and May 2022 and a dedicated modelling study was performed a posteriori for the same period. The data from the sensors was used as an input to a data assimilation procedure with the objective to update emission spatial and temporal patterns. The data assimilation procedure allowed a downscaling on the emission grid, updating of the average emission value and actualization of the time profiles associated with residential emissions.

Further to some lessons learnt from this first study, a similar study was carried out for Dungarvan. In the next Chapter we present an overview of the modelling and data assimilation methodology applied in both studies, followed by the results of each study in Chapter 3. Chapter 4 provides some key conclusions and recommendations.

[1] https://projects.au.dk/mapeire/

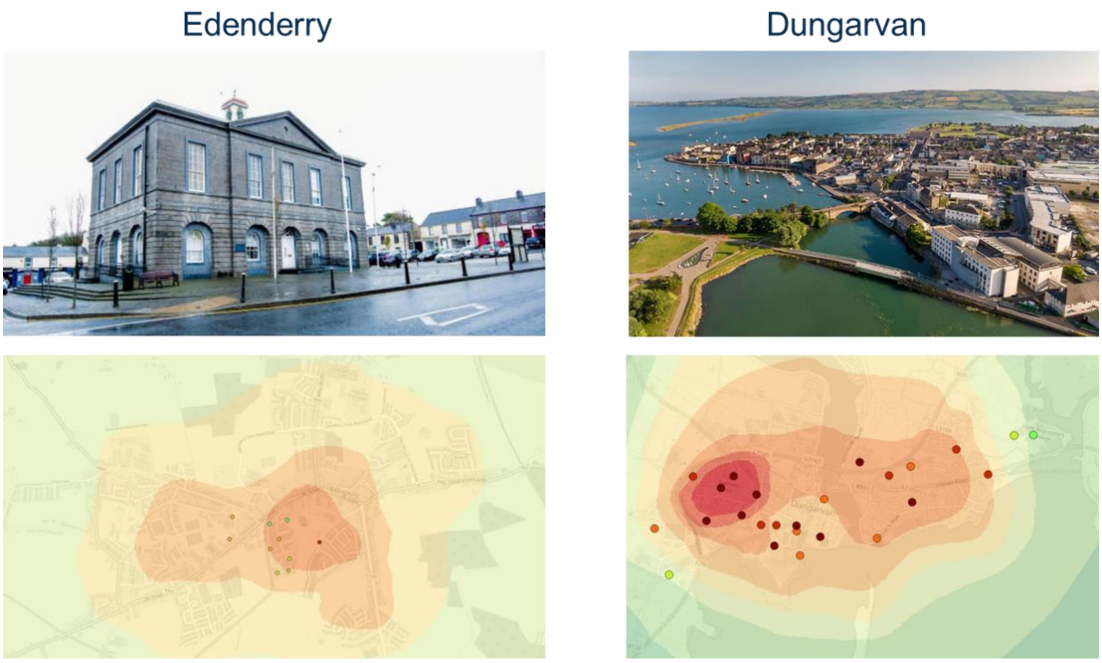


FIGURE 1 OVERVIEW OF THE MODELLING DOMAIN IN THE TWO LOCATIONS

2 Chapter 2 The Numerical Methodology

2.1 The Data Assimilation Approach

The Bayesian approach developed at VITO has been previously tested to improve urban wind field predictions with computational fluid dynamics and it has been validated with a full test campaign at the Stanford campus[2,3]. Recently this methodology was applied to Antwerp city and two different Chinese cites with the objective to identify locations of high emissions based on the IFDM dispersion model of ATMO-Street. The methodology uses as a backbone the ensemble Kalman filter with an augmented state for inverse problems [4,5]. The use of an augmented state indicates that the parameters (emission inputs) of the IFDM model are treated in the same fashion as the state field (concentration values) and are updated following the traditional Kalman update procedure as show in the following equation.

$$\overbrace{\begin{pmatrix}\alpha^{a}\\ \psi^{a}\\ \hat{d}^{a}\end{pmatrix}}^{\text{Updated}} = \begin{pmatrix}\alpha^{f}\\ \psi^{f}\\ \hat{\hat{d}}\end{pmatrix} + \overbrace{\begin{pmatrix}C_{\alpha d}\\ C_{\psi d}\\ C_{dd}\end{pmatrix}\left(C_{dd}+C_{\epsilon\epsilon}\right)^{-1}}^{\text{Kalman Gain}} \overbrace{\Big(d - \mathcal{M}\big[G(\alpha^{f})\big]\Big)}^{\text{Difference between model and obs.}},$$

α, ψ, $\hat{d}$ represent the augmented state, with the parameter value, state value and the state value at the observation locations, respectively. d represents the observation values while G(α) is a dispersion model that depends on the emission α. The *M* matrix is known as the measurement matrix that is 0 everywhere except where the observations are available. The equation updates the prior estimate of the augmented state by multiplying the Kalman gain by the difference between the model and observation. A fundamental step in this approach is the estimation of the covariances matrix for all the variables ($C_{\alpha d}$, $C_{\psi d}$ and C_{dd}) and of the experimental variance $C_{\varepsilon\varepsilon}$. Figure 2 illustrates the idea behind a Bayesian inference approach. Here the observation has a mean value and a variance associated with its uncertainty, additionally we have a predicted value with a mean and variance. The Bayesian step is supposed to generate an updated state where the variance is minimized. We can notice that in this illustration the update state moves closer to the model result due to its lower variance.

[2] Sousa, Jorge, Clara García-Sánchez, and Catherine Gorlé. "Improving urban flow predictions through data assimilation." Building and Environment 132 (2018): 282-290.
[3] Sousa, Jorge, and Catherine Gorlé. "Computational urban flow predictions with Bayesian inference: Validation with field data." Building and Environment 154 (2019): 13-22.
[4] Iglesias, Marco A., Kody JH Law, and Andrew M. Stuart. "Ensemble Kalman methods for inverse problems." *Inverse Problems* 29.4 (2013): 045001.
[5] Evensen, Geir. *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media, 2009

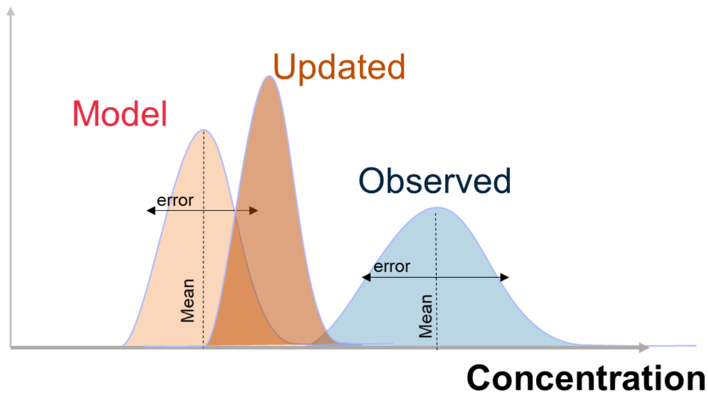


FIGURE 2 THE UPDATE PROCEDURE SHOWING PROBABILITY DENSITY FUNCTIONS (PDF) OF THE OBSERVED THE MODELLED VALUES AND THE FINAL UPDATED VALUES

A significant advantage of the ensemble approach is the fact that it relies on a physics-based dispersion model to compute the covariances, and there is no need to rely on empirical approximations. Hence, we can for example, include the street canyon effects or simple chemistry in the covariances. The following equation shows an example on how the covariances are computed with the ensemble approach.

$$C_{\alpha d} = \frac{1}{N}\sum_{j=1}^{N} u^j \left(d^j\right)^T - \bar{u}\left(\bar{d}\right)^T$$

As shown in Figure 2 we start with an initial guess on the emission value and its uncertainty (normal or uniformly distributed) and we run our model N times to compute, for a particular hour and meteorological conditions, the covariance matrix. At this step we have also propagated the uncertainty from the emissions to the concentrations, and therefore we have an estimate of the uncertainty in the concentrations. At the observation locations we compare the mean values of observations and model output as well as both model and observational uncertainty. Based on this information we can now perform the update step, where both the state values (concentrations) and the parameter values (emissions) are updated. In the approach not only is the mean value updated but also the uncertainty range, hence offering additional information on the accuracy of the estimated concentrations and emissions.

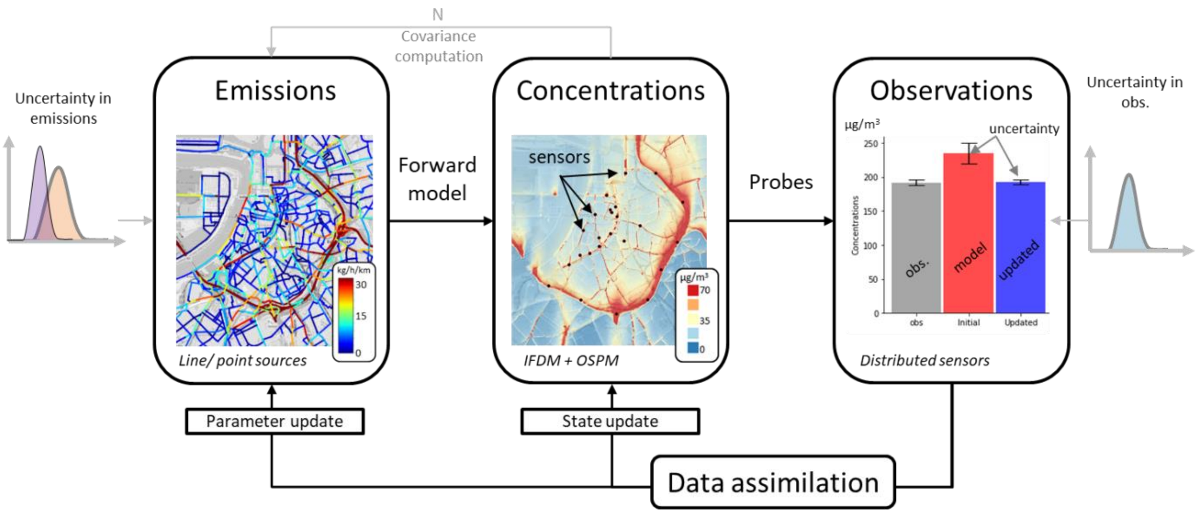


FIGURE 3 WORKFLOW OF THE DATA ASSIMILATION PRINCIPLE IN THE CONTEXT OF AIR QUALITY.

2.2 Preparation of the ATMO-Street IFDM Model and Data for Optimization of the Data Assimilation methodology

In this section we describe which data is being used for the studies in Edenderry and Dungarvan, and how the model and data are optimised to maximise the potential of the data assimilation methodology. When we mention the 'model' in the context of this report, we are referring to the core calculation kernel of the ATMO-Street-model, which is the bi-Gaussian IFDM plume model, designed to simulate non-reactive pollutant dispersion on a local scale.

The Gaussian dispersion parameters are dependent on the stability of the atmosphere and the wind speed following the Bultynck and Malet formulation based on the Bulk Richardson number (Bultynck and Malet, 1972). As ATMO-Street is a receptor-model, it can be used for every grid setup, whether it is regular or not.

As done in Action B.3 to produce the PM maps, residential emissions are taken from the output of the MapEIre project[6] which provides gridded residential emissions (sector C, Other Stationary Combustion) with a resolution of 1 km by 1 km. Gridded emission data for the year 2019 (for both modelling years) are used. To explore how the data assimilation methodology and the observations can be best applied the resolution the emissions dataset was first downscaled to a 100 by 100-meter grid as shown in Figure 4.

[6] See details on https://projects.au.dk/mapeire/

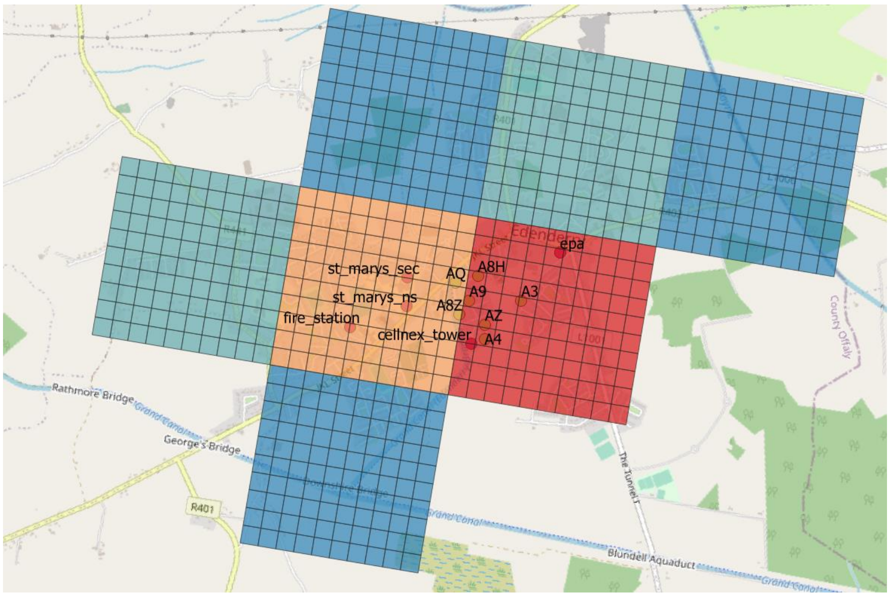


FIGURE 4 SCALED DOWN EMISSION GRID IN EDENDERRY

2.2.1 Time factors

The emission data is scaled up and down according to the time factors shown in Figure 5. In this plot the emission factors are displayed in function of time as described in the MapElre project.

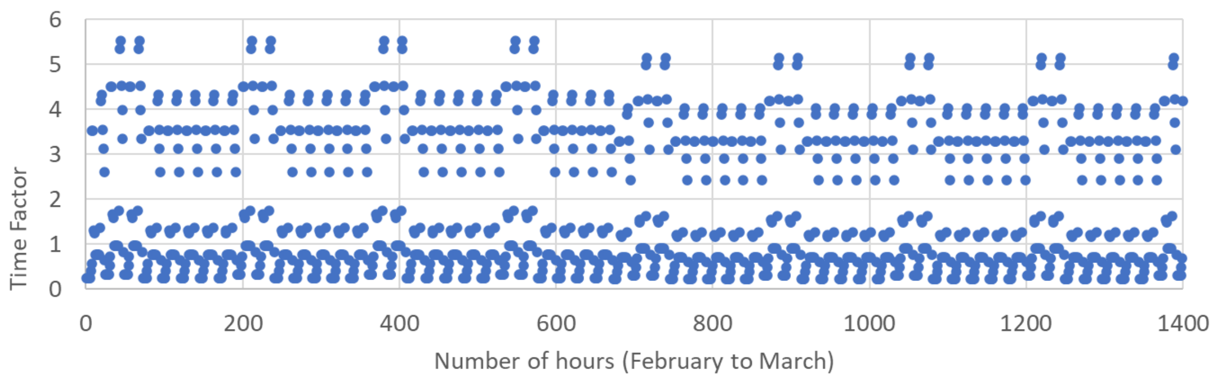


FIGURE 5 TIME FACTOR BETWEEN FEBRUARY AND MARCH

The time factors can also be divided into monthly, daily and hourly time factors as displayed in Figure 6.

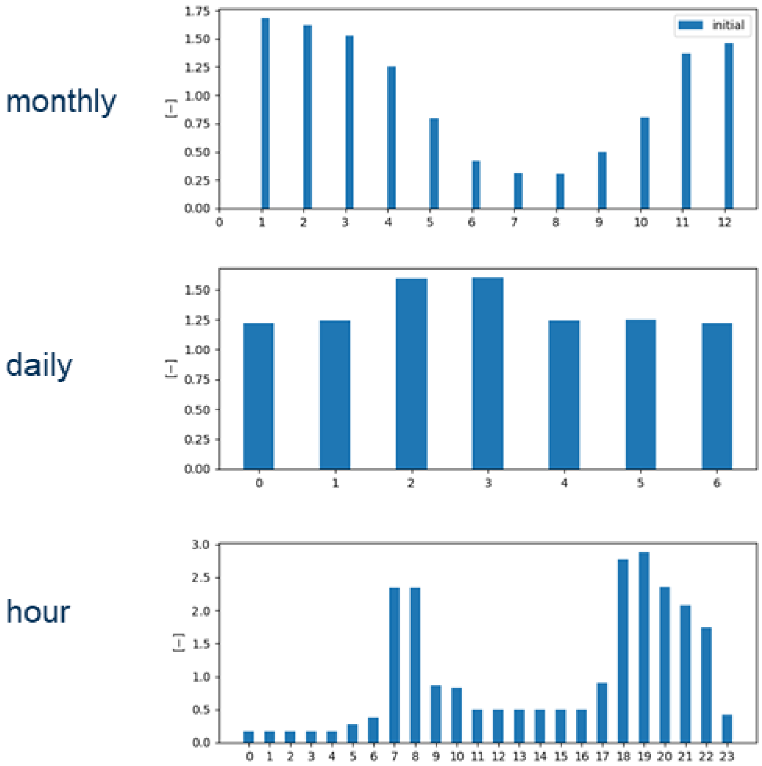


FIGURE 6 MONTHLY, DAILY AND HOURLY TIME FACTORS FOR RESIDENTIAL EMISSIONS

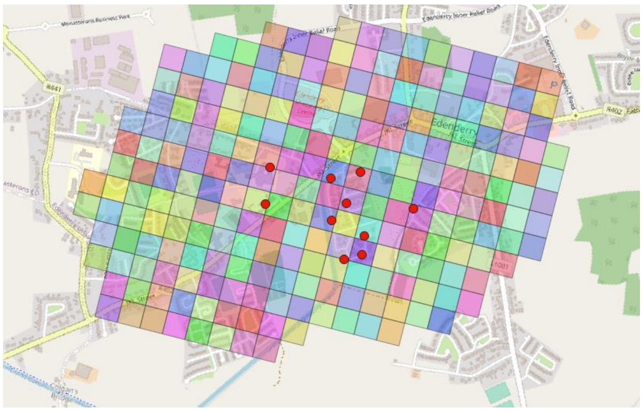
2.2.2 Background concentrations:

For the background concentrations two different options were considered. The first option uses the background RIO-maps, the second option relies on the measurements of a local background sensor. Both options were analysed for the Edenderry study case. Given the low resolution of RIO it was decided to rely on the measurements of a background station. In Edenderry, the A7 sensor has been used, whereas in Dungarvan the minimum value of all the sensors was used as a background value.

2.2.3 Aggregation of the Emission data:

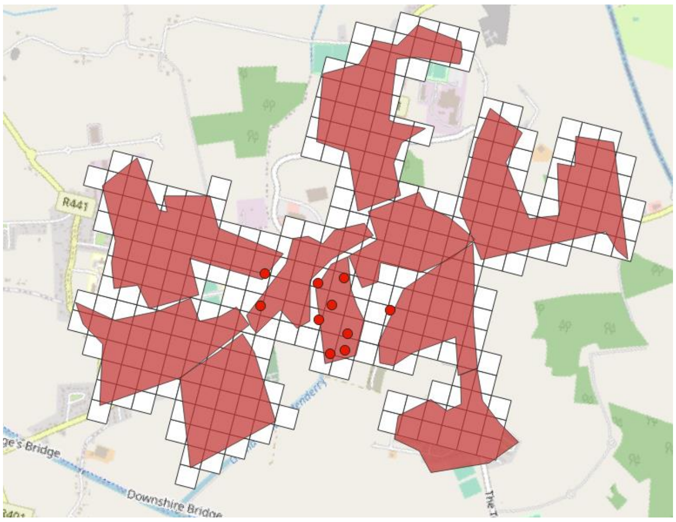
After downscaling the emissions to 100 m, one of the first tasks was to explore ways of aggregating the MapEIre gridded emissions data based on local residential emissions knowledge in Edenderry (the first study case), to maximise the potential of the data assimilation (DA) method (to minimise the 'unknowns'). This aggregation can consider the sensor data locations and any local knowledge available e.g., housing characteristics and heating use information. Three different options are shown below.

1) Raw usage of the emissions grid:



In this approach one assumes that the residential emissions for each grid cell can behave independently of the other neighbourhood cells which leads to many unknowns (an unknown per grid cell). Given the few measurement locations available this generates a fairly rigid inversion problem. On the other hand, this method would pinpoint to hyperlocal corrections with a high resolution, however, with a high level of uncertainty. Finally, the emissions knowledge would be rather difficult to extrapolate given the difficulty to generalize the residential areas associated to each grid cell.

2) Grouping emissions per neighbourhood:

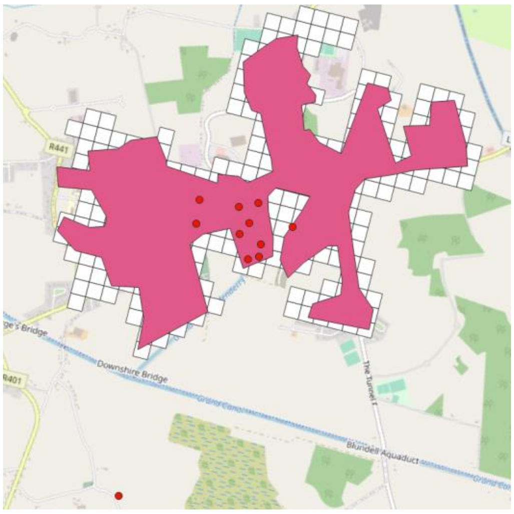


This approach reduces considerably the unknowns by grouping the emission cells into groups associated to each of the neighbourhoods in Edenderry. Thereby considering that the residential emission is an average value per neighbourhood. Additionally, only cells that overlap residential areas are considered to have an emission value. Any cell outside these areas, is deemed to have no residential emissions and is thus null.

Since, the grouping is based on urban neighbourhoods, any prior knowledge of residential heating use (for example provided by resident surveys) per urban zone could be used as prior knowledge, for an initial guess of the emissions. Finally, given the properties of each zone some generalization of the emissions and time factors can be extracted. As a drawback, hyperlocal identification of hot spots is not possible. This neighbourhood division was not based on any prior knowledge as none (based on local surveys) was available for Edenderry. Aerial images were used to make the spatial divisions.

Note: at this stage of the assessment detailed housing information (building age, renovation, heating appliances etc…) from the Central Statistics Office (CSO)/Building Energy Rating[7] (BER) databases was not used. Later, as explained in sub-section 4.2.3 "Proposed methodology for emission extrapolation" this information is used to assess how the methodology could be extended to improve the whole MapEIre residential emissions database used for modelling the PM concentrations for Ireland.

3) Uniform emissions over the urban area:



This approach reduces the unknowns to 1 variable, the average residential emission over Edenderry. This results in a rather easy inversion problem; however local corrections are not possible. The spatial pattern of $PM_{2.5}$ concentrations becomes rather spatially uniform and hence, it is difficult to match the different observations and the underlying spatial pattern.

The three different approaches have been experimented with. Option 2 was selected as it is considered to compromise between the difficulty of the inverse problem and the potential

[7] BER is an assessment of the energy efficiency of a building in Ireland, providing information on energy consumption and carbon dioxide emissions.

ability to generalize the knowledge associated with the residential emissions per neighbourhood type.

3 Chapter 3 Test case: Edenderry

3.1 Experimental Campaign

Several sensors were deployed in the town of Edenderry. Both Clarity nodes and Purple Air sensors were deployed by University College Cork. The location of each sensor is shown in Figure 7. High density was achieved in the town centre, while A7 sensor was placed further away, as a background station.

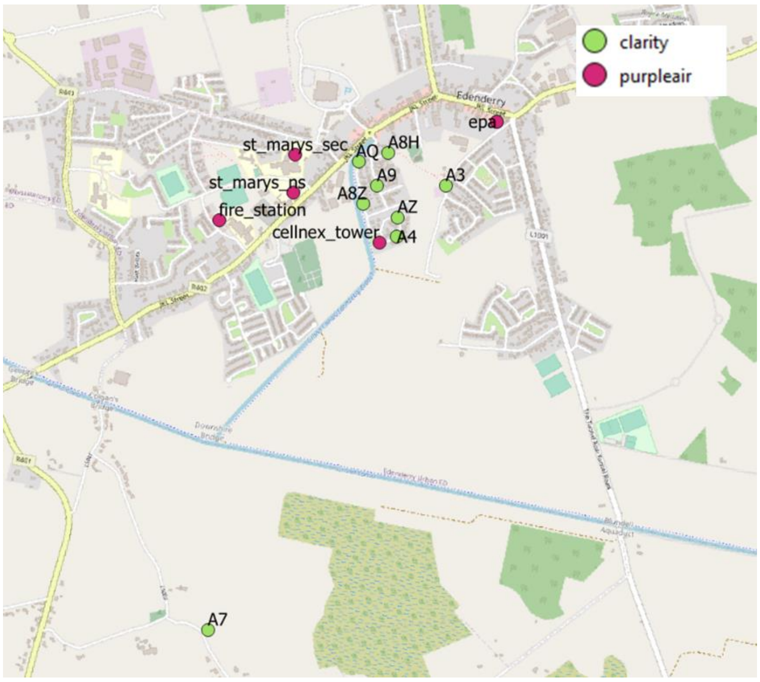


FIGURE 7 SENSOR NETWORK DEPLOYED IN EDENDERRY

During a period of 1 month both a Purple Air and a Clarity sensor were collocated with the reference station (IE005OY). The result of the collocation study is shown in the figures 7 - 8. Both sensors show a similar performance with a Pearson correlation of around 0.75, an absolute bias below 0.03 µg/m^3 and a root mean square error below 7 µg/m^3.

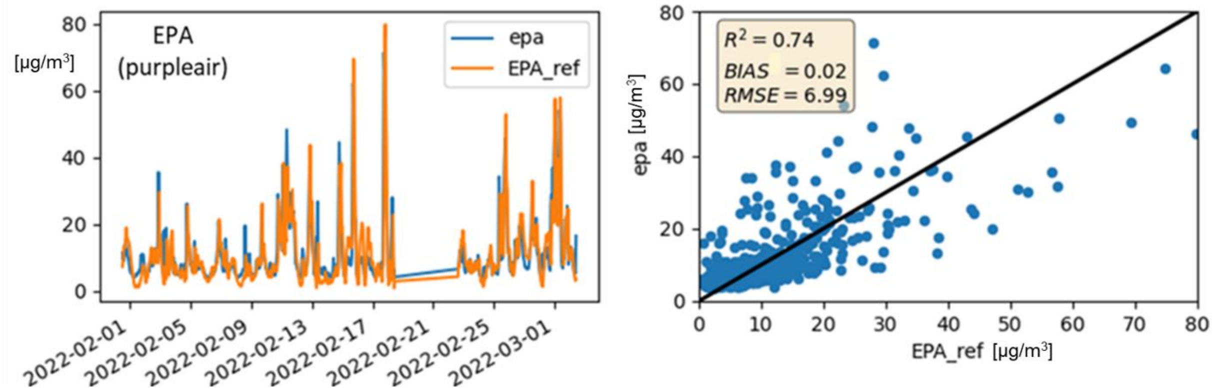


FIGURE 8 PURPLE AIR SENSOR COLLOCATED WITH THE REFERENCE STATION (IE005OY)

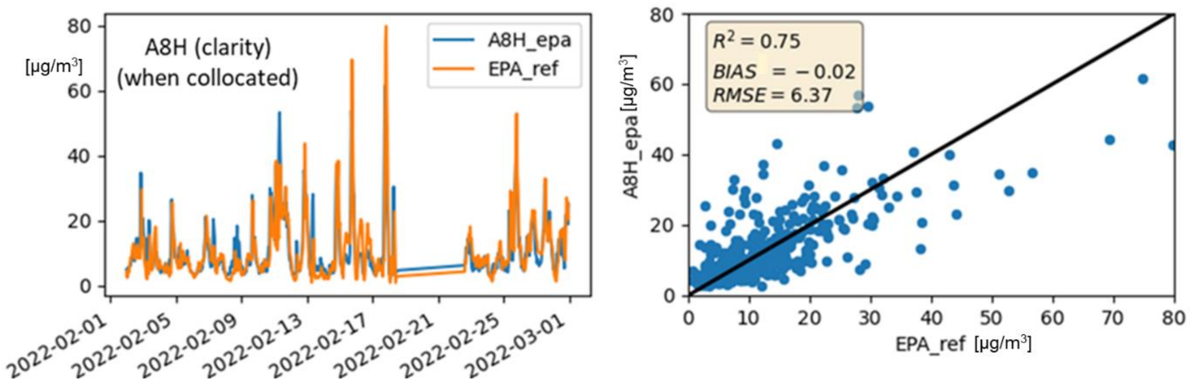


FIGURE 9 CLARITY SENSOR WHEN COLLOCATED WITH THE REFERENCE STATION (IE005OY)

3.2 Results

3.2.1 Initial model evaluation

In this section we discuss the results of the initial modelling exercise that was performed based on the initial emission map, downscaled to 100 m and the time factors shown before in Section 2.2. The statistical evaluation of the predictions is summarized in Figure 10. The R^2 is above 0.5 in all the stations expect for the fire station. There is an overall negative bias with a maximum value also observed at the fire station. High RMSE values (above 10 µg/m^3) are found for all the stations except for A7.

FIGURE 10 STATISTICAL ANALYSIS OF THE INITIAL RESULTS WITH ATMO-STREET

A time series comparison between the modelled results and the observations values is shown in Figure 11 and Figure 12. These plots represent the general trend in the rest of the stations.

We can conclude that concentration peaks are more pronounced in the evening particularly between 18:00 and 20:00 and the ATMO-Street model chain is not able to reproduce these high values. The morning peaks are much less pronounced which contradicts the daily time emission factors (Figure 6) where morning peak emissions and evening peaks have similar maximum values.

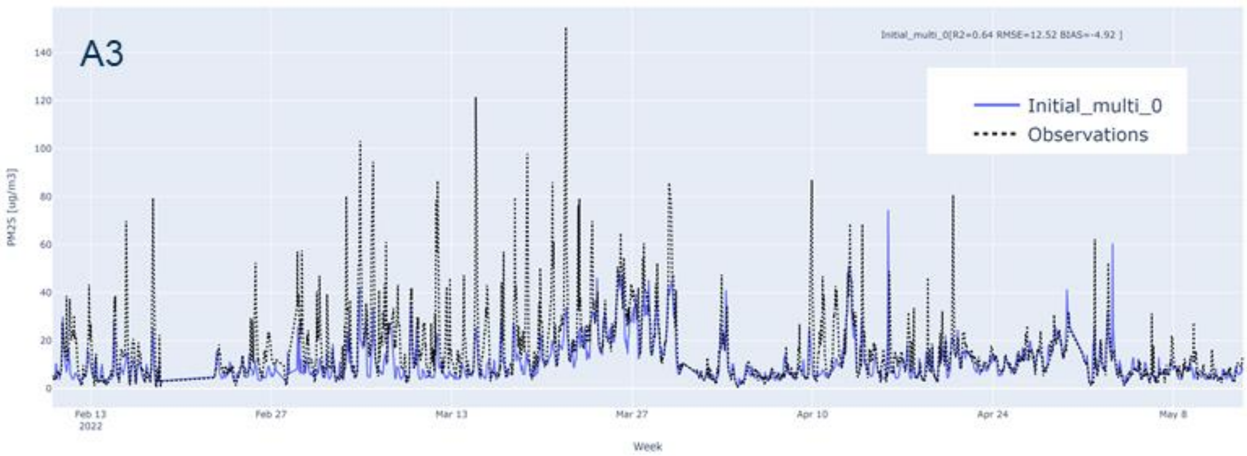


FIGURE 11 COMPARISON BETWEEN THE TIME SERIES OF OBSERVATIONS AND MODELLED RESULTS FOR STATION A3

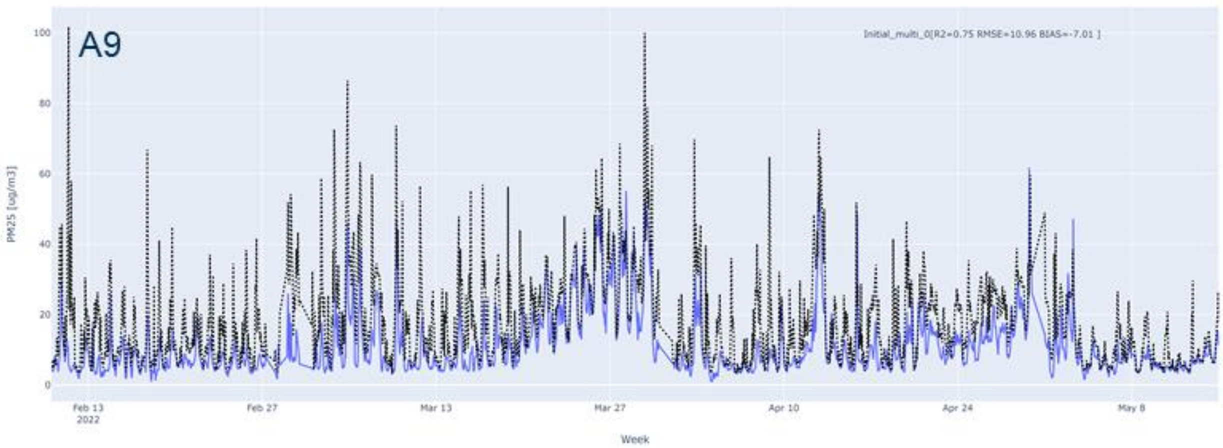


FIGURE 12 COMPARISON BETWEEN THE TIME SERIES OF OBSERVATIONS AND MODELLED RESULTS FOR STATION A9

The resulting average $PM_{2.5}$ map during the period of the experimental campaign is shown in Figure 13. Due to the low resolution of the initial emission grid, the resulting map has a limited spatial gradient.

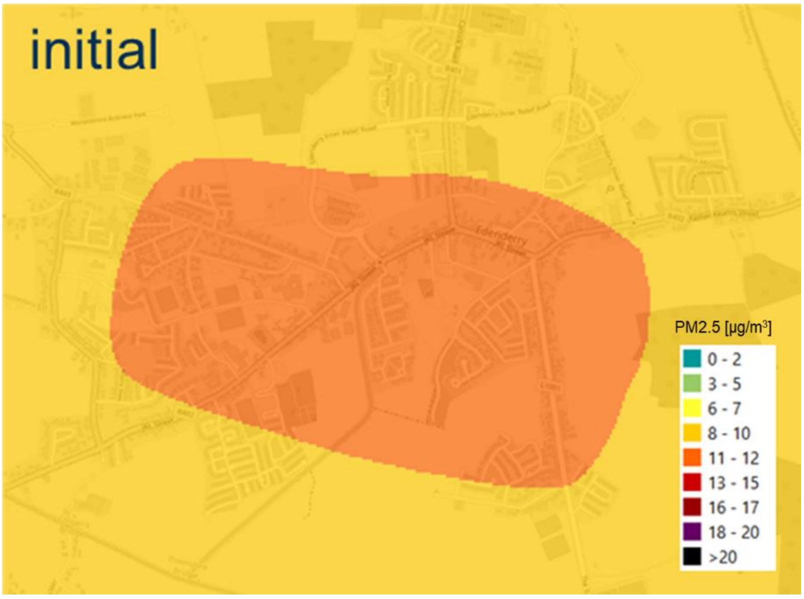


FIGURE 13 INITIAL $PM_{2.5}$ CONCENTRATIONS($\mu g/m^3$)

3.2.2 Results of the data assimilation procedure

In this section we analyse the results of the data assimilation procedure. Data was assimilated from all the sensors, except for A7 which was instead used as background values. As shown in Figure 14 below, the initial predictions (time series in blue) are unable to follow the observed concentration particularly during peak events. Following the data assimilation step the updated concentrations, shown in red, deliver a more accurate representation of the concentration peaks.

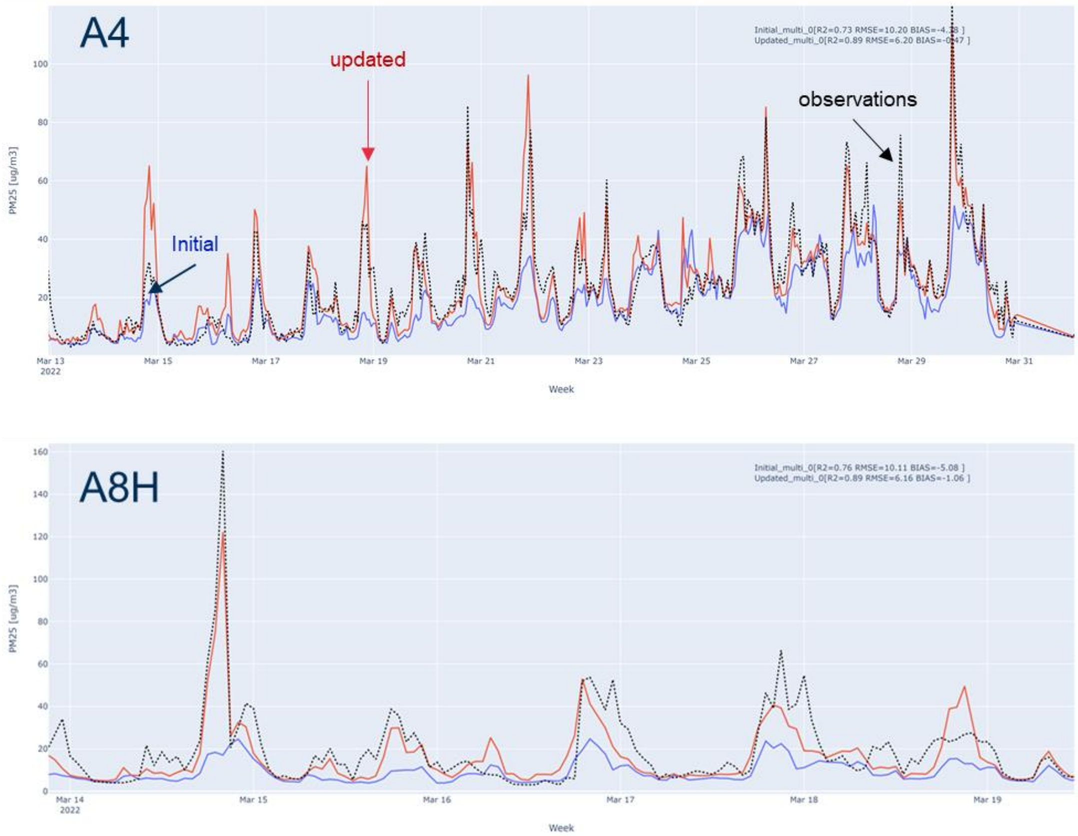


FIGURE 14 TIME SERIES OF CONCENTRATIONS BEFORE (IN BLUE) AND AFTER THE CORRECTION STEP (IN RED) FOR STATION A4 (TOP) AND A8H (BOTTOM)

In terms of statistical performance metrics, the R^2, the RMSE and the Bias are summarized in Figure 15. For all the stations there is an increase in correlation with the observations and a decrease of the random error and bias. However, for the sensor close to the fire station, the updated concentration prediction remains rather poor with a low correlation with observations and high RMSE and bias.

FIGURE 15 STATISTICAL CHARACTERIZATION OF THE INITIAL AND UPDATED CONCENTRATION TIME PROFILES.

The time profile of the fire station sensor is shown in Figure 16. The pattern is different from the surrounding stations and therefore the updated emission value in this neighbourhood appears to be mainly influenced by the remaining stations. This situation indicates that there might be a very local source influencing this sensor that is not representative of the rest of the area. Consequently, the data assimilation methodology used discards the contribution from this sensor.

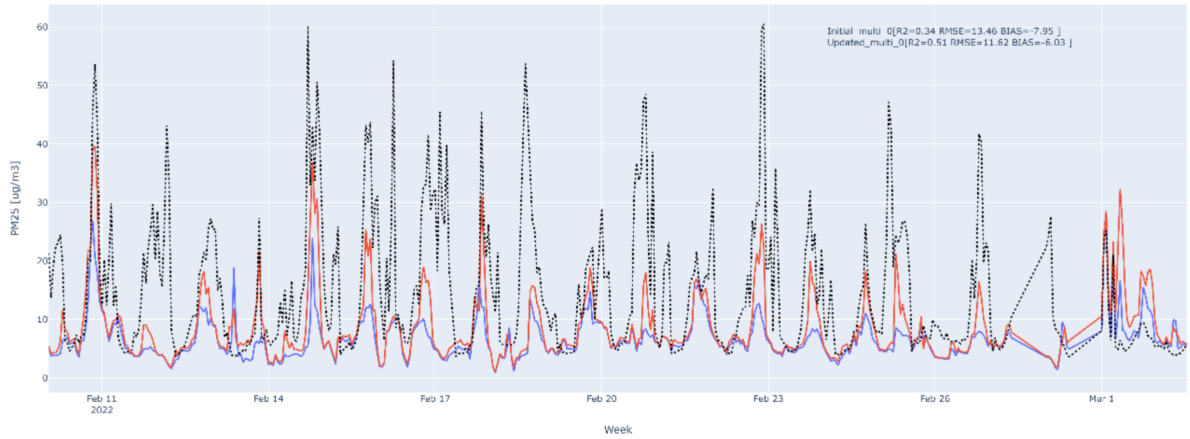


FIGURE 16 CONCENTRATION TIME PROFILE OF THE FIRE STATION SENSOR

3.2.3 Spatial Pattern

Finally, the average $PM_{2.5}$ concentration map (between 8 February to the 14 of March) before and after the correction step is shown in Figure 17. The updated map shows a stronger spatial pattern with increasing higher values towards the centre of Edenderry with a maximum average value above 10 µg/m³.

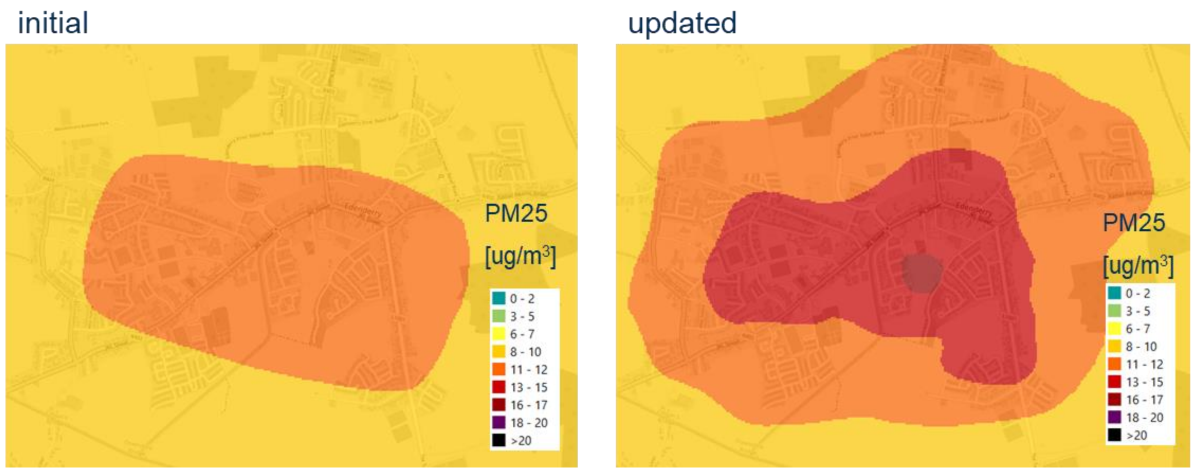


FIGURE 17 $PM_{2.5}$ CONCENTRATION MAP BEFORE (LEFT) AND AFTER (RIGHT) CORRECTION STEP

As shown in Figure 18 the new residential emission map has been limited to cells that overlap a built environment and divided into different neighborhoods. Besides higher absolute values, stronger emission gradients are also present in the new grid.

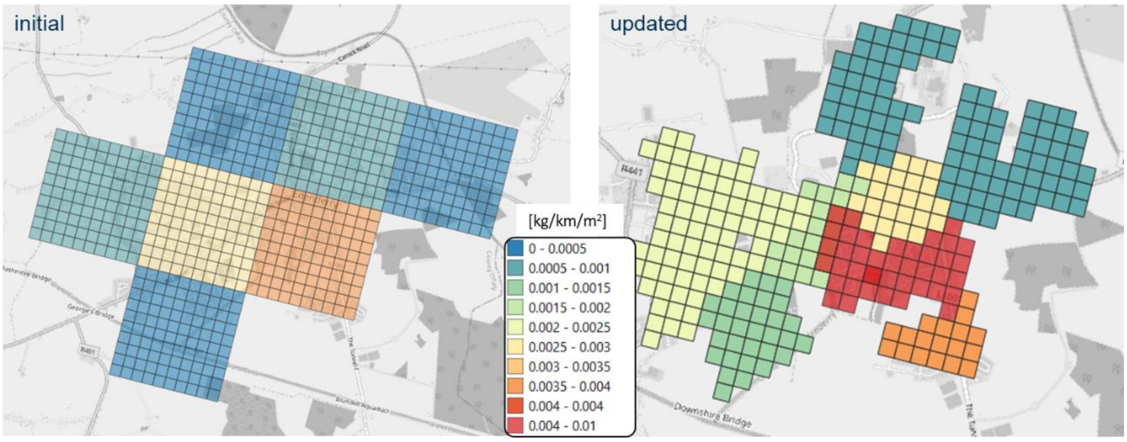


FIGURE 18 $PM_{2.5}$ EMISSION MAP BEFORE (LEFT) AND AFTER (RIGHT) CORRECTION STEP

Figure 19 depicts the initial and the updated emission values in function of the months. Since measurements were only performed from February to May updated emissions are not available for the remaining months. Depending on the month the average ratio varied between 2.6 (in March) and 1.2 (in May). Another relevant fact was a much stronger decline in emissions from the winter to the spring months. Initially the ratio between February and May was c. 1.9 while in the updated version it is c. 4.

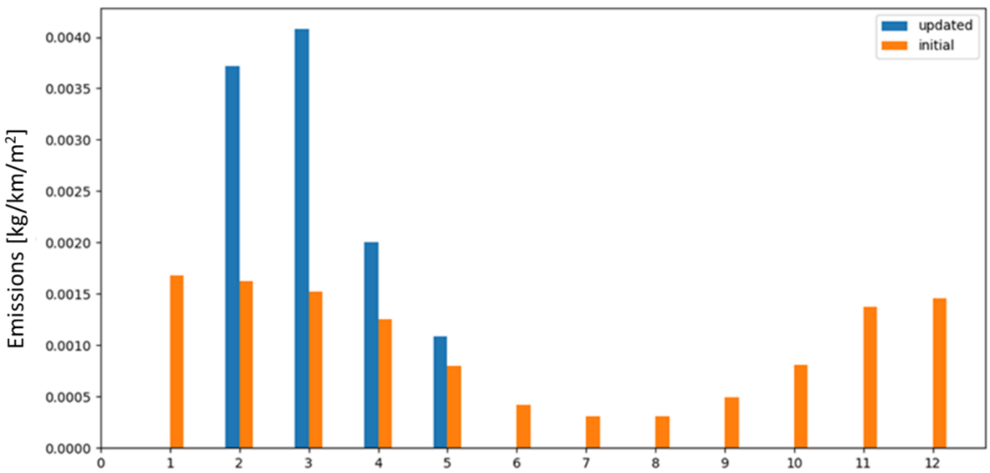


FIGURE 19 TOTAL EMISSIONS FOR THE NEIGHBOURHOOD WITH THE HIGHEST EMISSIONS BEFORE AND AFTER THE DATA ASSIMILATION CORRECTION.

3.2.4 Temporal Pattern

In Figure 20 the initial and updated time factors have been aggregated according to the hours of the day. As seen in Figure 18 there is an increase in the absolute emissions by a factor 2 in the zone with the highest emission value. The morning time factor peak (7 and 8 am) is about 40% smaller than initially expected, while the afternoon appears to be smaller at 18:00

reaching a similar value at 19:00 but remaining higher than expected until 23:00. The early morning hours and mid-afternoon hours remain similar in terms of time factors.

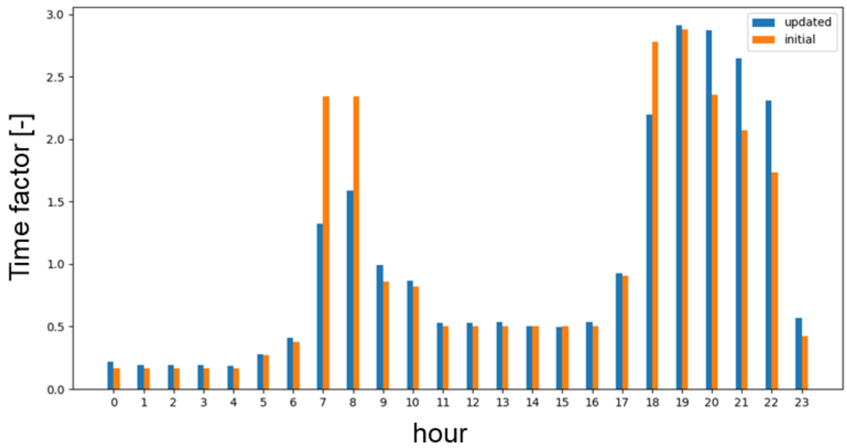


FIGURE 20 INITIAL AND UPDATED TIME FACTORS FOR THE DAILY PROFILE.

Figure 21 shows that the weak pattern is rather uniform in accordance with the original profile, with a variance below 10%. Given the small variance over the weekdays and the few numbers of tested months, we assume this variance is within the uncertainty of the methodology.

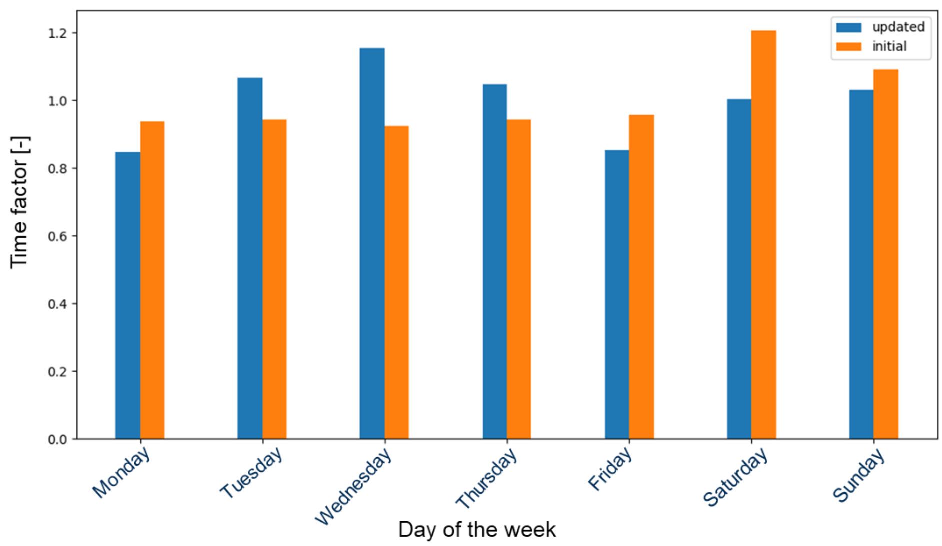


FIGURE 21 INITIAL AND UPDATED TIME FACTORS FOR THE WEEKLY PROFILE.

Contrary to the initial setup, the yearly pattern indicates that March was the month with the highest emissions according to Figure 22. This could be explained by the fact that in March the minimum temperature was around –2 ºC, while in February no negative temperature was observed according to the weather forecast model.

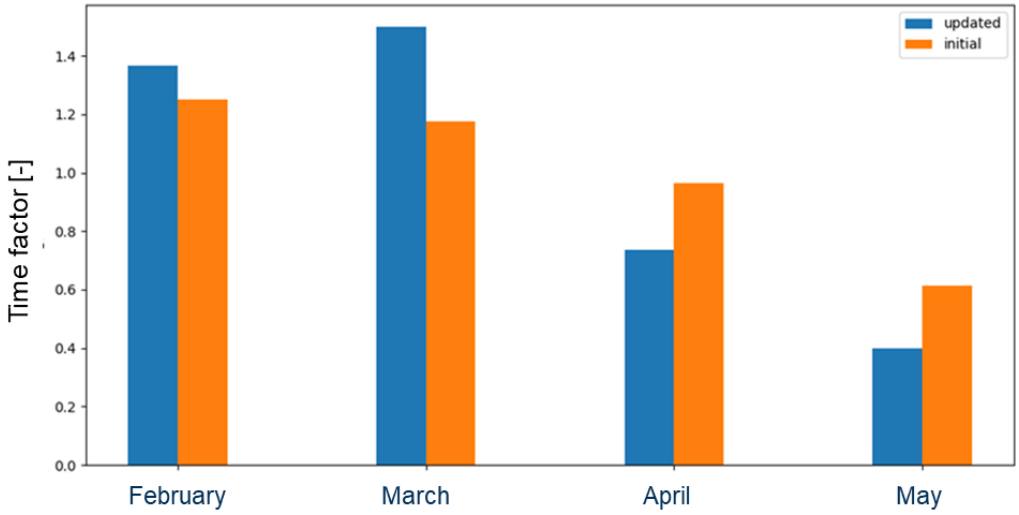


FIGURE 22 INITIAL AND UPDATED TIME FACTORS FOR THE YEARLY PROFILE.

4 Chapter 4 – Test case: Dungarvan

A second test campaign has been performed by University College Cork in Dungarvan during the winter of 2022/2023. For this campaign the lessons learnt from the Edenderry campaign were applied. The locations of the sensors have been better distributed according to the neighbourhoods, allowing for a better representation of the full town. Additionally, some prior information was available based on a survey which corroborated the findings associated with the data assimilation.

4.1 Experimental campaign and survey

A survey was also deployed by the University College Cork where different questions were asked associated with the buildings, stove age and human behaviour surrounding the use of residential heating. Each user participation survey was associated to a region within Dungarvan, as indicated in Figure 23, allowing for a spatial location of each answer. A total of 212 participants answered in this survey. Since the statistical significance might not be sufficient for quantitative results, we used these results for qualitative purposes.

FIGURE 23 REGIONS IN DUNGARVAN DEFINED IN THE SURVEY

For the current analysis the most relevant questions in survey were the following:

- **In which area of the map is your house located. Select numbered area from the map?**
- **What appliances do you use to heat your home? Select all that apply and rank their importance.**
- **If you use a solid fuel stove/appliance, how old is it?**
- **How old is the house/building you live in? If you don't know please select your best guess.**
- **What time of day do you normally burn solid fuel? Select all that apply.**
- **If your home has central heating, what time do you usually have it on?**
- **What time of day do you normally burn solid fuel?**

Following these questions, the answers are grouped according to the regions in the map. The analysis is mainly focused on Building age, Stove age and Heating choice. Figure 24 to Figure 27 display the answers for region 1, 2, 3 and 4. According to the available answers, Regions 1 and 2 appear to have older stoves and buildings. It is noteworthy that stoves that were more than 30 years old have only been identified in regions 1 and 2. Furthermore, according to the answers most of the oldest buildings (more than 30 years old) are located in region 2.

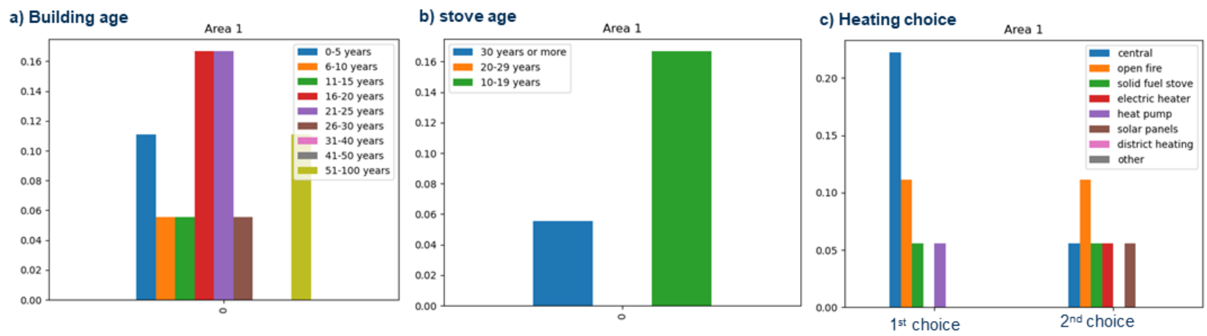


FIGURE 24 A) BUILDING AGE B) STOVE AGE C) PRIMARY AND SECONDARY HEATING CHOICE IN AREA 1

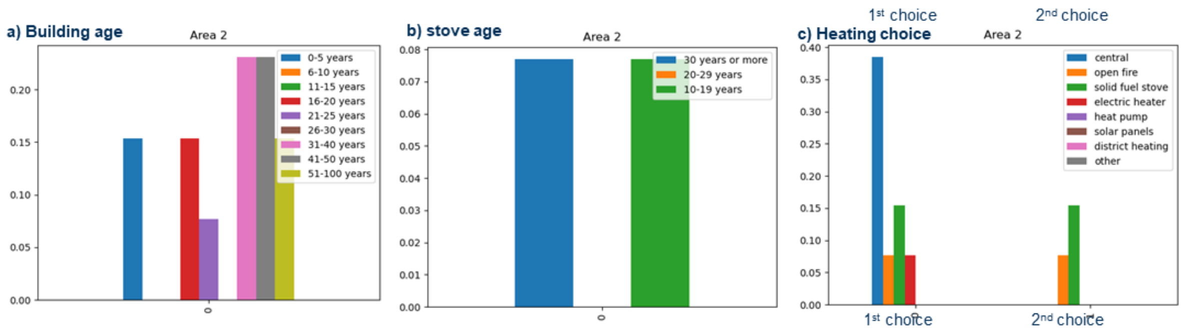


FIGURE 25 A) BUILDING AGE B) STOVE AGE C) PRIMARY AND SECONDARY HEATING CHOICE IN AREA 2

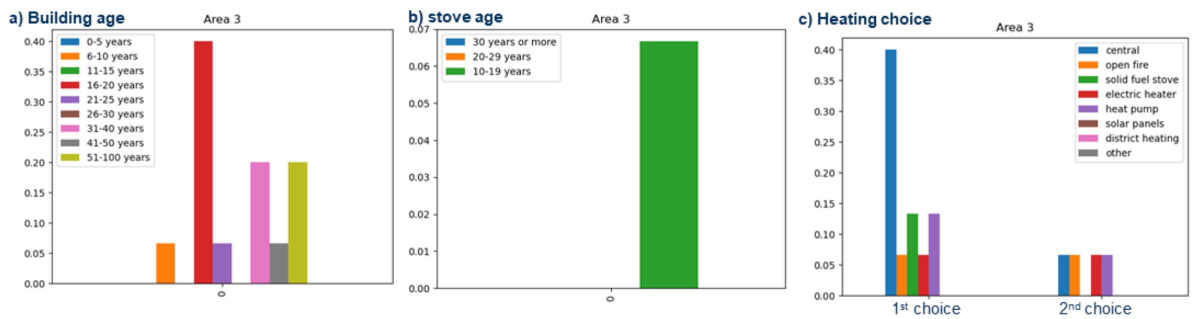


FIGURE 26 A) BUILDING AGE B) STOVE AGE C) PRIMARY AND SECONDARY HEATING CHOICE IN AREA 3

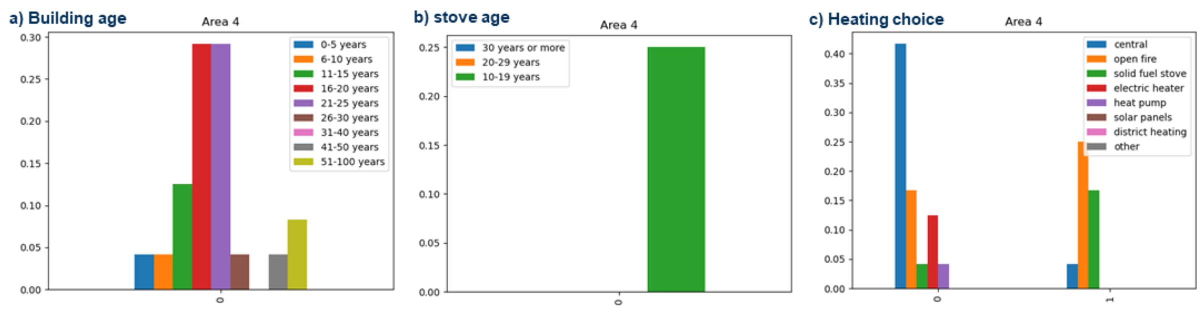


FIGURE 27 A) BUILDING AGE B) STOVE AGE C) PRIMARY AND SECONDARY HEATING CHOICE IN AREA 4

As seen in Figure 28 the usage of central heating is predominantly in the early morning and evening. However, a different conclusion can be extracted for solid fuel burning, where clearly the preferred usage time is in the evening.

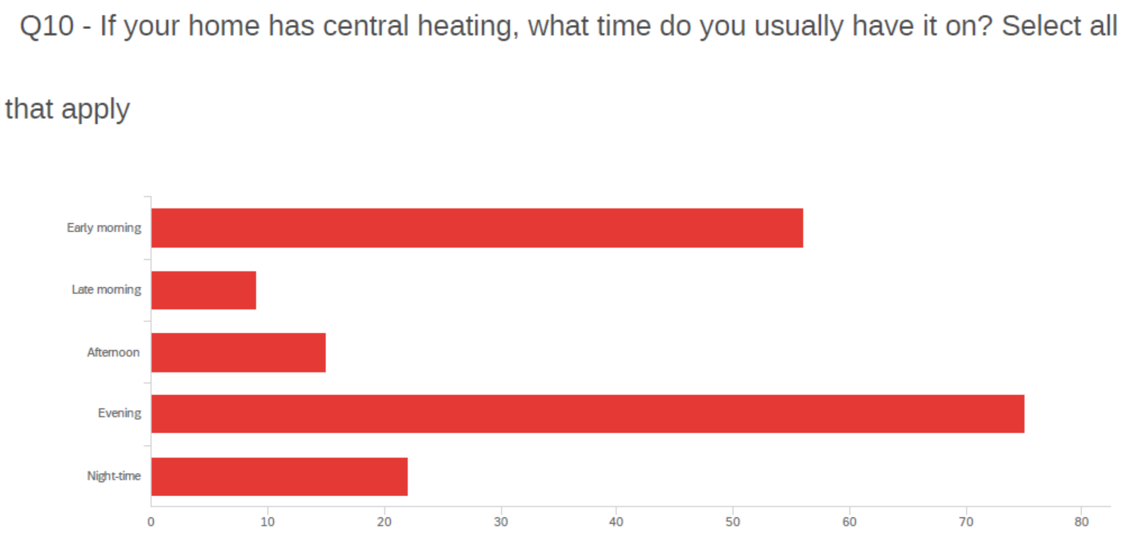


FIGURE 28 TIME PROFILE ACCORDING TO THE SURVEY OF CENTRAL HEATING USAGE

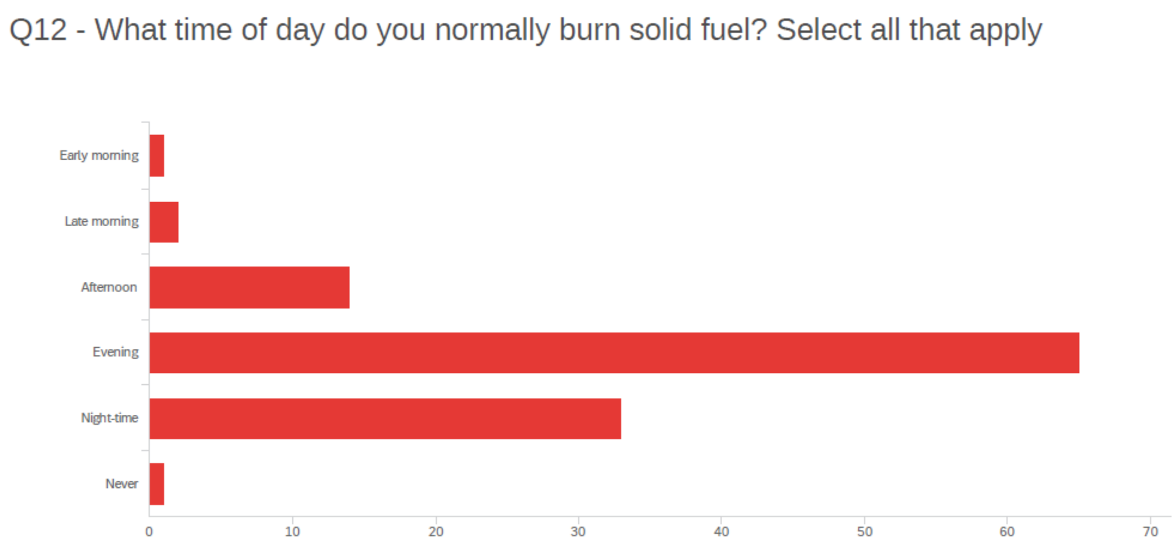


FIGURE 29 TIME PROFILE ACCORDING TO SURVEY OF SOIL FUEL USAGE

In total 26 $PM_{2.5}$ sensors were deployed in Dungarvan covering a wide region of Dungarvan as displayed in Figure 30.

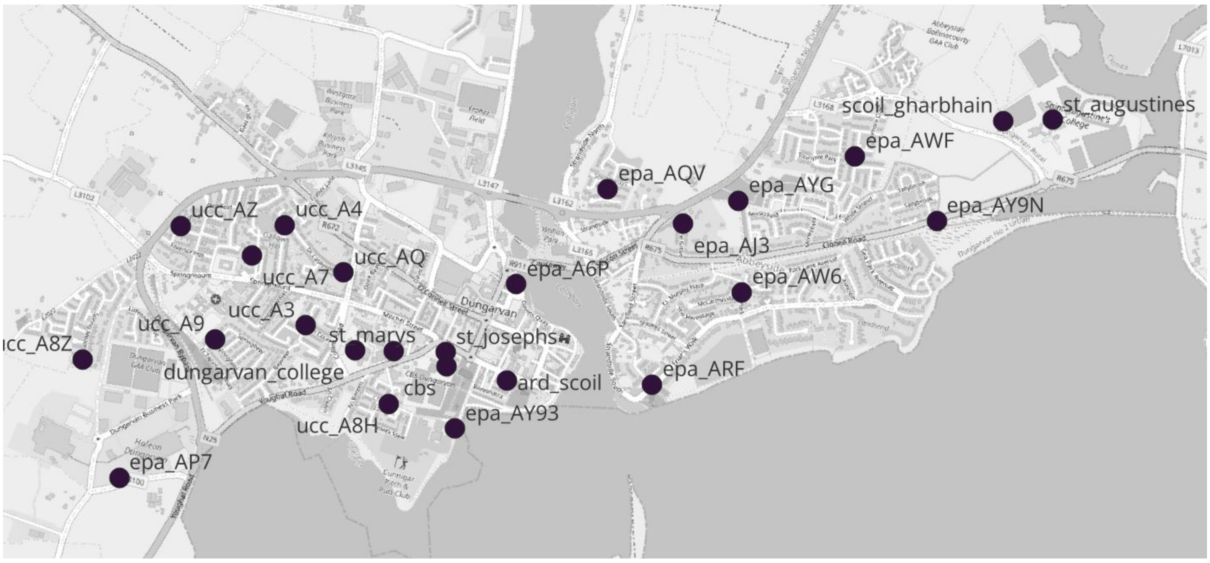


FIGURE 30 LOCATION OF THE SENSORS IN DUNGARVAN

4.2 Initial model evaluation

The initial ATMO-Street setup used the MapEIre emission spatial pattern as seen in Figure 31 with the same time factors as used in Edenderry. The average modelled concentration map shows limited spatial variability with values below 10 µg/m³. However, the measured average values (dots in Figure 31) are significantly higher, with values above 18 µg/m³ and with a stronger spatial pattern. The most westerly and easterly stations observed the lowest concentrations, and therefore the urban canopy lying between shows as a hotspot.

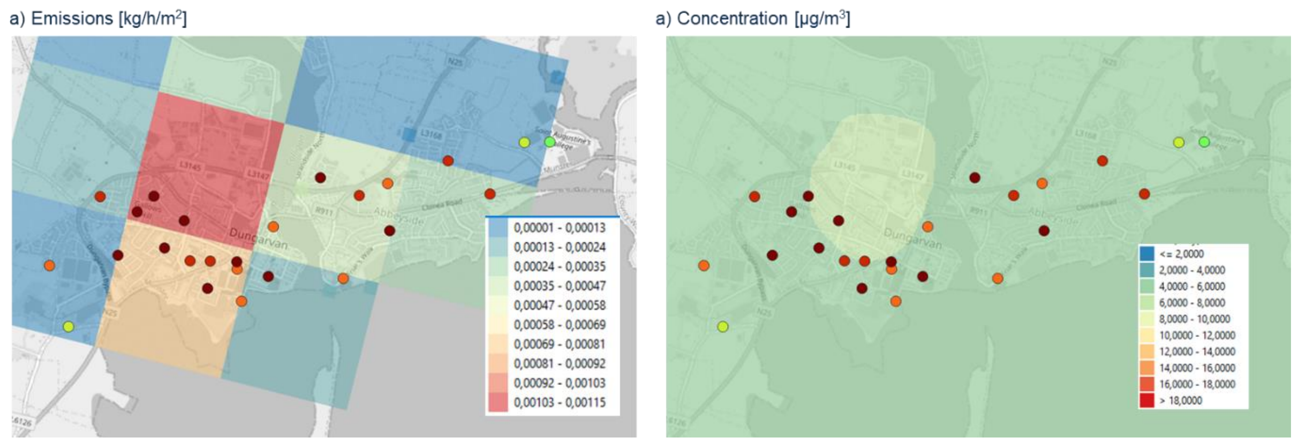


FIGURE 31 A) EMISSION PATTERN USED IN THE INITIAL ATMO-STREET RUN B) RESULTING CONCENTRATION AVERAGE OVER THE EXPERIMENTAL CAMPAIGN PERIOD. THE DOTS REPRESENT THE CONCENTRATIONS MEASURED BY THE SENSORS. THE SAME COLOUR SCALE IS USED FOR THE MODELLED VALUES AND THE SENSOR OBSERVED VALUES.

The underestimation of the initial modelling is corroborated by the bias for each station as seen in Figure 32. A negative bias is represented for all the stations where is some cases the values are as negative as -20 µg/m³.

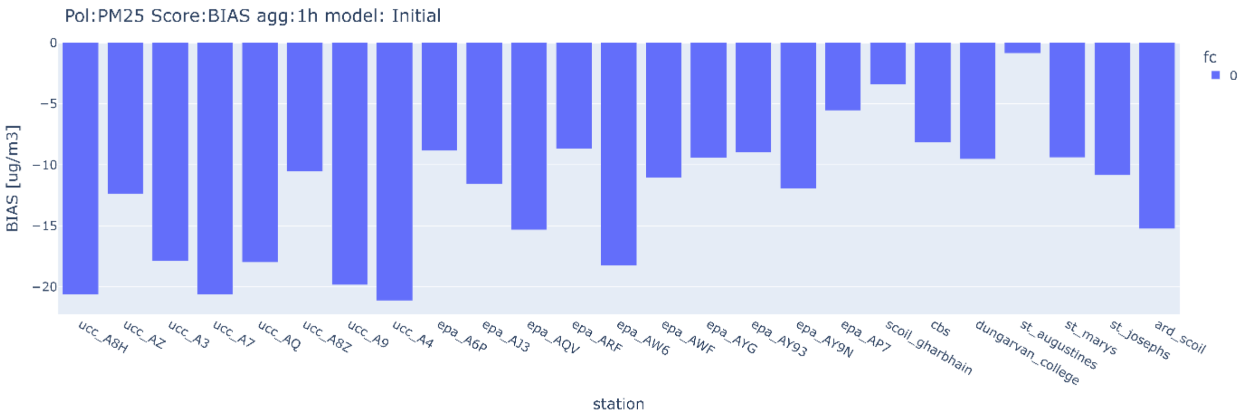


FIGURE 32 BIAS FOR EACH STATION OF THE INITIAL ATMO-STREET RESULTS

Figure 33 and Figure 34 depict the time series profile for 2 randomly selected stations. There is a model underestimation over the full period, which is especially significant during the evening peaks, corresponding to the time where most solid fuel was burned according to the survey.

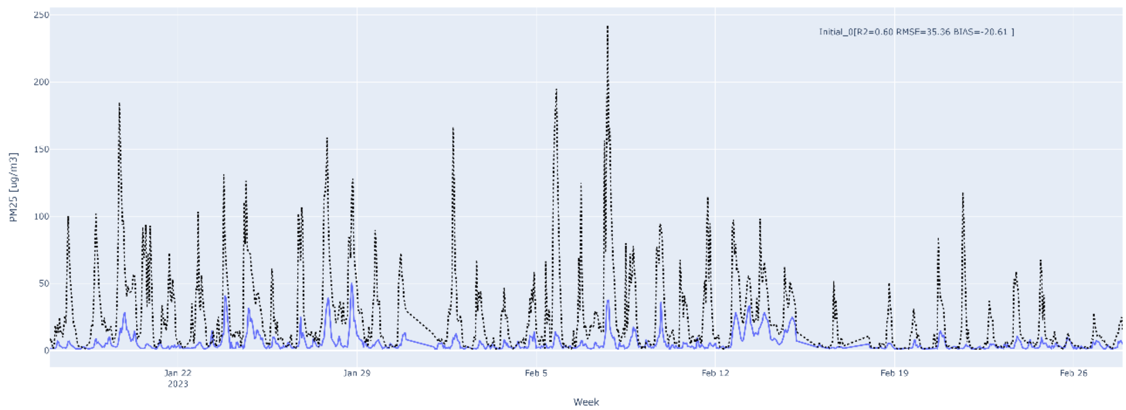


FIGURE 33 TIME SERIES OF THE MEASURED AND INITIAL PREDICTED (BLUE) CONCENTRATION FOR STATION A8H

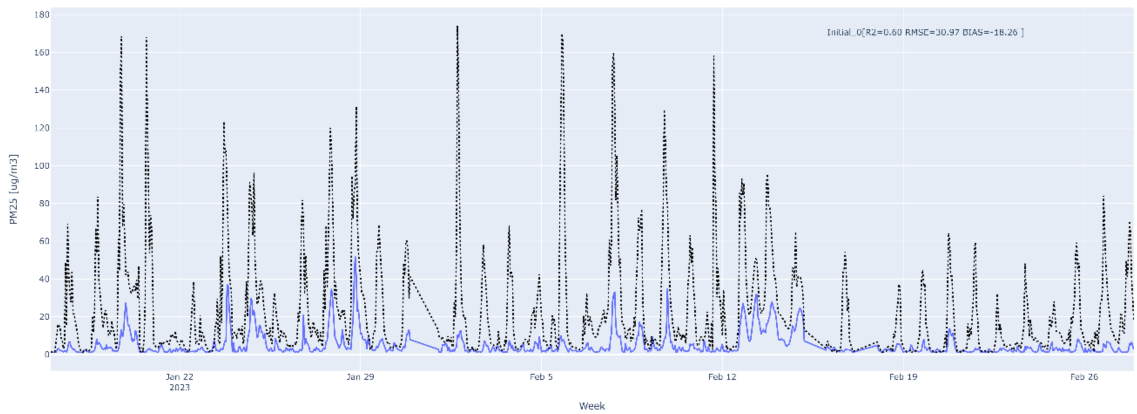


FIGURE 34 TIME SERIES OF THE MEASURED AND INITIAL PREDICTED (BLUE) CONCENTRATION FOR STATION AW6

4.2.1 Results of data Assimilation

In order to address this negative bias, the data assimilation was applied to the Dungarvan case following the same approach used in Edenderry. As shown in Figure 35 the emissions have been spatially split into the different regions as applied in the survey in Figure 23.

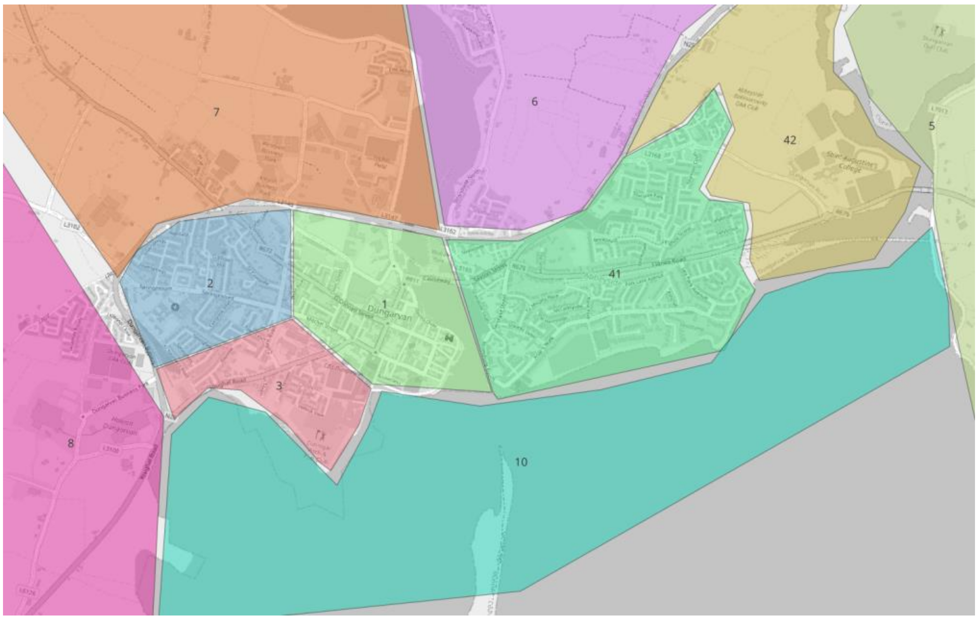


FIGURE 35 POLYGONS REPRESENTING THE AREAS BY WHICH THE EMISSION HAVE BEEN SPLIT

The results from the data assimilation step are show in Figure 36. The updated winter average emission values are shown in Figure 36 a), where higher emission values are predicted with stronger spatial patterns. Region 2 shows the highest emission value. In terms of concentrations as shown in Figure 36 b), the average peak value reached up to 18 ug/m^3 and shows a closer match between the measured and predicted $PM_{2.5}$ concentrations.

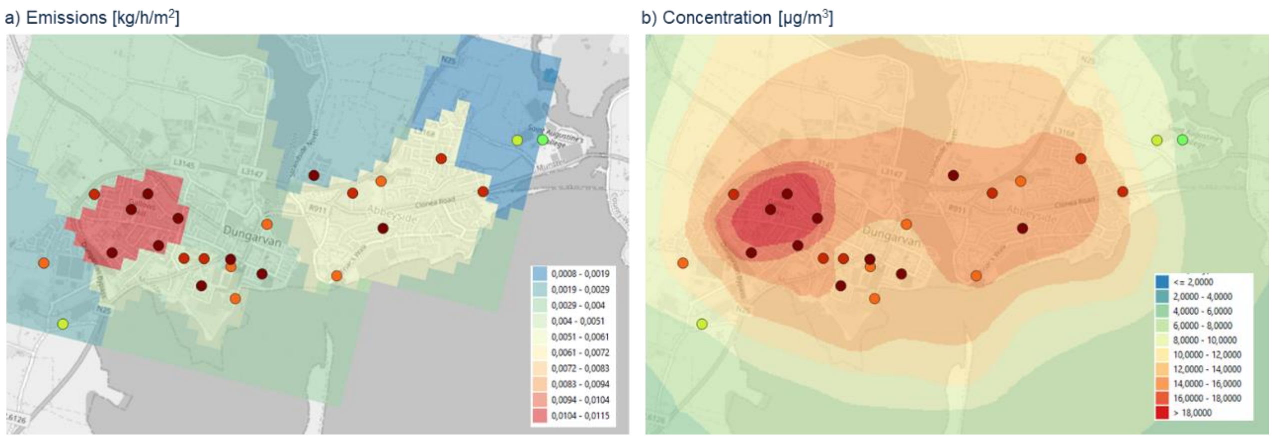


FIGURE 36 A) UPDATED EMISSIONS AFTER THE DATA ASSIMILATION STEP B) THE RESULTING UPDATED CONCENTRATIONS RESULTING FROM THE DATA ASSIMILATION STEP

Figure 37 shows the absolute initial and updated emission values over the months where the data assimilation was successful. In December 2022 there were insufficient sensor data to perform the data assimilation step. In general, the updated emissions are about 10 times higher than the initial emissions.

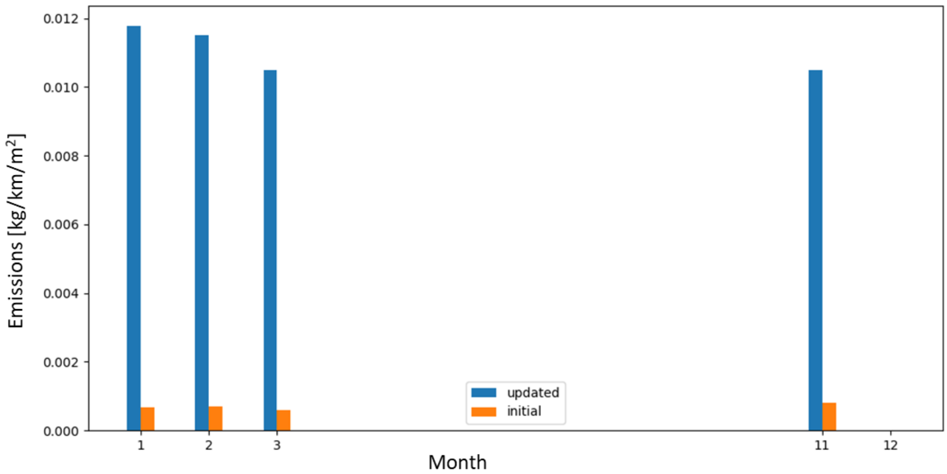


FIGURE 37 SPATIAL AVERAGE INITIAL AND UPDATED EMISSIONS OVER THE DIFFERENT MONTHS

For these months the new time factors can be extracted as shown in Figure 38. As observed in Edenderry, the morning peak is significantly reduced relative to the evening emission peak. This is, however, more evident in Dungarvan. This is also in line with the solid fuel burn time profile provided by the survey as shown in Figure 29.

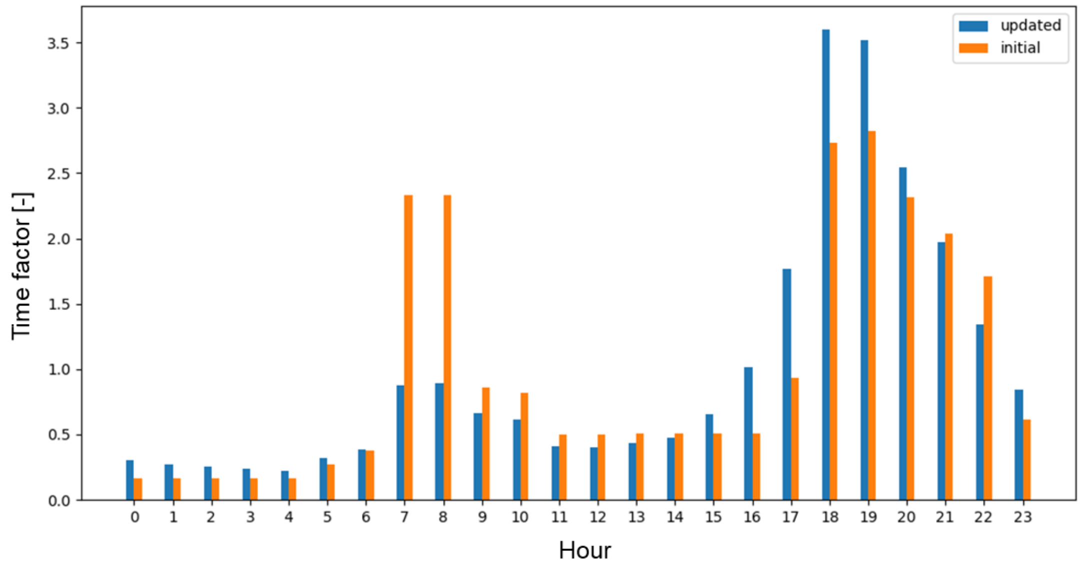


FIGURE 38 ORIGINAL AND UPDATED HOURLY TIME FACTORS

The weekly profile shows few variations with +/- 20% across the week. The MapEIre profile peaks on the weekends, while the updated profile peaks on a Wednesday. However, such small variation can be within the uncertainty associated with the data assimilation step and might be influenced by specific meteorological conditions during the measurement period.

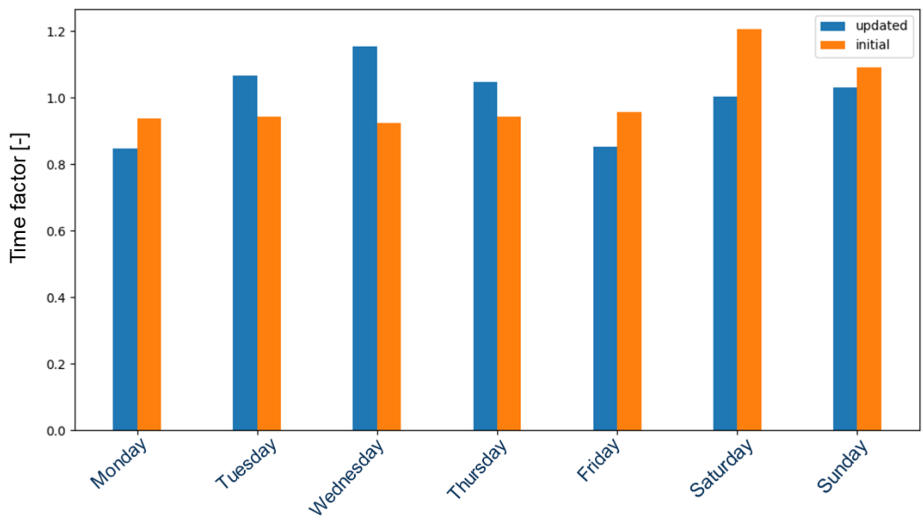


FIGURE 39 INITIAL AND UPDATED WEEKLY TIME FACTORS

4.2.2 Validation with updated emissions and time factors

Since no other measurement is available in Dungarvan, the best validation available for the updated time profiles and emission maps is to re-run the same period but now relying on the updated mean emissions and time profiles. In this step no data assimilation is carried out and only the ATMO-Street inputs have been modified. The resulting statistics are shown in Figure 40. When using the updated ATMO-Street parameters the RMSE decreases by about 22%, while the bias was reduced by about 75%.

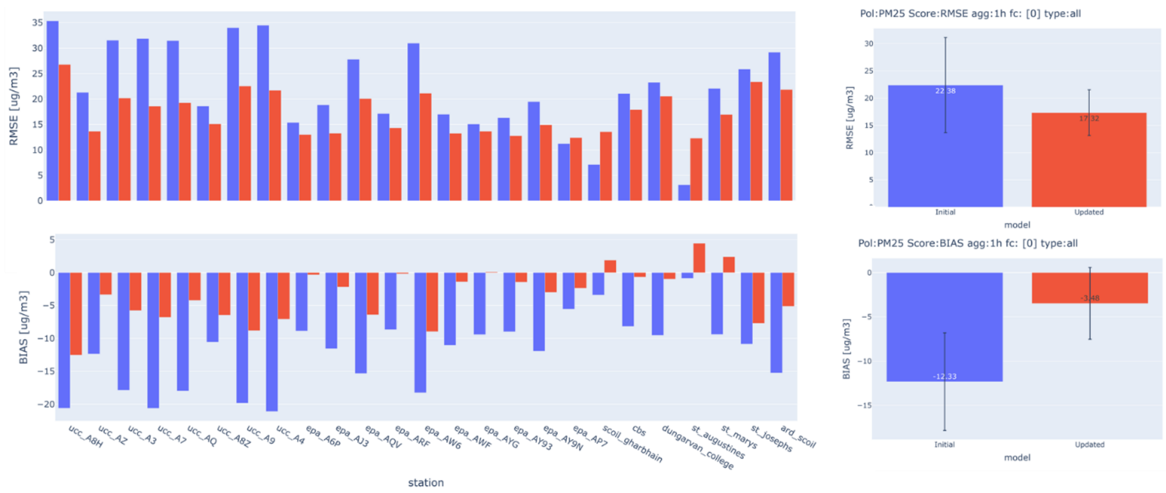


FIGURE 40 INITIAL AND UPDATED STATISTICS (RMSE, BIAS): LEFT - FOR EACH STATION; RIGHT - AVERAGE OVER ALL STATIONS

In terms of the FAIRMODE[8] assessment target plot shown in Figure 41 the MQI has improved from 1.042 to 0.859. Hence, the update parameters in the model allow it to achieve the model quality objective set by FAIRMODE.

[8] FAIRMODE is an initiative aimed at improving air quality modelling in Europe by fostering collaboration, harmonizing methodologies, and providing tools and guidance for model evaluation and intercomparison.

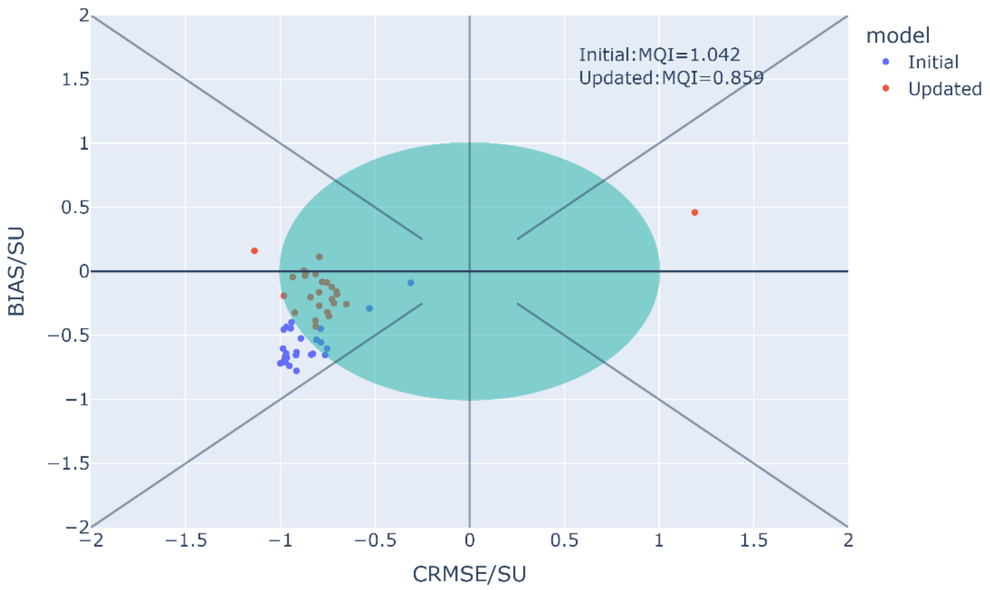


FIGURE 41 ASSESSMENT TARGET PLOT FOR THE INITIAL AND UPDATED ATMO-STREET PREDICTIONS

In Figure 42 to Figure 46 the observed hourly concentrations are compared with the initial and the updated predictions. Overall, there is clearly an improvement in the ability to predict the daily peak occurring typically in the evening. Furthermore, not only is the peak value better predicted, but also the remaining hours with lower values. This improvement is observed for all the stations.

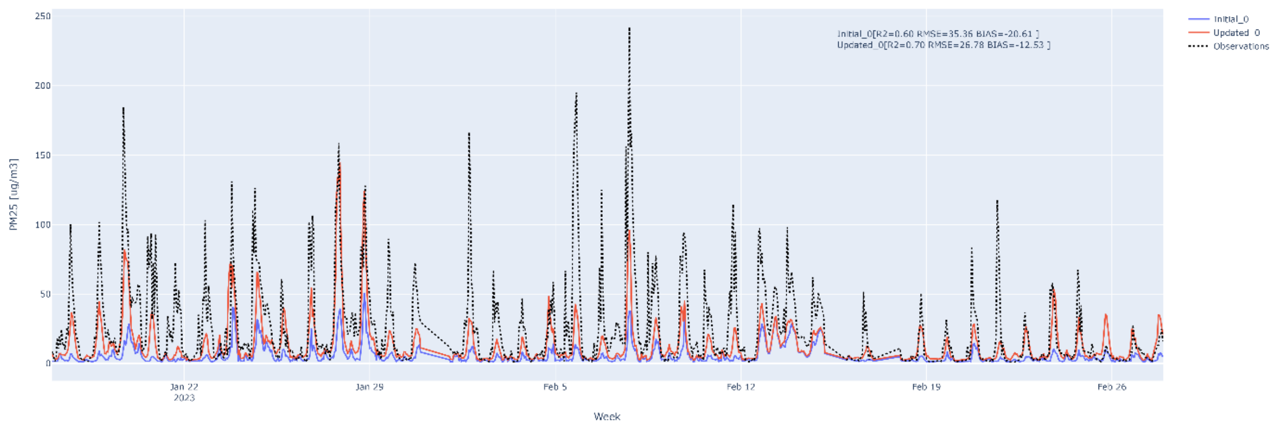


FIGURE 42 COMPARISONS BETWEEN OBSERVATIONS WITH INITIAL AND UPDATED TIME SERIES PREDICTIONS A8H

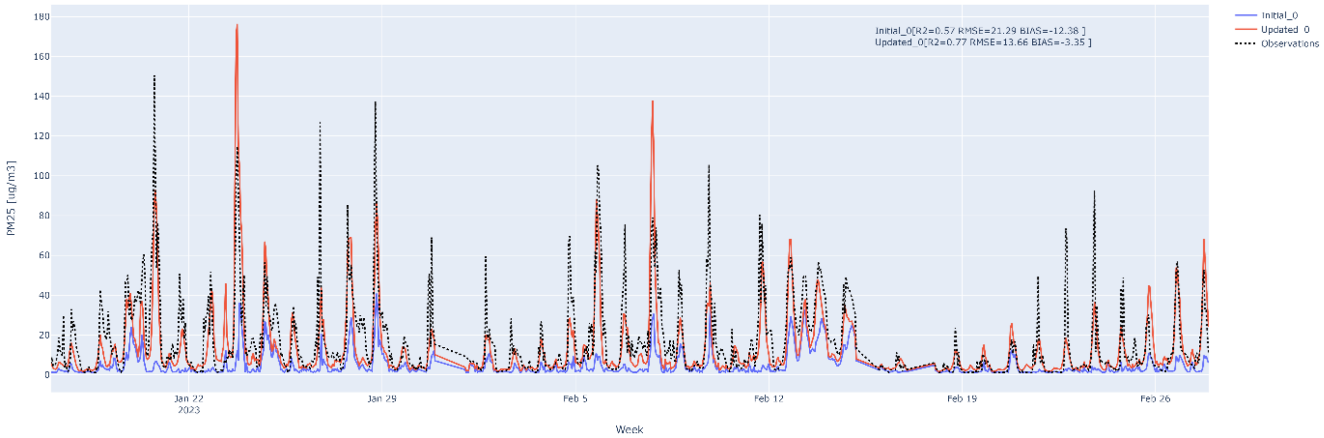


FIGURE 43 COMPARISONS BETWEEN OBSERVATIONS WITH INITIAL AND UPDATED TIME SERIES PREDICTIONS AZ

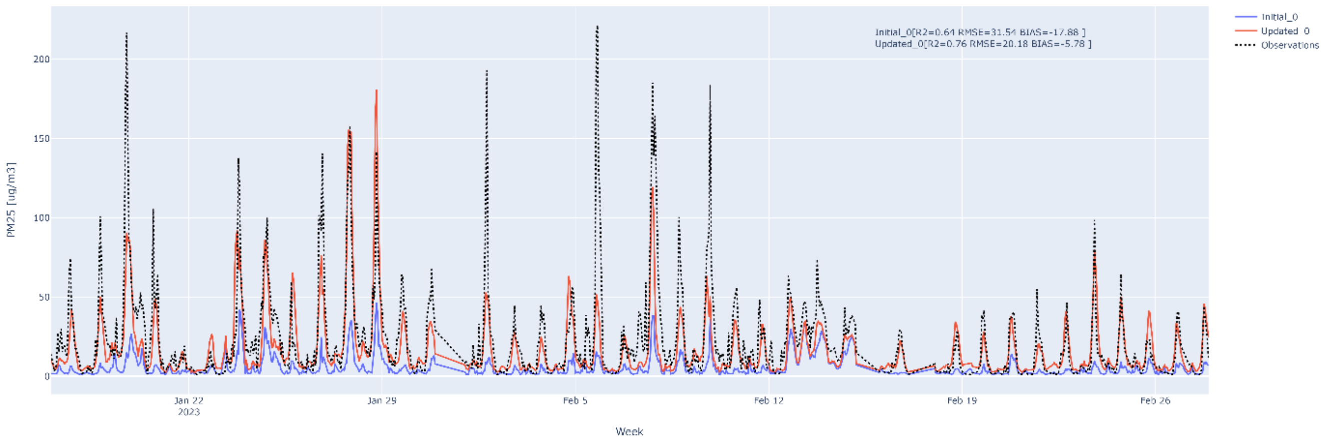


FIGURE 44 COMPARISONS BETWEEN OBSERVATIONS WITH INITIAL AND UPDATED TIME SERIES PREDICTIONS A3

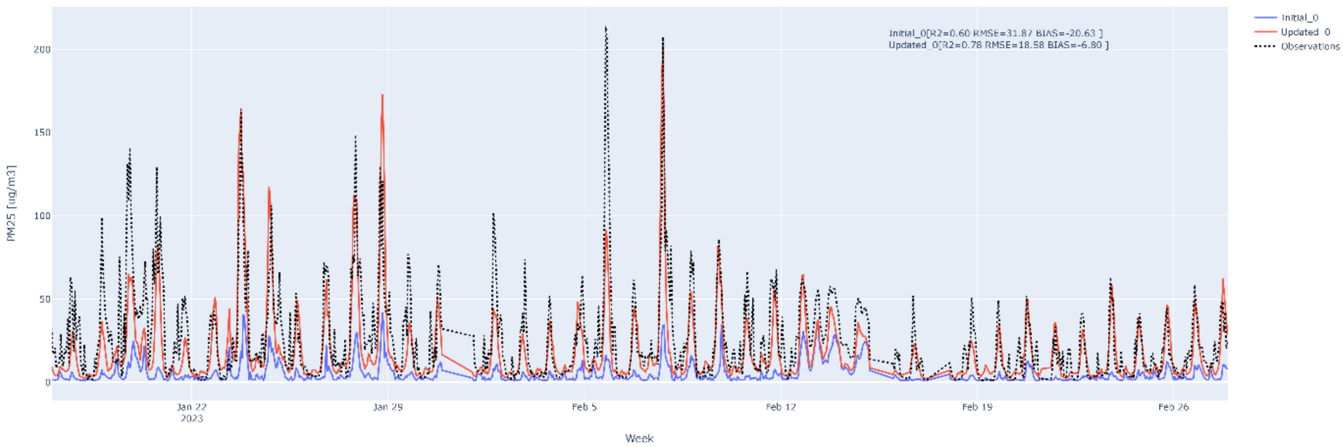


FIGURE 45 COMPARISONS BETWEEN OBSERVATIONS WITH INITIAL AND UPDATED TIME SERIES PREDICTIONS A7

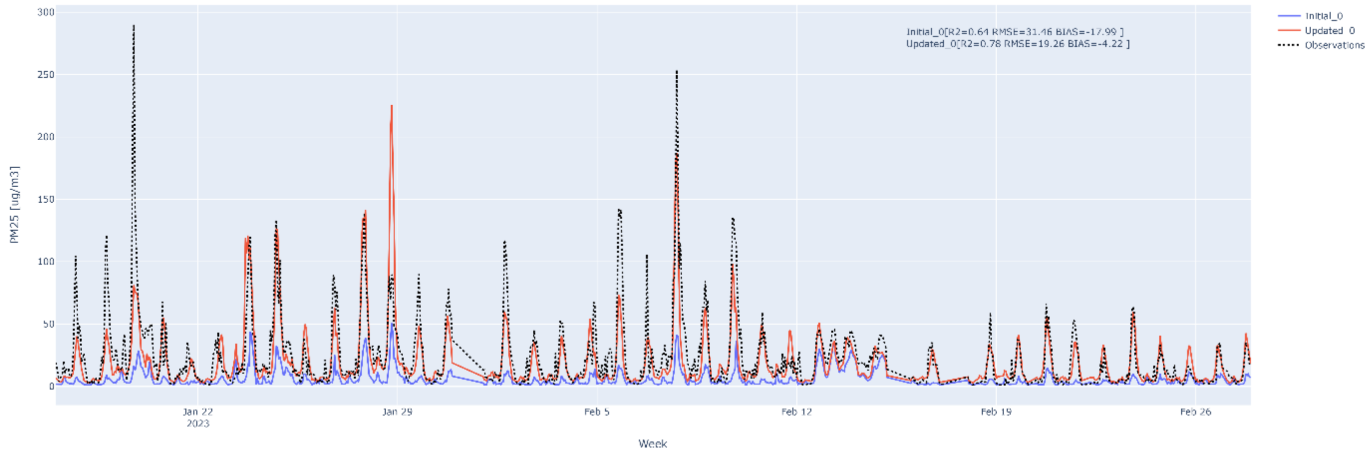


FIGURE 46 COMPARISONS BETWEEN OBSERVATIONS WITH INITIAL AND UPDATED TIME SERIES PREDICTIONS AQ

This validation exercise cannot be considered an independent validation because it covers the same period for which the data assimilation has been performed. However, it does not use any data assimilation step and relies solely on the updated input parameters. In particular, the updated annual average emission values per neighbourhood and the updated time factors. It proves the ability of the proposed methodology to generate a better emission map and time factors. It is, however, still not possible to conclude if these results can be generalized to other regions in Ireland and to other time frames i.e., other months of the year, other years etc. It does, however, highlight where the uncertainties are in the MapEIre dataset when applied to these two towns and it demonstrates that it would be beneficial to have a higher resolution residential emission data to enhance the reliability of the concentration maps.

4.2.3 Proposed methodology for emission extrapolation

Ideally, we would like to be able to use the knowledge gained in these studies to downscale and improve the MapEIre dataset when used for modelling the PM concentrations across all of Ireland. Next, the potential methodology that was explored and tested on the Dungarvan study case is described. This methodology looks to exploit the high resolution statistical CSO[9] data that is available on building age and building density (number of buildings per m^2) across the country.

Figure 47 hereunder shows the modelled emissions together with the CSO data for building density, average building age and primary (central) heating source for the Dungarvan study case. The location with the highest emission appears to overlap with locations with the highest building density, the oldest buildings, and the specific central heating types.

[9] https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/

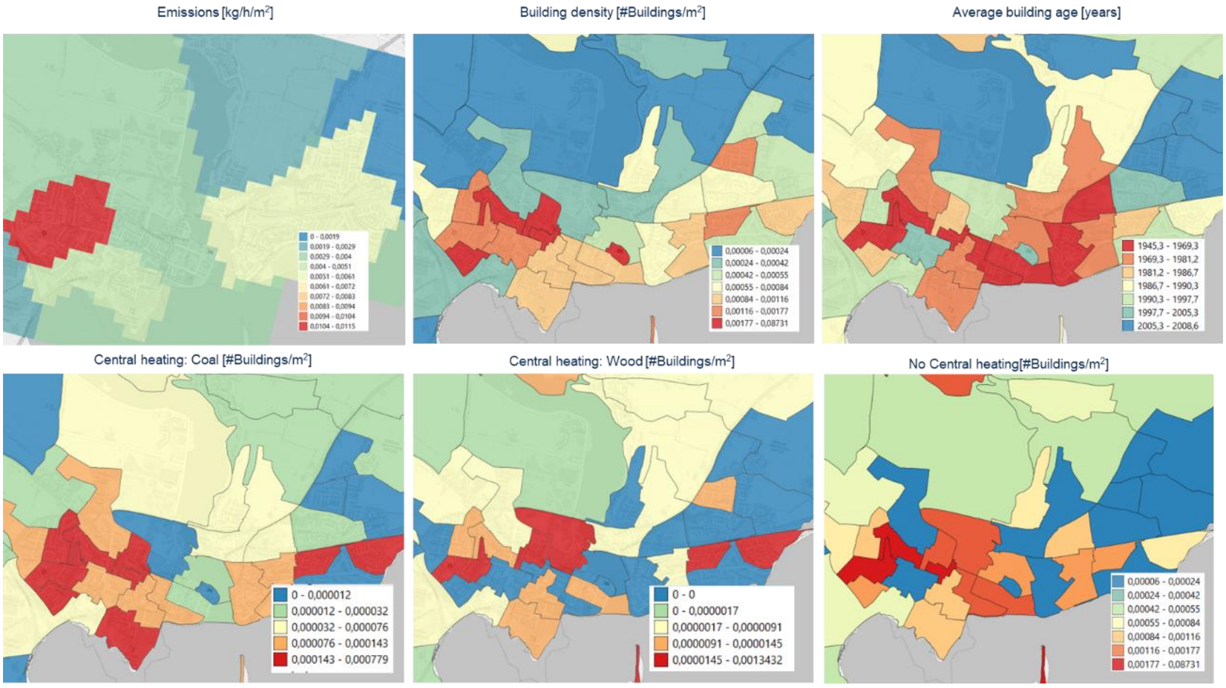


FIGURE 47 EMISSION MAP (TOP LEFT CORNER) AND DIFFERENT INFORMATION AVAILABLE IN THE CSO DATABASE.

To further demonstrate this correlation, Figure 48 shows the correlation matrix between emissions and some of the CSO variables. One can see that the estimated emissions have a negative correlation with building age (-0.35) and a positive correlation with house density (0.1), the usage of coal (0.49) and the No Central Heating variable (0.35).

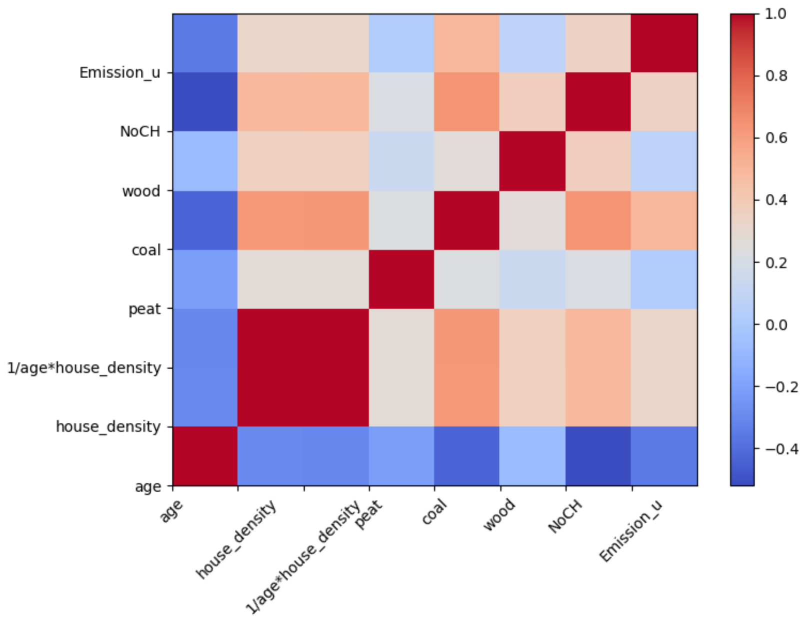


FIGURE 48 CORRELATION MATRIX BETWEEN ALL EMISSION AND THE CHARACTERISTICS IN EACH SMALL AREA

Hypothetically, these variables can therefore be used to train an emission model that estimates the emissions all over Ireland. Nonetheless, to train such a model accurately more spatial and temporal data would be required and in addition the statistical small area should be representative enough throughout the whole country. To investigate this possibility, a multi-linear regression model was trained with all the variables shown above. The CSO data was used for the whole country in terms of building age, building density, central heating sources and wood burning to extrapolate the predicted emissions to the whole country. The predicted winter residential emissions with this approach of this analysis are show in Figure 49.

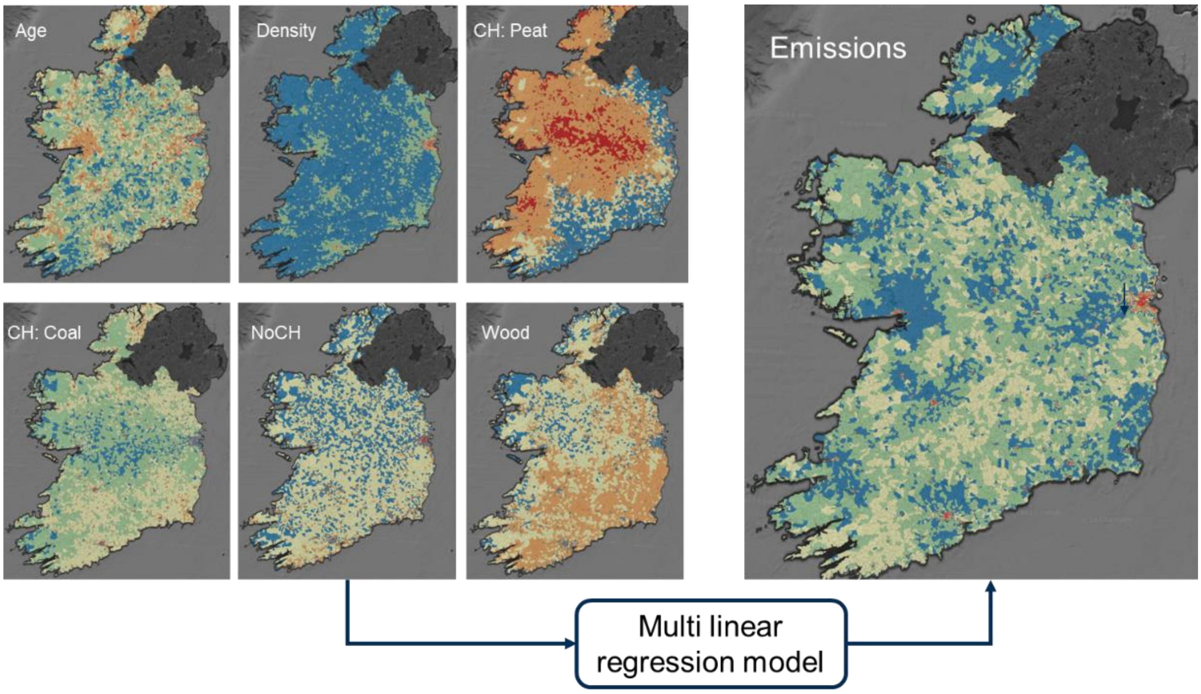


FIGURE 49 USAGE OF A MULTI LINEAR REGRESSION MODEL TRAINED WITH DUNGARVAN DATA TO EXTRAPOLATE EMISSIONS TO THE WHOLE COUNTRY

Since the data assimilation step has been only used during the winter months and covers a small sample, the extrapolated emissions can only serve as an example on how this emission extrapolation methodology could be undertaken. For this approach to have further reliability, longer experimental campaigns would be required, covering a wider range of the variable situation across the county.

5 Chapter 5 - Conclusions and Recommendations

5.1 Summary

Two experimental study cases were developed in Edenderry and Dungarvan where a data assimilation approach, using low-cost sensors, was used to correct the pollutant concentrations and residential emissions. A set of Purple Air and Clarity sensors were deployed by the University College Cork in both town centres during the winter of 2021-2022 (Edenderry) and the winter of 2022-2023 (Dungarvan). In tandem, building upon the ATMO-Street model configuration for Ireland, which uses the MapEIre residential emissions, a dedicated Gaussian dispersion model was set-up for this study, running on an hourly basis. For the data assimilation methodology, the emission grid was downscaled and grouped according to different neighbourhoods in which each of the 'neighbourhood' regions was assumed to have a uniform emission value. On an hourly basis, prior to the data assimilation step, a significant negative bias was observed in both cases. This statistical mismatch was used by the data assimilation methodology to correct the $PM_{2.5}$ emissions per neighbourhood.

This process was run over a period of 3 months for each case, which resulted in an hourly updated residential emission grid. The hourly corrected emission maps were averaged over the 3-month period. This resulted in a residential emissions grid with stronger, more representative spatial gradients and higher absolute values. In addition, the time profile was modified, where most notably the morning peak became less pronounced, and the afternoon peak was delayed by an hour.

Finally, an emission generalization methodology is described that could facilitate the extrapolation of the information taken at local scale over the whole country. This methodology uses the deduced emissions from the Dungarvan case and trains a multi linear regression model using CSO data as drivers. Once this model is trained, it could be used to update the emissions for the whole country. However, the accuracy of methodology depends upon the representativeness of the data assimilated deduced emissions. As we have only tested this methodology with data from two towns, it was concluded that further local exploration across the country is required before the extrapolation methodology could be applied to scale up the emissions for the whole country. Nonetheless, when more data becomes available (i.e., more localised measurement campaigns are set-up), extrapolation could become a viable solution to update the emission inventory.

With regards to future data assimilation exercises this study concluded the following:

- During this test case different approaches were explored in terms of downscaling the emissions. To reduce the number of unknowns for the inverse problem, the preferred methods assumed that the emissions are uniform over each neighbourhood. This approach proved to be the best in the light of the limited number of sensors.

- In future experimental and data assimilation campaigns it is therefore recommended to follow a similar setup, where a prior split of the neighbourhoods should be performed followed by a sensor deployment, where at least two sensors are placed in each region. This was concluded with the Edenderry test case and successfully applied in Dungarvan. The optimised distribution of the sensors has helped to minimize the uncertainty in the data assimilation step.
- At least one sensor of each kind should be collocated with the closest reference station (if available) in order to track sensor accuracy over the period of investigation. As no reference station was available in Dungarvan this was only possible in Edenderry.
- To better improve the prior knowledge of the residential emissions and their uncertainties, it is recommended to extract prior information on residential emissions through questionnaires and/or surveys. This could allow a better split between the different neighbourhoods according to residential housing construction type and household heating behaviour. Additionally, the survey could better inform the distribution across the time profiles. The Dungarvan survey was available prior to the modelling exercise. However, due to the uncertainties associated with the statistical significance of the survey in Dungarvan, this data was not used as prior knowledge. Nonetheless, the results from the survey corroborated the updated emission spatial gradients with a particular neighbourhood showing higher emissions. The higher importance of the evening peak was also corroborated by the survey, where it concludes that soil fuel usage occurs predominately in the evening.
- In order to obtain an independent validation, it is recommended to repeat the experimental campaign in the following year and compare the observations with the modelling outputs when using the updated average emission maps and time profiles. This was not possible in any of the experimental sites due to time and budget constraints.

5.2 Conclusions regarding the MapElre residential emission maps and subsequent policy decisions

- These studies provide very useful insights into the flaws (e.g., spatial distribution & time factors) in the 2019 MapElre residential emissions data when being used as input for high resolution modelling of PM concentrations.
- Compared to the initial daily average emissions, following the data assimilation, the updated emissions for the winter months are about two times higher for Edenderry and about three times for Dungarvan.
- In addition, it seems that the hourly evening peaks between 18:00 and 21:00 hours are not significantly underestimated.
- The updated daily time factor patterns demonstrate that the morning peak is less relevant than the evening peak. Additionally, the evening peak between 19:00 and 20:00 remains at a constant level and starts to decay slowly between 21:00 and 22:00 resulting in a higher time factor compared to the original values during these hours.

- For cities, towns and villages across Ireland the current standard ATMO-Street model that is used in Action B.3 of the LIFE Emerald project is underestimating the associated spatial gradients. If the emissions could be improved and higher resolution emission maps are applied, stronger spatial gradients within the urban areas would be expected.

5.3 Conclusions for air quality modelling purposes

- For locations with predominant residential emissions the modelled $PM_{2.5}$ concentrations based on the initial MapEIre emission grid, have a significant negative bias, a high random error and a low correlation with the observations.
- The updated time factors could be useful to better capture hourly peak episodes. However, it will only have a limited impact on the annual average maps. This would be the case if also the bias correction factors would be applied.
- Based on the two studies at hand, it is, however, impossible to infer whether the underestimations are similar in other periods of the year (spring, summer, autumn), or in other regions of Ireland.

5.4 Conclusions for FAIRMODE WG4

The work developed during this exercise is in line with FAIRMODE's WG4 task (Microscale assessment), where a large network of low-cost sensors is being investigated with the view of improving the air quality concentration maps for those areas. There is one major difference to note. In this project high quality sensors (Purple Air and Clarity nodes) were used when compared to the ones being considered in WG4 (SDS0111). The conclusions in this project demonstrated the following:

- Co-location analysis shows an acceptable performance of the Purple Air and Clarity nodes when compared to the uncertainties associated in the emissions.
- This study supports the concept that sensor networks can be used to further inform emission data and improve air quality maps.
- The presence of the low-cost sensors provides the opportunity to downscale the emissions and consequently the concentration maps.
- The study demonstrates the applicability of Bayesian inference methodology where both uncertainties in the input parameters and in the measured values can be considered.

5.5 Recommendations for future research studies after LIFE Emerald

- Further studies to validate both approaches:
  - Re-run the Edenderry case for 2023 using the updated emissions and time factors and compare the results with the observations from the Edenderry reference station.
  - Run the ATMO-Street model for a new town where observations are already available or are planned. Use the updated emissions based on the multilinear

regression model developed for Dungarvan and emissions factors based on those derived in Edenderry and Dungarvan.

- Repeat data assimilation studies in other regions and during other periods of the year.
- Once sufficient studies covering a wide range of region in the country becomes available, the proposed generalisation of the emissions methodology could be validated.

5.6 Implications for Action B3 and the ATMO-Street Maps

- Until further research studies are carried out, no changes will be made to the MapEIre residential emissions based on this report's findings.
- During the final phase of the project, it is recommended to liaise with t EPAs emissions team, who are currently making an update to clarify their methodology and any recent changes.
- In addition, based on this report's findings a scope of works for a large-scale research project that aims improving the residential emissions data for use in the ATMO-Street model could be prepared.